

Development of lysine-reactive covalent inhibitors and chemoproteomic probes

by
Adolfo Cuesta

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Jack Taunton

Jack Taunton

0FD9D71201FC43C...

Chair

DocuSigned by:

Kevan Shokat

Kevan Shokat

DocuSigned by:

David Agard

David Agard

328308DD7E5546A...

Committee Members

Copyright 2019
by
Adolfo Cuesta

Acknowledgements

For me, completing a doctoral dissertation was a huge undertaking that was only possible with the support of many people along the way. First, I would like to thank my PhD advisor, Jack Taunton. He always gave me the space to pursue my own ideas and interests, while providing thoughtful guidance. Nearly every aspect of this project required a technique that was completely new to me. He trusted that I was up to the challenge, supported me throughout, helped me find outside resources when necessary. I remain impressed with his voracious appetite for the literature, and ability to recall some of the most subtle, yet most important details in a paper. Most of all, I am thankful that Jack has always been so generous with his time, both in person, and remotely. I've enjoyed our many conversations and hope that they will continue. I'd also like to thank my thesis committee, Kevan Shokat and David Agard for their valuable support, insight, and encouragement throughout this project.

My lab mates in the Taunton lab made this such a pleasant experience, even on the days when things weren't working well. I worked very closely with Tangpo Yang on the mass spectrometry aspects of this project. Xiaobo Wan taught me almost everything I know about protein crystallography. Thank you as well to Geoff Smith, Jordan Carelli, Pat Sharp, Yazmin Carasco, Keely Oltion, Nicole Wenzell, Haoyuan Wang, Steve Sethofer, and Shyam Krishnan, Shawn Ouyang and Qian Zhao. Sharing an office with this wonderful group of people has been a highlight of my graduate career.

About half-way through this PhD project, it became clear that mass spectrometry was going to be a central component. As a result, I joined the Burlingame lab to learn everything I could about it. Thank you to Al Burlingame for letting me join the lab and giving me the opportunity to learn from the many talented people there. In particular, I would like to thank Giselle Knudsen, Robert Chalkley and Mark Burlingame who all volunteered to sit down with me and a few lab new lab members to teach us all the very basics about mass spectrometry. Thank

you as well to Juan Oses-Prieto, Jason Maynard, Jim Wilkins, Mike Trnka, Anatoly Urisman who have been especially generous with their time and expertise. Many thanks to Nancy Phillips, Krista Kaasik and Shenheng Guan.

From long before coming to UCSF, my family was instrumental in their unconditional support on this journey. Thank you to Adolfo Sr. and Edilma for their loving encouragement at every stage, and the many sacrifices they have made to ensure that I (and my siblings) had everything we needed. Thank you to Julian and Melina for their support, and the very often needed comic relief. Thanks as well to Jan, David and Simone, who have been endlessly generous and supportive over the last 5 years. I value our time together.

A special thanks to Mollie. Her positive energy, devoted, caring, and loving demeanor have made this process so much easier. Her optimism and excitement on the good days, and cheerleading and support on the bad days helped keep me going. Thank you.

Contributions

Chapter 1 of this thesis is reproduced in its entirety from previously published work cited below:

Cuesta, A. & Taunton, J. Lysine-Targeted Inhibitors and Chemoproteomic Probes. *Annu. Rev. Biochem.* **88**, 1–17 (2019)

Development of lysine-reactive covalent inhibitors and chemoproteomic probes

Adolfo Cuesta

Abstract

Covalent inhibitors have numerous applications as drugs, as tools for drug discovery, and as probes for chemical biology. This class of compounds depends on the availability of a suitably reactive nucleophile in the target protein of interest. Among the non-catalytic nucleophiles, cysteine is the most reactive at physiological pH, but also the least prevalent. Based on the analysis presented herein, ~80% of known binding sites in the human proteome lack a cysteine residue. In contrast, ~80% of known binding sites contain either a lysine or a tyrosine—nucleophilic residues that are much less reactive than cysteine at physiological pH. Novel methods of covalent inhibition are therefore necessary to target a larger proportion of the proteome.

We mined the Protein Data Bank (PDB) for small molecules within 5 Å of a nucleophilic atom (the epsilon amine in lysine, the phenol hydroxyl in tyrosine or the side chain thiol in cysteine). From the resulting compendium of proteins and ligands, we started with PU-H71 as an affinity scaffold to develop arylsulfonyl fluoride-based inhibitors that target a non-catalytic, non-reactive, surface-exposed lysine residue adjacent to the ATP binding site of Hsp90. We used high-resolution X-ray crystallography to decipher the molecular basis by which constrained, chiral linkers enantioselectively increase the reactivity of sulfonyl fluoride probes toward Hsp90 Lys58, without affecting the reversible binding affinity or intrinsic chemical reactivity.

The *in vivo* use of covalent inhibitors requires electrophiles with the proper balance of metabolic stability and reactivity for their targets. We used the salicylaldehyde group as a reversible covalent electrophile to develop probes that react with the catalytic lysine in kinases. Using these probes, in combination with mass spectrometry, we found that ~50% of the kinome

can react with the probes under equilibrium labeling conditions, including 80 kinases from living mice treated with the probes. Despite this promiscuity, we discovered that these probes exhibit remarkable kinetic selectivity for a small subset of kinases. The salicylaldehyde probes exhibit an apparent dissociation half-life of greater than 6 hours from a handful of kinases. These are among the first lysine-reactive kinase probes to exhibit the requisite stability and reactivity for in vivo applications. These novel probes demonstrate how to achieve kinase selectivity, despite targeting the conserved, catalytically essential lysine residue.

Table of Contents

Chapter 1: Lysine-targeted inhibitors and chemoproteomic probes.....	1
Chapter 2: Structural bioinformatic mapping of ligandable nucleophiles.....	21
Chapter 3: Ligand pre-organization drives enantioselective covalent targeting of a surface-exposed lysine	35
Chapter 4: Development and characterization of lysine-reactive aryl aldehyde kinase probes by mass spectrometry.....	71
References.....	89
Appendix A: NMR spectra	104
Appendix B: Python scripts for data processing	131
1. AA_per_pocket.py	131
2. has_good_K_v3.3_CSV_Output.py	132
3. PP_LFQ_v5.py.....	142
4. format_and_combine_protein_value-v2.py.....	146
5. batch_exp_fit_triplicates.py	150
Appendix C: Human proteins with cofactors $\leq 5\text{\AA}$ from a Lys e-amine.....	152

List of figures

Figure 1-1: Mining the PDB for lysines proximal to metabolite/cofactor binding sites.	7
Figure 1-2: Proteome-wide determination of lysine reactivity and ligandability using the clickable probe sulfotetrafluorophenyl pentynoate (STP-alkyne) and quantitative mass spectrometry.	11
Figure 1-3: Lysine-targeted chemoproteomic probes for kinases and other ATP binding proteins.	13
Figure 1-4: XO44, a broad-spectrum probe for quantifying kinase occupancy in living cells.	15
Figure 1-5: Targeting lysine and N-terminal amines with ortho-substituted aldehydes....	17
Figure 2-1: Analysis of amino acids in Swiss-prot and the representative “pocketome”. ..	24
Figure 2-2: Drug-like molecules within 5 Å of nucleophilic residues in human proteins ..	27
Figure 2-3: Histograms of nucleophile-to-ligand distances	28
Figure 2-4: Structures of ligands and their proximal lysine residues.	29
Figure 2-5: Tyrosine residues near ligands	30
Figure 2-6: Cysteine residues near ligands	30
Figure 2-7: Examples of “low-hanging fruit” for the synthesis of covalent inhibitors	31
Figure 3-1: Characterization of sulfonyl fluoride-based covalent inhibitors of Hsp90.	39
Figure 3-2: Concentration-dependent k_{obs} for the modification of Hsp90 by 1	40
Figure 3-3 MS/MS fragmentation spectra of probe 2-labeled peptides.	41
Figure 3-4: Constrained linkers result in enantioselective Hsp90 modification	42
Figure 3-5: Kinetic analysis of compound 5.	43
Figure 3-6: Crystal structure of 5 and 6 bound to Hsp90 α	44
Figure 3-7: Electron density maps and X-ray structures of Hsp90 α K58R NTD bound to 5 and 6	45

Figure 3-8: Optimized compound 5 engages Hsp90 in vivo in Skbr3 cells.	46
Figure 3-9: The effects of covalent Hsp90 modification after compound washout.	48
Figure 3-10: Direct measurement of inhibited Hsp90 turnover	49
Figure 4-1: Monitoring aldehyde washout by in-gel fluorescence	75
Figure 4-2: Kinase enrichment from Jurkat cells treated with aryl aldehyde probes	77
Figure 4-3: LC-MS/MS measurement of kinase probe half-life	78
Figure 4-4: Reproducibility of replicates and goodness of fit.	80
Figure 4-5: In vivo treatment with YTP-2-137	82

List of tables

Table 1-1: Lysine-reactive electrophiles cited in this chapter	4
Table 1-2: Number of unique human proteins (excluding kinases) with lysines ≤ 5 Å from the indicated metabolite.....	6

Chapter 1: Lysine-targeted inhibitors and chemoproteomic probes

Abstract

The development of covalent small-molecule inhibitors is a validated and widely used strategy in drug discovery and chemical biology. Although non-catalytic cysteines are frequently targeted by covalent inhibitors, many ligand-binding sites lack an accessible cysteine. Here, we review recent advances in the chemical biology of lysine-targeted covalent inhibitors and chemoproteomic probes. By analyzing crystal structures of proteins bound to common metabolites and enzyme cofactors, we identify a large set of mostly unexplored lysines that are potentially targetable with covalent inhibitors. In addition, we describe mass spectrometry-based approaches for determining proteome-wide lysine ligandability and lysine-reactive chemoproteomic probes for assessing drug-target engagement. Finally, we discuss the design of amine-reactive inhibitors that form reversible covalent bonds with their protein targets.

Introduction

The development of electrophilic small molecules that selectively react with a protein nucleophile is a major focus of chemical biology and drug discovery efforts.^{1,2} Covalent inhibitors usually function through a two-step process in which reversible noncovalent binding precedes a reversible or irreversible reaction between an electrophile on the small molecule and a proximal nucleophile on the protein target. Covalent bond formation can result in sustained target engagement (slow off-rate, long drug-target residence time) and increased pharmacodynamic potency in vivo. After an irreversible covalent drug has been administered, recovery of the target's function depends on its resynthesis rate rather than the drug clearance rate, potentially resulting in a prolonged duration of action despite rapid clearance.^{3,4} In addition, covalent ligands bearing an affinity tag or fluorophore are employed as chemoproteomic probes to quantify drug-target engagement and proteome-wide selectivity.⁵

Cysteine is the most intrinsically reactive amino acid, and even noncatalytic cysteines can react rapidly with weak electrophiles if the reacting atoms are optimally oriented. Structure-guided targeting of noncatalytic cysteines is now a prominent strategy in drug discovery. This strategy has led to the discovery of first-in-class inhibitors of protein-protein interactions, exemplified by covalent antagonists of the oncogenic Gly12Cys mutant of KRAS and the nuclear export receptor, XPO1.⁶⁻¹² In addition, targeting noncatalytic cysteines has been employed to increase the selectivity of protein kinase inhibitors,¹³⁻¹⁶ culminating in the recent FDA approval of covalent inhibitors of EGFR (afatinib, osimertinib), BTK (ibrutinib, acalabrutinib), and HER2 (neratinib).¹⁷⁻²¹ A limitation of this approach is that many ligandable sites lack a cysteine residue.

Lysine is one of the most prevalent amino acids in the proteome, with ~650,000 lysine residues distributed among ~20,000 human proteins. By contrast, there are only ~260,000 cysteines, many of which are engaged in structural disulfide bonds and therefore non-nucleophilic. Aryl sulfonyl fluorides,²²⁻²⁵ aryl fluorosulfates,^{26,27} Michael acceptors,²⁸⁻³³ dichlorotriazines,³⁴ activated esters/amides,³⁵⁻⁴² and aryl aldehydes⁴³⁻⁴⁷ have been shown to form covalent adducts with lysine residues when added to purified proteins, cell lysates, or intact cells (Table 1-1). However, compared to cysteine-targeted inhibitors, relatively few covalent inhibitors and probes (and no approved drugs) have been designed to target lysine (reviewed in⁴⁸). Hence the lysine ϵ -amine, and to a lesser extent the N-terminal amine of certain proteins, present a largely unexplored opportunity for covalent targeting (note that only ~20% of human proteins contain an unblocked N-terminal amine⁴⁹). This opportunity comes with challenges, including low intrinsic nucleophilicity – most lysines are predominantly protonated at physiological pH – and a selectivity challenge due to the ubiquitous distribution of lysines throughout the proteome, with each protein having 32 lysines on average. Although certain lysines are more intrinsically reactive than others (e.g., due to microenvironment-perturbed pK_a), identification of such lysines by computational prediction or empirical approaches is challenging.

In this review, we discuss strategies for (1) identifying ligandable lysines in proteins and proteomes, and (2) developing lysine-targeted covalent inhibitors and chemoproteomic probes.

Table 1-1: Lysine-reactive electrophiles cited in this chapter

Name	Electrophile	Lysine adduct	Major nucleophile	Reference(s)
Aryl sulfonyl fluoride			Lysine, tyrosine, serine	22–25
Aryl fluorosulfate			Lysine, tyrosine, serine	26, 27
Vinyl sulfone/sulfonamide			Cysteine, lysine	29, 31
Methyl fumaramide			Cysteine, lysine	30, 33
Acrylate			Cysteine, lysine	32
Dichlorotriazine			Lysine	34
Activated ester/amide			Lysine	35–42

Name	Electrophile	Lysine adduct	Major nucleophile	Reference(s)
Activated ester/amide (cont.)			Lysine	35–42
Salicylaldehyde			Lysine (N terminus)	43–45
2-Formyl phenyl boronic acid			Lysine	46, 47

Mining the Protein Data Bank for ligandable lysines

Structural analysis, along with empirical studies using chemoproteomic probes, suggest that each of the >500 human protein kinases has an active-site lysine that can react with a covalent inhibitor (described in more detail below). To expand this analysis to other protein classes, we searched the protein structure data bank (PDB) for lysines in close proximity to the following metabolites (and stable analogs), which are widely employed as enzyme substrates, cofactors, and allosteric modulators: ADP/ATP, GDP/GTP, α -ketoglutarate, NAD(P)/NAD(P)H, and S-adenosyl methionine (SAM). We limited our PDB analysis to crystal structures with a

resolution of 3 Å or better. Excluding protein kinases, we identified 470 human proteins (based on Uniprot ID) bound to one or more of these metabolites (Table 1-2). Strikingly, ~75% have at least one lysine within 5 Å of the metabolite (measured from the lysine ε-amine). The entire curated list, which includes the Uniprot ID, the bound metabolite, and the position of the proximal lysine(s), is presented in Supplementary Table 1. ATPases and GTPases are highly represented in this group, consistent with the known phosphate-binding and essential catalytic roles of lysine in these enzymes. In addition, most of the identified α-ketoglutarate binding proteins have a lysine that interacts with a carboxylate of the cofactor (e.g., JmjC-domain lysine demethylases) (Figure 1-1A).

Table 1-2: Number of unique human proteins (excluding kinases) with lysines ≤5 Å from the indicated metabolite

	Number of Lysine Residues			
	0	1	2	≥3
GTP	3	7	47	38
ATP	24	81	39	12
AKG	10	24	0	0
NAD(P)(H)	23	42	36	10
SAM	50	23	2	0

While phylogenetically diverse ATPases employ a conserved lysine to engage one or more phosphate oxygens, these proteins often differ in their fold, active-site structure, and specific interactions with ATP. For example, the AAA+ ATPase p97, DEAD-box helicase DDX19 and ABC transporter ABCB10 contact the β-phosphate of ATP using a similarly positioned lysine (Figure 1-1, B-D). Despite this shared feature, the adenine moiety forms distinct van der

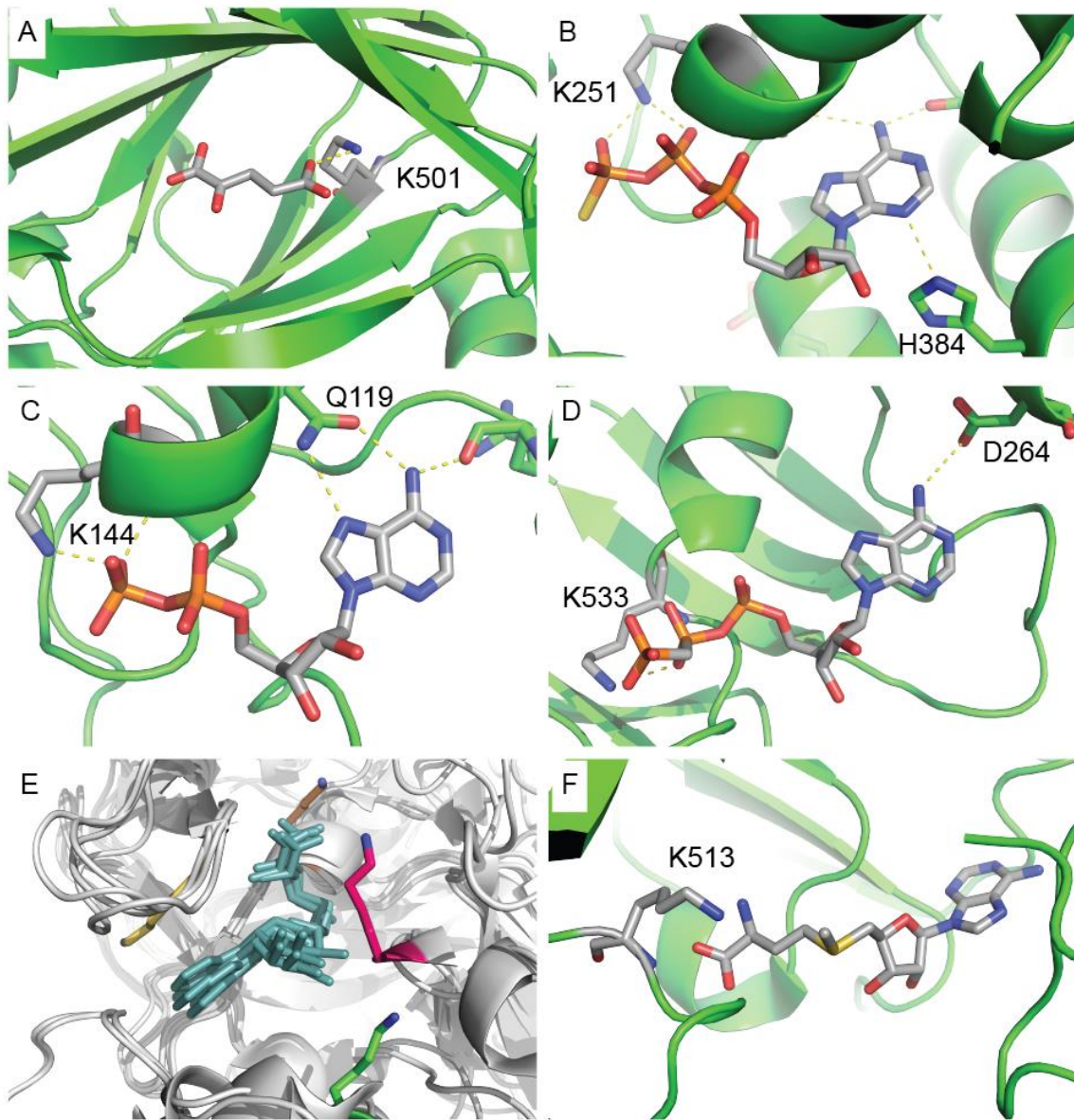


Figure 1-1: Mining the PDB for lysines proximal to metabolite/cofactor binding sites.

(a) A conserved lysine in the Jumonji C demethylase domain of KDM5A interacts with α -ketoglutarate (PDB: 5E6H). (b–d) Although a proximal lysine is found in phylogenetically divergent ATPases, they employ distinct ATP binding modes and hydrogen bonding interactions: (b) AAA+ ATPase p97/VCP (PDB: 4KO8), (c) DEAD-box helicase DDX19 (PDB: 3EWS), and (d) mitochondrial ABC transporter ABCB10 (PDB: 4AYT). (e) Overlay of four SET domains with spatially distinct lysines proximal to SAM or S-adenosylhomocysteine (cyan): KMT5A (yellow; PDB: 2BQZ), SETD7 (orange; PDB: 3CBM), SUV39H2 (magenta; PDB: 2R3A), and EHMT1 (green; PDB: 4I51). (f) SAM-proximal Lys513 in METTL3, a non-SET domain methyltransferase for N⁶-adenosine (PDB: 5IL1).

Waals and hydrogen bonding interactions with the protein in each case (Figure 1-1, B-D). By exploiting these structural differences in the ATP binding site, it should be possible to design covalent inhibitors that target this lysine, yet discriminate among distinct ATPase families. SET-family methyltransferases use SAM to methylate lysine residues on histones, transcription factors, and other proteins. Of the 26 human SET-family methyltransferase domains in the PDB, 9 have a lysine at various locations in the SAM binding site (Figure 1-1E). Targeting a poorly conserved lysine with a SAM-competitive, covalent inhibitor could serve as a strategy to increase selectivity among SET-family methyltransferases. Outside the SET family, other methyltransferases contain a SAM-proximal lysine, including the *N*6-adenosine methyltransferase METTL3, which was recently implicated in acute myeloid leukemia^{50,51} (Figure 1-1F).

Because we limited our analysis to structurally characterized proteins and a small set of metabolites/cofactors, it likely represents the tip of the iceberg in terms of ligandable lysines in the human proteome. To expand the 'ligandable lysinome' further, we propose the following: (1) structural bioinformatic analysis of all protein structures in the PDB – as well as homologous proteins not in the PDB – with the aim of identifying lysines proximal to all potentially ligandable sites, including allosteric sites and protein-protein interfaces; (2) empirical interrogation of the proteome using lysine-reactive chemical probes (see below).

Determinants of lysine reactivity and ligandability

In addition to its proximity to a small molecule binding site, a lysine's intrinsic nucleophilicity is a critical determinant of ligandability governed, in part, by its pK_a . In solution, the pK_a of the protonated ϵ -amine of lysine is 10.6, while in proteins it can range from ~5 to ~11 depending on the local microenvironment (e.g., local dielectric, proximity of negatively and positively charged side chains).⁵² While it's difficult to measure, let alone predict the pK_a of

specific lysines, one can measure their relative reactivity toward simple electrophiles using quantitative mass spectrometry. In a recent study, acetyl CoA and acetyl phosphate were used to measure the relative reactivity of 90 lysines across 8 purified proteins. The second-order rate constants spanned three orders of magnitude, with 30% of the lysines being unreactive.⁵³ Decreased reactivity correlated in many cases with ionic interactions between the Lys and a proximal Asp or Glu (e.g., ACAT1 Lys84). The most reactive lysines (e.g., ACAT1 Lys304, GDH Lys503) were on the protein surface, close to other positively charged residues. The increased reactivity of these lysines can be explained by their perturbed pK_a , as well as accessibility to the charged electrophiles. This raises a potential caveat to these types of measurements: the reactivity of solvent-exposed surface lysines toward small, charged active esters may not reflect reactivity toward lipophilic, sterically demanding electrophiles. That is, a lysine's 'intrinsic reactivity' may vary depending on the specific physicochemical properties of the test electrophile.

Using advanced chemoproteomic methods, it should be feasible to measure the relative reactivity of every lysine in the proteome toward a model electrophilic probe; in practice, this remains challenging due to the massive number of lysines. In a recent study,⁴¹ proteome-wide lysine reactivity was estimated by treating cell lysates with two concentrations of a clickable lysine-reactive probe, sulfotetrafluorophenyl pentynoate (STP-alkyne, Figure 1-2A). A similar study used a clickable NHS ester as the probe.⁴² After copper-catalyzed click conjugation to a cleavable biotin-azide and affinity purification with streptavidin-agarose beads, probe-modified peptides were identified by mass spectrometry. The difference in labeling intensity between the two probe concentrations was used to infer relative reactivity; those lysines with roughly equivalent labeling intensity under both conditions were deemed 'highly reactive' (labeled to saturation at the lower probe concentration), whereas those with increased labeling proportional to the probe concentration were deemed to have moderate or low reactivity. Using this workflow, the relative reactivity of >4,000 lysines was estimated using proteomes derived from

three human cancer cell lines. Hyper-reactive lysines (<10% of lysines modified by STP-alkyne) were enriched in enzyme active sites and ligand binding sites.

Competition experiments using STP-alkyne and a panel of 32 lysine-reactive fragments (NHS, pentafluorophenyl, and other active esters) were used to identify lysines proximal to saturable fragment binding sites, i.e. 'ligandable lysines'. Only a minor fraction (<2%) of the STP-alkyne-modified lysines were competed by pretreating lysates with one or more active ester fragments. Some of the ligandable lysines were found in challenging 'classically undruggable' targets, including protein-protein interaction domains. For example, several fragments modified the transcriptional repressor Sin3A at Lys155, which inhibited Sin3A binding to TGIF1, a DNA-binding transcription factor. The dataset of 121 ligandable lysines in 113 human proteins can potentially serve as a resource for structure-based design of new covalent probes. As an example, we highlight the E3 ubiquitin ligase subunit CUL2. As shown in the supplementary dataset accompanying the Hacker et al. study,⁴¹ CUL2 Lys489 was labeled by two active ester fragments, including compound **1** (Figure 1-2A). A recent crystal structure of a multi-subunit E3 ligase complex reveals that CUL2 Lys489 is adjacent to a shallow pocket at the interface of CUL2 and the E2-binding subunit, RBX1 (Figure 1-2B). Covalent modification of this lysine with small-molecule ligands that bind to this interface could potentially inhibit CUL2 E3 ligase activity; alternatively, targeting this site with a bifunctional ligand could be used to promote ubiquitination and degradation of neosubstrates.⁵⁴⁻⁵⁷

The dataset reported by Hacker et al.⁴¹ provides the most comprehensive view of lysine reactivity to date. Nevertheless, it captures only a small fraction of the total number of lysines in the proteome. For example, although STP-alkyne labeled 50 kinases, only 12 probe-modified lysines correspond to the kinase catalytic lysine. By contrast, the catalytic lysine in >300 protein kinases has been shown to react rapidly with active site-directed acyl phosphate and sulfonyl fluoride probes (see below).^{24,25,40} Efficient modification of the kinase catalytic lysine, situated in

a deep hydrophobic pocket, may require tailored probes with higher intrinsic affinity for the ATP binding site. We speculate this may be true for many other ligandable lysines.

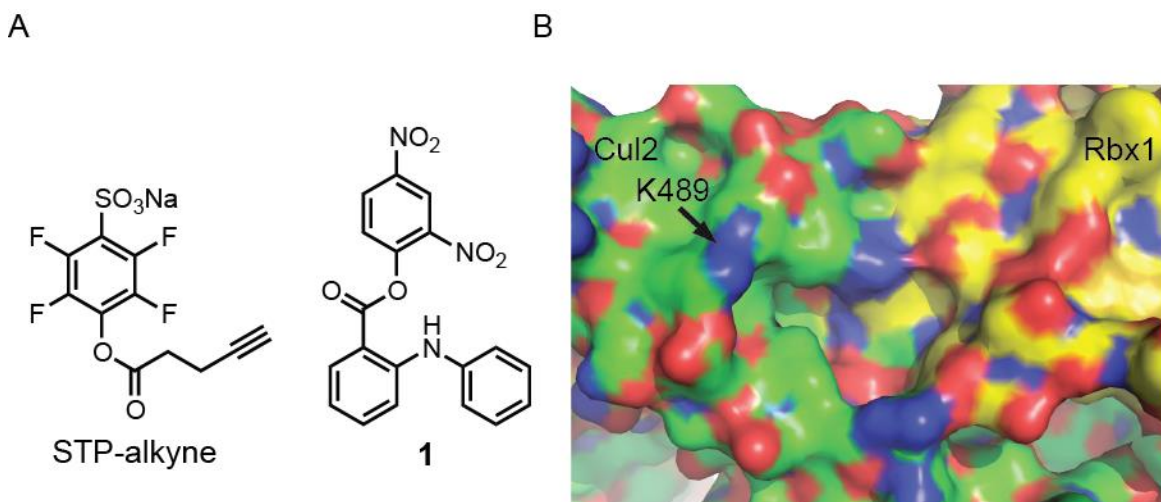


Figure 1-2: Proteome-wide determination of lysine reactivity and ligandability using the clickable probe sulfotetrafluorophenyl pentynoate (STP-alkyne) and quantitative mass spectrometry.

(a) Chemical structure of STP-alkyne and reactive ester fragment **1**. Fragment **1** was found to prevent covalent modification of E3 ubiquitin ligase subunit CUL2 at Lys489 by STP-alkyne (41). (b) CUL2 Lys489 is adjacent to a pocket at the interface with RBX1 in an E3 ligase complex (Protein Data Bank: 5N4W).

Chemoproteomic probes targeting the catalytic lysine in protein kinases

Protein kinases comprise a large family of signal transducing enzymes (>500 in humans) that regulate all cellular processes. Kinases are validated therapeutic targets for cancer, rheumatoid arthritis, and pulmonary fibrosis, and they continue to be a central focus of drug discovery efforts for many other diseases. In addition to their prominent role as drugs, selective kinase inhibitors often provide new mechanistic insights that are not easily obtained through genetic approaches.⁵⁸

Given the highly conserved ATP binding site and the presence of hundreds of kinases in any given cell type, a key question inevitably arises with the use of any kinase inhibitor: when the inhibitor is added to cells at a concentration that induces a phenotype of interest (e.g., cancer cell death), which kinases are occupied by the inhibitor and to what extent? Although

selectivity is typically assessed by profiling a panel of purified recombinant kinases or kinase domains, recent chemoproteomic methods have been developed to interrogate hundreds of endogenously expressed kinases in a single experiment, either in cell lysates or intact cells.^{40,59,60} Such methods have the potential advantage of quantifying inhibitor/kinase interactions in a more physiologically relevant context.

ADP- and ATP-desthiobiotin are acyl phosphate-based, chemoproteomic probes that were designed to react with a proximal lysine residue in protein kinases and other ATP-utilizing enzymes.³⁹ In protein kinases, the primary modification site is the conserved and catalytically essential lysine, which coordinates the ADP leaving group during the phospho-transfer reaction (Figure 1-3A). Other lysines just outside the kinase active site are also modified to varying extents. The acyl phosphate probes contain a desthiobiotin moiety attached via its carboxylic acid to the β -phosphate of ADP or the γ -phosphate of ATP. A lysine in close proximity to the terminal phosphate of the reversibly bound probe attacks the carbonyl to form a stable desthiobiotin carboxamide, with expulsion of the ADP/ATP leaving group (Figure 1-3B).

In a typical chemoproteomic profiling experiment, ATP-desthiobiotin (5 μ M) is added to gel-filtered (ATP-depleted) cell lysates in the absence or presence of a competing kinase inhibitor. After exhaustive trypsinization, desthiobiotinylated peptides are enriched with streptavidin beads and analyzed by quantitative mass spectrometry. By comparing MS intensities of kinase-derived, desthiobiotinylated peptides from control vs. inhibitor-treated lysates, one can estimate inhibitor occupancy of up to 150-200 endogenous kinases from a single cell line and >300 kinases across multiple cell lines.^{40,61-67} Reproducible detection and quantification of hundreds of unique kinases – each of which can contribute only 1 or 2 probe-modified peptides – requires a significant time investment to develop a robust mass spectrometry pipeline.⁶⁸⁻⁷⁰ In addition to kinases, ADP/ATP-desthiobiotin covalently modifies

hundreds of non-kinase proteins, including HSP70 and HSP90 chaperones, in an ATP-competitive manner.⁷¹

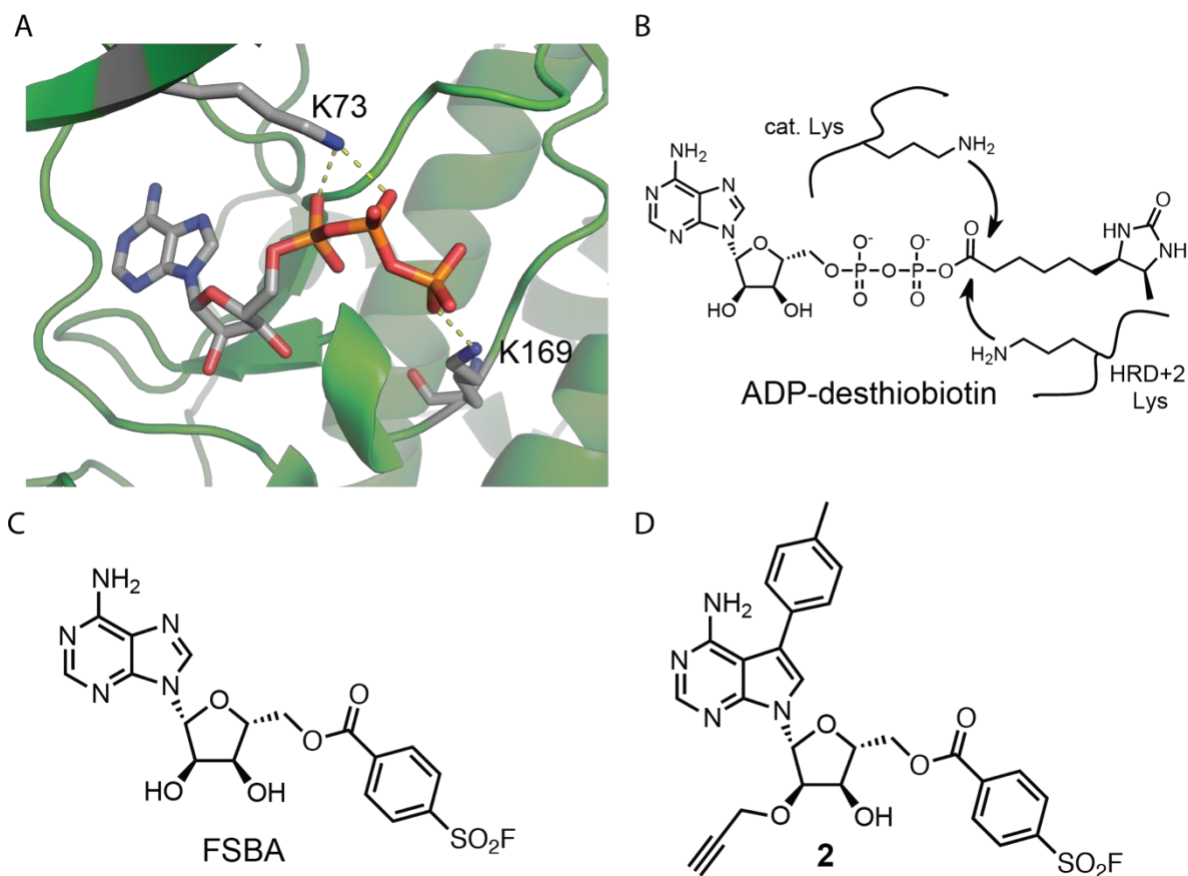


Figure 1-3: Lysine-targeted chemoproteomic probes for kinases and other ATP binding proteins.

(a) Crystal structure of protein kinase A bound to ATP showing the conserved catalytic Lys73 and proximal Lys169 (Protein Data Bank: 1ATP), which is conserved in serine/threonine kinases. (b) ADP- and ATP-desthiobiotin probes acylate the kinase catalytic lysine and other ATP-proximal lysines in cell lysates. Chemical structures of (c) 5'-fluorosulfonylbenzoyl adenosine (FSBA) and (d) clickable sulfonyl fluoride probe 2, which potently labels SRC family kinases in cells.

Sulfonyl fluorides react with multiple protein nucleophiles, yet they are relatively stable to hydrolysis at physiological pH.^{72–74} Reactions with amines and hydroxyls (Lys, Tyr, Ser, Thr) produce stable sulfonamides and sulfonates, whereas reactions with thiols and imidazoles (Cys, His) result in adducts that may be labile and difficult to isolate.^{74–77} Sulfonyl fluorides have been used extensively as serine hydrolase inhibitors and nucleoside-based affinity probes for myriad

enzymes. However, there are relatively few examples of sulfonyl fluoride probes designed to target a specific Lys in a protein active site.

5'-fluorosulfonylbenzoyl adenosine (FSBA) binds the ATP site of numerous proteins, including protein kinases, with affinities ranging from high micromolar to low millimolar (Figure 1-3C). Although FSBA selectively modifies the catalytic lysine of purified protein kinases,⁷⁸ its utility as a chemoproteomic kinase probe is limited. To enable covalent, lysine-targeted kinase inhibition in cells, we designed the sulfonyl fluoride probe **2** (Figure 1-3D), a hybrid of FSBA and the potent SRC-family kinase inhibitor, PP1.²⁵ The *p*-tolyl group on its pyrrolopyrimidine core was designed to fill a hydrophobic pocket in SRC and other kinases with a small amino acid at the gatekeeper position, thereby increasing the affinity by ~10,000-fold relative to FSBA. It also contains a 2'-propargyl ether, enabling copper-catalyzed click conjugation to biotin or rhodamine azides. Probe **2** rapidly labeled SRC in live cells at nanomolar concentrations (IC₅₀ ~300 nM), and labeling was not observed when the catalytic Lys was mutated to Arg.

To expand the number of endogenous kinases amenable to detection by cell-permeable sulfonyl fluoride probes, we designed XO44, which links a phenylsulfonyl fluoride to a highly promiscuous kinase-binding scaffold (Figure 1-4A).²⁴ Covalent modification of the catalytic lysine was confirmed by mass spectrometry as well as co-crystal structures of XO44 bound to SRC and EGFR (Figure 1-4B). When added to live cells, XO44 modified 133 endogenous kinases, as revealed by mass spectrometry analysis (Figure 1-4C). This represents by far the largest fraction of the kinome labeled by a single probe in live cells. Using XO44 and label-free quantitative mass spectrometry, we observed selective engagement of endogenous ABL1, BLK, SRC, and LCK by clinically relevant concentrations of the myeloid leukemia drug, dasatinib.

A remarkable feature of XO44, as shown by chemoproteomic analysis, is its 'selective promiscuity'. Other than protein kinases, most of which were labeled to saturation by XO44

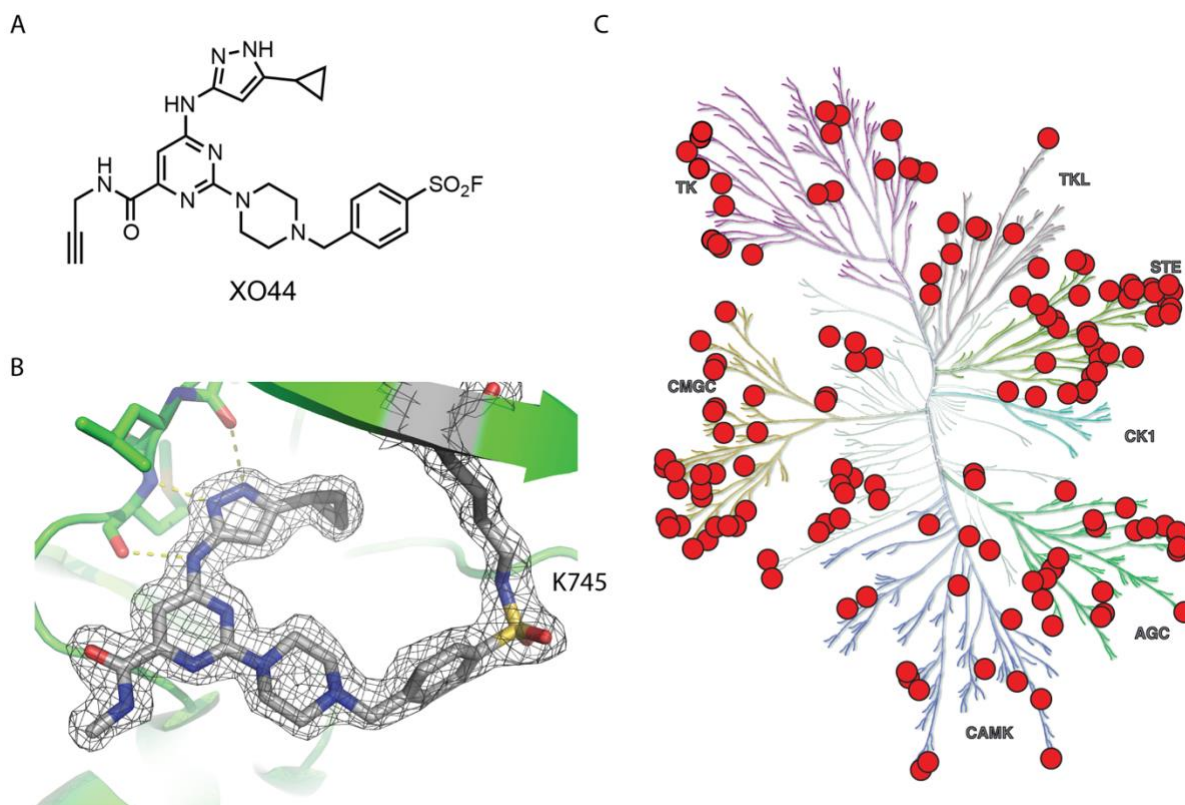


Figure 1-4: XO44, a broad-spectrum probe for quantifying kinase occupancy in living cells.

(a) Chemical structure of XO44, comprising an aminopyrazole noncovalent kinase-recognition moiety, a piperazine linker, and a phenylsulfonyl fluoride. (b) Crystal structure of XO44 covalently bound to the EGFR kinase domain (Protein Data Bank: 5U8L) (24). (c) Kinase dendrogram showing kinases covalently labeled by XO44 in Jurkat cells. Figure adapted with permission from Reference 24; copyright 2017 American Chemical Society.

within 30 min, very few proteins were specifically enriched and identified by mass spectrometry. These results suggest that sulfonyl fluorides such as XO44 occupy a 'sweet spot' in electrophile space, efficiently reacting with an appropriately positioned lysine, while reacting slowly with most off-target nucleophiles when added to cells at low micromolar concentrations. As mentioned above, sulfonyl fluoride probes have been designed to target other protein nucleophiles, including a tyrosine in the mRNA decapping enzyme, Dcps.⁷⁹ In the case of Dcps, there is a lysine within 7 Å of the modified tyrosine. It remains to be determined which specific features in the protein microenvironment favor sulfonyl fluoride modification of tyrosine vs. lysine.

Targeting noncatalytic lysines and N-terminal amines with reversible covalent inhibitors

Reversible covalent inhibitors have the potential to exhibit sustained target engagement, high selectivity, and a reduced propensity to form adducts with off-target nucleophiles. An example of this approach is provided by cyanoacrylamide-based kinase inhibitors, which covalently modify a noncatalytic cysteine proximal to the ATP-binding site.^{80–83} Imine (Schiff base) formation between an aldehyde-containing small molecule and a protein-derived amine – either a lysine ϵ -amine or an unblocked amino terminus – represents a conceptually related strategy for developing reversible covalent inhibitors with potential application to proteins that lack a suitably positioned cysteine. In this section, we highlight three recent examples of *ortho*-substituted benzaldehydes that form an imine with a specific amine on their protein target.

The *o*-hydroxy benzaldehyde (salicylaldehyde) GBT440 is in late-stage clinical development for sickle cell disease. GBT440 forms an imine with the N-terminal amine of α -hemoglobin (Figure 1-5A), stabilizing the oxygen-bound state and preventing polymerization of mutant α/β -hemoglobin tetramers.⁸⁴ GBT440 analogs that lack the aldehyde are inactive, while those that lack the *o*-hydroxyl are less potent *in vitro* and show reduced erythrocyte occupancy *in vivo*.⁴³ The crystal structure of GBT440 bound to the $\alpha_2\beta_2$ hemoglobin tetramer reveals an intramolecular hydrogen bond between the *o*-hydroxyl and the imine nitrogen (Figure 1-5A), which may contribute to GBT440's long residence time *in vivo*. Kinetic studies of model benzaldehyde-derived imines have shown that an *o*-hydroxy group decreases the hydrolysis rate by at least 40-fold compared to an *o*-methoxy or *m*-hydroxy group.⁸⁵ This profound kinetic stabilization effect has been attributed to intramolecular hydrogen bonding in the context of a phenoxy-iminium zwitterion (Figure 1-5A). Similar trends have been observed in kinetic measurements of transamination reactions comparing benzaldehyde- vs. salicylaldehyde-derived imines.⁸⁶ GBT440 also forms an intermolecular hydrogen bond between a pyrazole

nitrogen and Ser132 on the second α -chain, as well as hydrophobic interactions on both sides of the α -hemoglobin dimer interface. These interactions prevent binding of a second GBT440

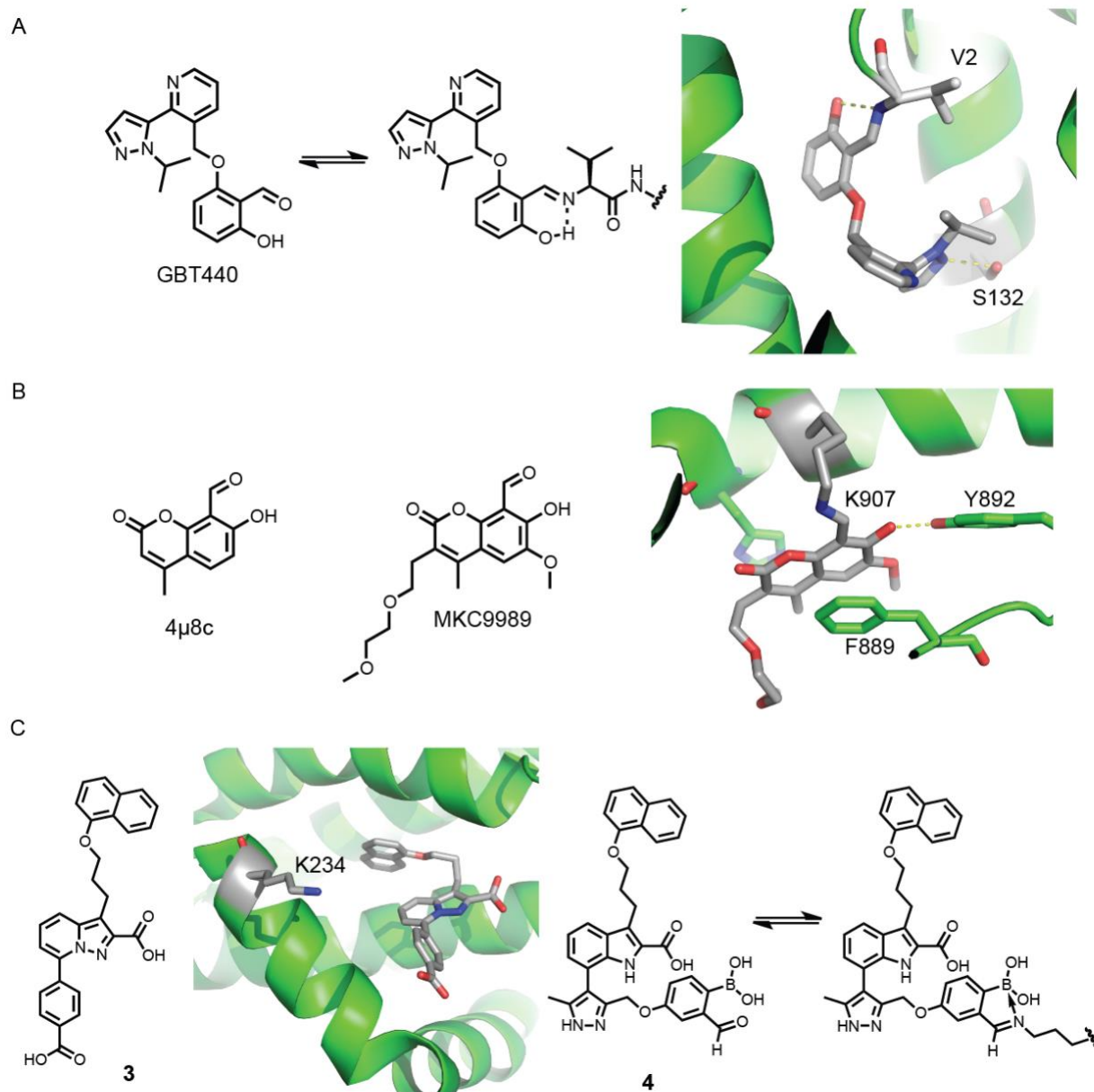


Figure 1-5: Targeting lysine and N-terminal amines with ortho-substituted aldehydes.

(a) Salicylaldehyde inhibitor GBT440 forms a reversible imine with the N-terminal amine (Val2) of α -hemoglobin [Protein Data Bank (PDB): 5E83]. An intramolecular hydrogen bond may stabilize the covalent adduct. (b) Salicylaldehyde inhibitors 4 μ 8c and MKC9989 reversibly target Lys907 in the RNase domain of IRE1 (PDB: 4PL3). The imine is shielded from solvent by Phe889. (c) The crystal structure of compound 3 bound to MCL1 (PDB: 3WIX) guided the design of o-boronic acid benzaldehyde 4, a reversible covalent MCL1 inhibitor that targets Lys234.

molecule to the symmetry-related site; this differs from earlier hemoglobin stabilizers, most of which bind with 2:1 stoichiometry.^{87–89} Hence, similar to cysteine-targeted reversible covalent kinase inhibitors,^{80,83} covalent and noncovalent interactions cooperatively stabilize a long-lived complex between GBT440 and the hemoglobin tetramer.

Salicylaldehydes (and salicylaldimines) have also been discovered to inhibit the unconventional RNA splicing activity of IRE1,^{90,91} a transmembrane protein kinase and RNase that signals in response to unfolded protein accumulation in the endoplasmic reticulum. Compound 4μ8c (8-formyl-7-hydroxy-4-methylcoumarin) and other related salicylaldehydes form reversible imines with Lys907, a noncatalytic lysine in the RNase domain of IRE1 (Figure 1-5B).⁴⁴ Benzaldehydes that lack the *o*-hydroxyl group are inactive.^{45,91} Molecular dynamics simulations suggest that Lys907 has a shifted pK_a of ~7 (compared to a pK_a of 10-11 for fully solvated lysine).⁹² Lys907 is buried in a hydrophobic pocket and shielded from solvent, potentially contributing to its decreased pK_a and increased nucleophilicity. Crystal structures of IRE1 bound to salicylaldehyde inhibitors confirm imine formation with Lys907,⁴⁵ but the orientation of the imine nitrogen precludes intramolecular hydrogen bonding (Figure 1-5B). Instead, the imine nitrogen forms a hydrogen bond with the catalytically essential Tyr892. The imine is further shielded from solvent by a pi-stacking interaction with Phe889. Solvent accessibility of the imine – which must undergo nucleophilic attack by a water molecule prior to dissociation of the inhibitor – is likely a major determinant of residence time for amine-targeted aldehyde inhibitors, in addition to intra- and intermolecular stabilizing interactions.

The salicylaldehyde-based ligands described above provide evidence that a hydroxyl substituent *ortho* to the electrophilic aldehyde can have a profound impact on potency. A similar concept, based on intramolecular imine stabilization, was employed in the design of benzaldehydes (and acetophenones) with an *o*-boronic acid substituent. *o*-Boronic acid aldehydes react with amines to form iminoboronates, whose thermodynamic stability is thought to derive from an intramolecular dative bond between the imine nitrogen and boron⁴⁶ or an ionic

interaction between the protonated imine and the negatively charged boronate.⁹³ Starting with the noncovalent ligand **3**, *o*-boronic acid aldehyde **4** was designed to target a noncatalytic surface lysine (Lys234) near the BH3 binding groove of Mcl-1, an anti-apoptotic regulator that promotes cancer cell survival⁴⁷ (Figure 1-5C). Compound **4** binds Mcl-1 tightly ($IC_{50} \sim 5$ nM), and mutation of Lys234 to Ala results in a ~ 30 -fold reduction in binding affinity. Both the aldehyde and the boronic acid are critical for potency, since analogs lacking either exhibit 20-50-fold lower binding affinity for Mcl-1 and >20 -fold lower cellular potency. Compared to cysteine-targeted, reversible covalent kinase inhibitors with residence times of hours to days,^{80,83} *o*-boronic acid aldehyde **4** dissociates from Mcl-1 much faster ($t_{1/2} \sim 38$ s). Only limited kinetic data are available for other amine-targeted aldehyde-based inhibitors, including the salicylaldehyde ligands for IRE1 and hemoglobin described above. In this regard, it would be of interest to directly compare the hydrolysis rates of model imines (in pH ~ 7 buffer) and protein/ligand adducts derived from *o*-boronic acid aldehydes, salicylaldehydes, and other substituted benzaldehydes.

In summary, aldehyde-based covalent inhibitors – in particular, salicylaldehydes and *o*-boronic acid benzaldehydes – have been designed or discovered to reversibly target an amine (N-terminal amine or lysine ϵ -amine) proximal to a ligand-binding site on α -hemoglobin (allosteric site), IRE1 (RNAse active site), and Mcl-1 (protein-protein interaction site). Affinity, selectivity, and residence time are likely determined by a combination of tunable properties, including the intrinsic stability of the product imine – its solvent accessibility, bound conformation, and ability to form intramolecular interactions – as well as noncovalent intermolecular interactions between the ligand and the protein. It should be possible to optimize these cooperative stabilizing interactions by structure-based design. Moreover, the strategy of targeting lysines and N-terminal amines with reversible covalent inhibitors should be applicable to any target with a suitably reactive amine proximal to a ligand binding site.

Conclusions and perspective

Broad application of covalent inhibitors and probes to protein sites that lack a ligandable cysteine requires new approaches in medicinal chemistry, chemical biology, and chemoproteomics. The total number of proteins with a 'ligandable lysine' is currently unknown. Nevertheless, our analysis of the PDB suggests that the majority of enzyme cofactor binding sites – and perhaps by extension, other ligand binding sites – contain at least one proximal lysine. Although quantitative chemoproteomic studies have begun to reveal the reactivity and ligandability of lysines on a proteome-wide scale,^{41,42} we argue that the amine-reactive probes deployed thus far have captured only a small fraction of the 'ligandable lysinome', implied by the absence of many active-site lysines in published chemoproteomic datasets. Which lysines can efficiently capture a proximal electrophile in living cells and animals, and which electrophile is best suited for a given lysine are key unanswered questions for future research. Innovative chemistry – for the discovery of new electrophiles and their precise installation onto diverse noncovalent recognition scaffolds – will play a critical role in the expansive application of covalent inhibitors and chemoproteomic probes to lysines, as well as other protein nucleophiles like tyrosine, serine, and histidine. Finally, advances in multiplexed quantitative mass spectrometry and chemoproteomic probe design will be required to fully explore the vast uncharted universe of ligandable protein nucleophiles in cells and organisms.

Acknowledgements

We thank members of the Taunton lab, including Geoff Smith, Phillip Sharp, and Jordan Carelli, for comments on the manuscript. This work was supported by Pfizer (J.T.) and the National Cancer Institute (A.C., grant number F31CA214028).

Chapter 2: Structural bioinformatic mapping of ligandable nucleophiles

Abstract

Covalent inhibitors frequently react with a catalytic nucleophile in the target protein's binding site or with a non-catalytic cysteine, both of which exhibit high intrinsic reactivity. Rarely are they designed to react with potentially less reactive non-catalytic nucleophiles like lysine, tyrosine, serine or histidine. The fraction of the proteome that lacks ligandable cysteines is currently unknown. In this chapter, we performed an analysis of previously described pockets in the Protein Data Bank (PDB) and find that ~80% of such pockets lack a cysteine residue. In contrast, ≥80% of pockets contain either a lysine or tyrosine residue, motivating the development of chemical approaches to covalently target these less reactive amino acids. To identify all small-molecule ligands (excluding enzyme cofactors) in the PDB situated within 5 Å of the nucleophilic atom of lysine, tyrosine, and cysteine in human proteins. We highlight several protein targets with ligands with low affinity or poor selectivity, which could potentially be improved by conversion into a covalent inhibitor.

Introduction

Irreversible and reversible covalent inhibitors have numerous advantages over their noncovalent counterparts, including increased potency and prolonged residence time on their targets. Regardless of the specific application, the use of covalent inhibitors requires a suitably positioned reactive amino acid, usually a nucleophile, within the protein target of interest. As a result, covalent inhibitors are frequently designed to react with residues involved in catalysis, including Cys, Ser, Thr, and Lys (e.g., activity-based probes)⁹⁴. These catalytic residues are often more intrinsically nucleophilic than the same residue in a non-catalytic context. An alternative approach, which has been widely used by our lab and others over the past 10 years, is to design electrophilic inhibitors that react with a non-catalytic cysteine,^{13,14,58,62,63,80–82,95} which has the highest intrinsic chemical reactivity of the 20 common amino acids. There are

comparably fewer examples of covalent inhibitors designed to react with other potentially nucleophilic residues (e.g., Lys, Tyr, Ser, His), especially those not involved in enzyme catalysis. While the development of chemistry to selectively react with non-cysteine nucleophiles in proteins is an area of active research, the potential impact of such efforts – that is, the total number of “ligandable nucleophiles” in a given proteome – is unknown. In this chapter, we mine the Protein Data Bank (PDB) to prospectively identify such nucleophiles, which can potentially be targeted with electrophilic covalent inhibitors. More specifically, we analyzed the occurrence of nucleophilic residues in previously identified pockets in the PDB, and we measure their distance to drug-like small molecules, excluding the common cofactors and enzyme substrates discussed in Chapter 1. The resulting dataset provides a resource for generating new probe development projects, as well as potential chemical starting points for the design and synthesis of targeted covalent inhibitors.

Results

Cysteine is one of the least prevalent nucleophilic residues in the human proteome (Fig. 2-1a). Moreover, many cysteines are non-nucleophilic due to the formation of kinetically stable, structural disulfides in the extracellular and luminal domains of secretory and membrane proteins. Lysine and tyrosine are 2-3 times as prevalent as cysteine (Fig. 2-1a), yet most lysines and tyrosines are less reactive than cysteine at physiological pH. Nevertheless, electrophiles such as arylsulfonyl fluorides^{22,24,79} and aryl fluorosulfates^{26,27} react with lysine and tyrosine in certain contexts. While their overall prevalence implies that lysine and tyrosine-targeted covalent inhibitors can react with many more proteins and protein sites than cysteine-targeted inhibitors, the amino acid composition of certain functional, “ligandable” pockets may be subject to distinct evolutionary pressures that limit the occurrence of specific nucleophilic residues.

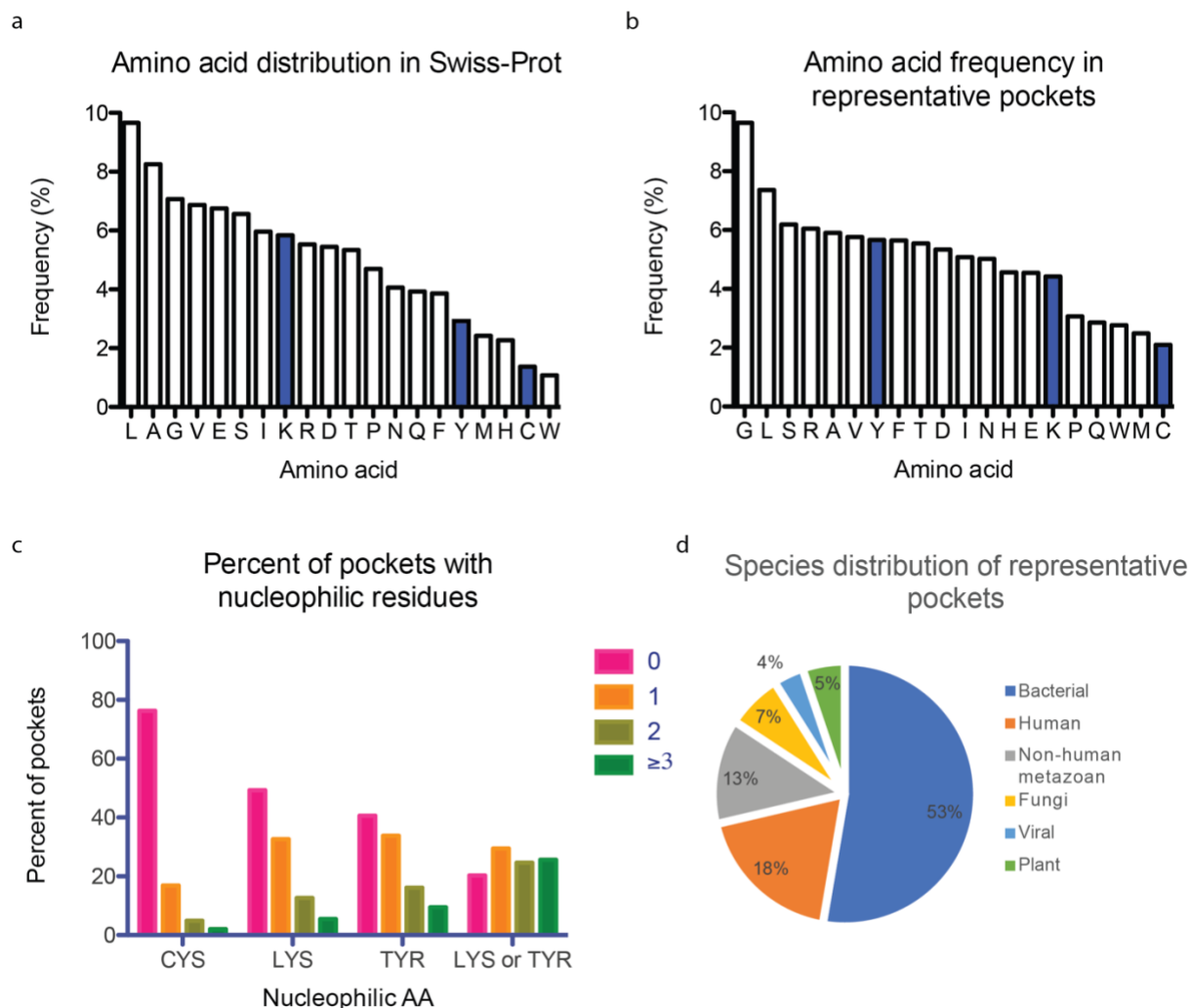


Figure 2-1: Analysis of amino acids in Swiss-Prot and the representative “pocketome”

a) The frequency of amino acids in Swiss-Prot is shown. Cysteine is one of the least prevalent nucleophilic residues, while lysine and tyrosine are 2-3 times as prevalent as cysteine. b) The frequency of cysteine, lysine and tyrosine in the 6,761 representative pockets is comparable to their overall frequency in Swiss-Prot. c) ~80% of the representative pockets lack a cysteine residue and approximately half lack a lysine or tyrosine residue. In contrast, 80% of the representative pockets contain either a lysine or a tyrosine. Many of these pockets can theoretically react with a sulfonyl fluoride or aryl fluorosulfate-based probe. d) Distribution of species in the 6,761 representative pockets.

Recent work from the Chandra group used 3 complementary pocket prediction algorithms (SiteHound, PocketDepth and FPocket) to identify pockets within X-ray structures in the PDB.⁹⁶⁻⁹⁹ SiteHound finds protein pockets by identifying clusters of energetically favorable interactions between the protein and a small test probe (e.g. a methyl group).⁹⁷ FPocket

generates “alpha spheres,” which are spheres with a 3-6 Å radius that contact 4 protein atoms simultaneously. These spheres are clustered and pockets are identified based on cluster size and the hydrophobicity of the amino acids contacting the cluster of alpha spheres.⁹⁹ Lastly, PocketDepth identifies pockets by drawing a line between all pairs of atoms that are 2-15 Å apart and contact the protein surface. Lines that pass through the interior of the protein are prohibited. Regions that exhibit the highest density of connecting lines are identified as pockets by their algorithm.⁹⁸ To be included in the Chandra group’s study, a pocket must be identified by all three methods.

Starting from all 113,486 structures in the PDB (from May 2015), they clustered structures with $\geq 95\%$ sequence identity to reduce data complexity and redundancy. Of the resulting 30,457 protein clusters (and $\sim 249,000$ pockets), they selected 6,761 clusters containing at least two proteins and at least one liganded protein to validate their pocket predictions. They found that the representative pocket in each cluster correctly defined the pocket found in 83% of the 55,386 member proteins in the 6,761 clusters. The amino acid sequence of the 6,761 pockets was reported in the supplementary dataset accompanying their study.

Using this dataset containing 6,761 representative pockets from all domains of life (Fig. 2-1d), we analyzed the occurrence of all amino acids in each representative pocket and found that cysteine is the least prevalent residue, while lysine and tyrosine each occur almost twice as frequently as cysteine in these pockets (Fig. 2-1b), similar to the distribution of these residues across the entire proteome. Approximately 80% of these pockets are devoid of cysteines and are therefore not amenable to a cysteine-targeted covalent inhibitor approach (Fig. 2-1c). By contrast, lysine and tyrosine residues are present in $\sim 50\%$ and $\sim 60\%$ of the pockets, respectively. Together, $\sim 80\%$ of the identified pockets contain either a lysine, a tyrosine or both, suggesting that sulfonyl fluoride or fluorosulfate-based covalent inhibitors could be developed

for a large fraction of computationally defined pockets, and perhaps a large fraction of the proteome. This compares favorably to cysteine, which is found in only ~20% of defined pockets.

While lysine and tyrosine are prevalent within pockets, it is not possible to accurately predict which specific residues will be sufficiently reactive for covalent modification. Empirical measurements of proteome-wide reactivity using small, hydrophobic test electrophiles fail to identify many known reactive sites in the proteome, suggesting that many of these sites can only be discovered with a directed ligand.^{41,100} Since the discovery of high-affinity, non-covalent scaffolds that can direct an electrophile to any site of interest is challenging, we set out to identify all of the drug-like molecules in the PDB adjacent to nucleophilic residues. We propose that the resulting compendium of ligand/protein structures represents a "low-hanging fruit" resource of small-molecule leads, which can be further elaborated into covalent inhibitors using structure-based design.

We analyzed 42,689 metazoan structures in the PDB that were solved by X-ray crystallography to a resolution of ≤ 3 Å, all of which contain a bound ligand. These ligands were filtered in an attempt to exclude buffers, precipitants, ions, detergents and biological cofactors, resulting in 15,902 compounds and 2,492 proteins after applying these filters. In each structure, the shortest distance between the small molecule ligand and the target nucleophilic heavy atom was determined (epsilon amine in lysine, hydroxyl in tyrosine, or thiol in cysteine). Only those structures of human proteins with a nucleophile-to-ligand distance of ≤ 5 Å were kept for further analysis. We identified a total of 1412 unique human proteins (including 227 protein kinases) that meet these criteria. The entire curated list, which includes the Uniprot ID, the PDB code, the ligand name and id, and the position of the proximal lysine(s), is include as an electronic Supplementary Table.

Approximately 40% of these 1,412 human proteins contained a cysteine residue within 5 Å of a small-molecule ligand (Fig. 2-2a), whereas nearly twice as many proteins were identified with lysine or tyrosine near a small molecule, including ≥ 500 proteins that lack a proximal

cysteine (Fig. 2-2a). Similarly, on a per-compound basis, only 20% of compounds are within 5 Å of a cysteine, while nearly 60% of all compounds analyzed are within 5 Å of a lysine or tyrosine (Fig. 2-2b). In principle, cysteines identified through this analysis can be targeted with a covalent inhibitor, due to the intrinsic reactivity of the partially deprotonated thiol at physiological pH. While many more lysine and tyrosine residues are found near small molecules than cysteine residues, it remains to be determined which of lysines and tyrosines can react efficiently with appropriately positioned electrophiles (e.g., $t_{1/2} \leq 10$ min), particularly if the side chain is on the protein surface. Solvation and surface exposure is expected to result in decreased nucleophilicity for lysines and perhaps increased nucleophilicity for tyrosines. Nevertheless, our analysis identifies a large set of compounds near lysine and tyrosine residues on specific protein targets. We propose that these compounds can be used as starting scaffolds to design lysine and tyrosine-directed covalent ligands, for example, by structure-guided installation of sulfonyl fluoride, fluorosulfate, or aldehyde electrophiles.

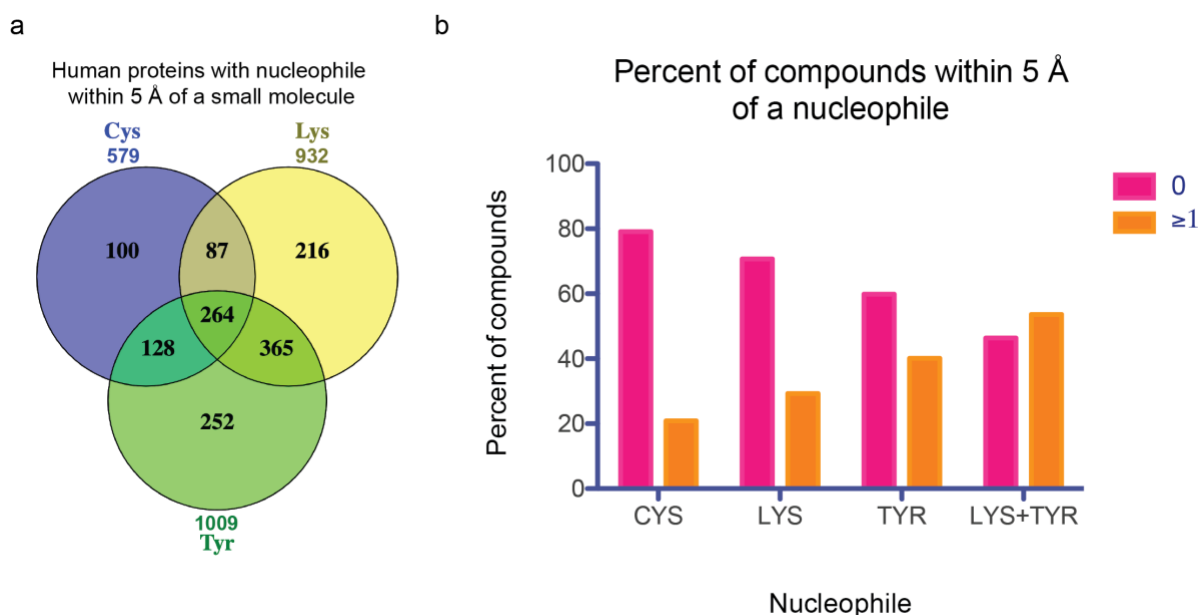


Figure 2-2: Drug-like molecules within 5 Å of nucleophilic residues in human proteins

a) Venn diagram showing the overlap of human proteins containing small molecules within 5 Å of the indicated residue. b) Bar graph showing the percentage of compounds within 5 Å of the indicated residue.

The distribution of nucleophile-to-ligand distances implies the formation of stabilizing interactions between the nucleophile and the ligand in certain cases. For example, covalent adducts between lysine or cysteine and a ligand were identified by this analysis as a small but distinct peak at 1.2 Å and 1.8 Å in the lysine and cysteine distributions, respectively (Fig. 2-3a and b, 2-4a and b, 2-5a and b). Additionally, the prominent peak at ~2.8 Å in the lysine/ligand distance distribution likely reflects hydrogen bonding (Fig. 2-4d) or salt bridge formation (Fig. 2-4c) with the proximal ligand. We propose that such lysines are attractive targets for structure-based design because of their close proximity to the ligand. Alternatively, compounds positioned >3 Å from the lysine could more easily incorporate a larger electrophile (e.g., arylsulfonyl fluoride or benzaldehyde) to bridge the gap between the ligand and the target nucleophile (Fig. 2-4e and f). The intrinsic flexibility of the lysine side chain makes structure-based design more challenging, on the one hand, and more forgiving on the other, depending on the specific context.

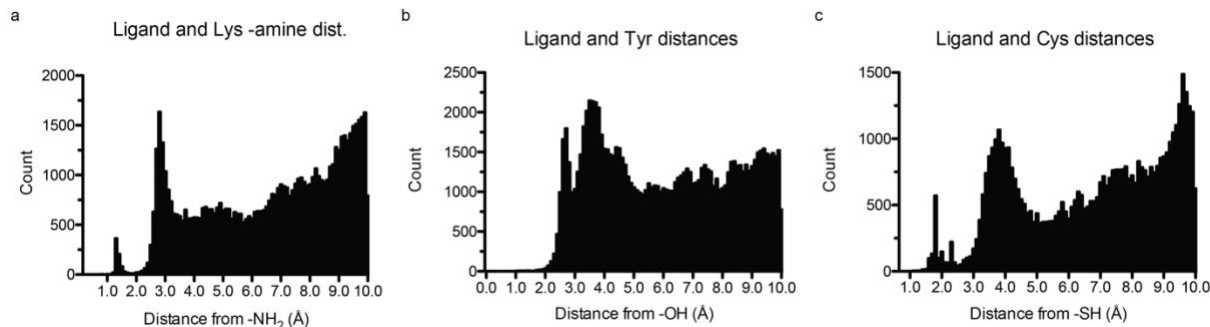


Figure 2-3: Histograms of nucleophile-to-ligand distances

Distribution of distances from a small-molecule ligand to a) the epsilon-amine of a lysines, b) the hydroxyl of tyrosines, and c) the thiol of cysteines.

Tyrosine/ligand distances exhibit two prominent peaks at ~2.8 Å and ~3.8 Å (Fig. 2-3b). The peak at 2.8 Å likely reflects hydrogen bonding between the phenolic hydroxyl and the ligand (Fig. 2-5a and b), while the peak at 3.8 Å likely reflects hydrophobic/van der Waals interactions (Fig 2-5c and d). Converting small molecules that interact with the tyrosine through hydrophobic

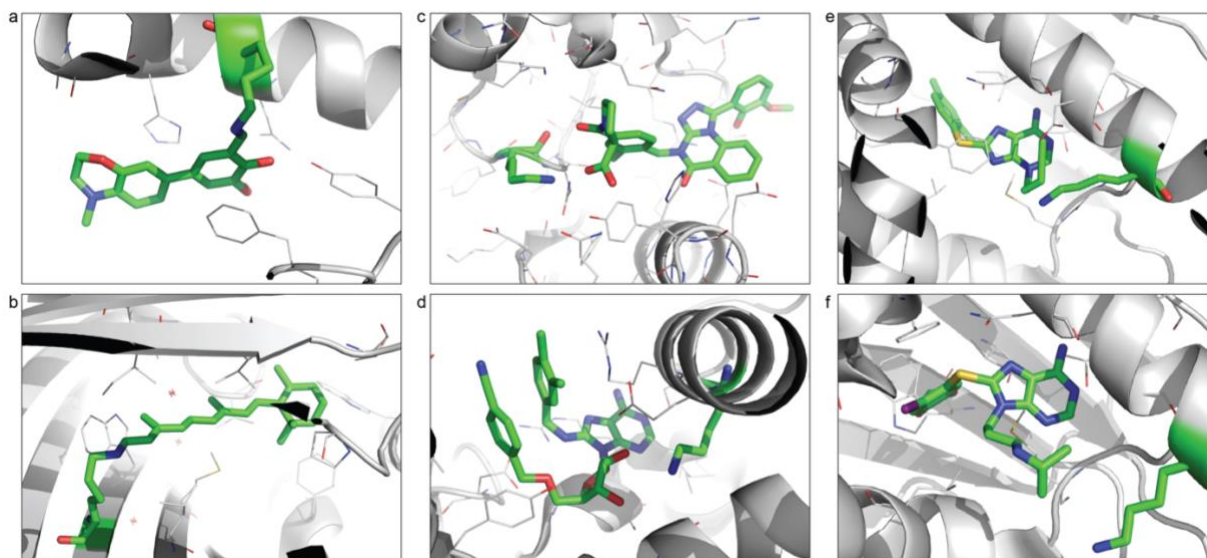


Figure 2-4: Structures of ligands and their proximal lysine residues.

a) and b) Small molecules 1.2 Å from the lysine epsilon-amine are covalently bound as in a) IRE1 bound to a reversible covalent inhibitor (PDB code: 4PL4) or b) Retinal covalently bound to a lysine in retinoic acid binding protein (PDB code: 4I9R). c) and d) Small molecules ~2.8 Å from the lysine epsilon-amine form a c) hydrogen bond as with PTPN11 Lys280 (PDB code: 6BMX) or d) salt bridge as with BIP Lys296. e) and f) Lysines ≥ 3.3 Å from the small molecule are unlikely to directly interact through polar interactions. e) Grp94, PDB code: 3O2F and f) Hsp90 α , PDB code: 2FWZ

interactions into electrophiles that react with the side-chain hydroxyl may require a significant re-design of the ligand to optimally orient the electrophile. The peak at 3.8 Å in the cysteine/ligand distance distribution also suggests the presence of hydrophobic interactions with the proximal small molecule (Fig. 2-3b). Since many of these interactions are mediated by the thiol itself, such cysteines may react more readily with electrophilic derivatives of a known ligand (Fig. 2-6c and d). Here again, it may be necessary to significantly engineer the ligand to create space to accommodate the electrophile.

From this analysis, we discuss two specific protein targets for which a covalent approach could improve inhibitor potency or selectivity. Poly ADP-ribose polymerase 14 (PARP14)

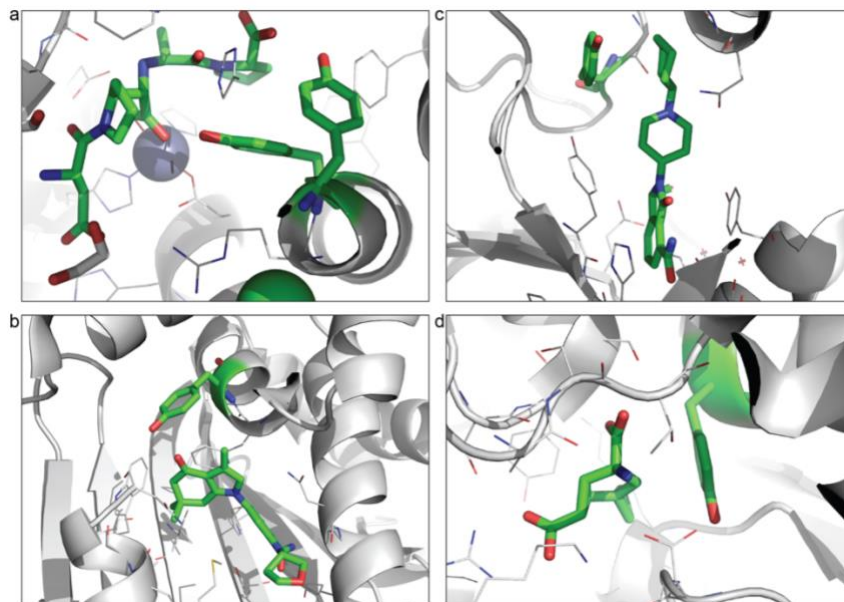


Figure 2-5: Tyrosine residues near ligands

a and b) Tyrosine residues within 2.8Å of the nearest small molecule likely form polar interactions through the phenolic hydroxyl group a) Angiotensin converting enzyme (PDB Code: 6EN5) with two tyrosine residues interacting with the bound small molecule. b) Grp94 (PDB Code: 4NH9) bound to an inhibitor and interacting with Tyr200. c and d) Tyrosine residues and small molecules within 3.8 Å may interact through hydrophobic interactions as in c) Parp1 (PDB Code 6I8M) or d) the metabotropic glutamate receptor 2 (PDB Code 4XAS).

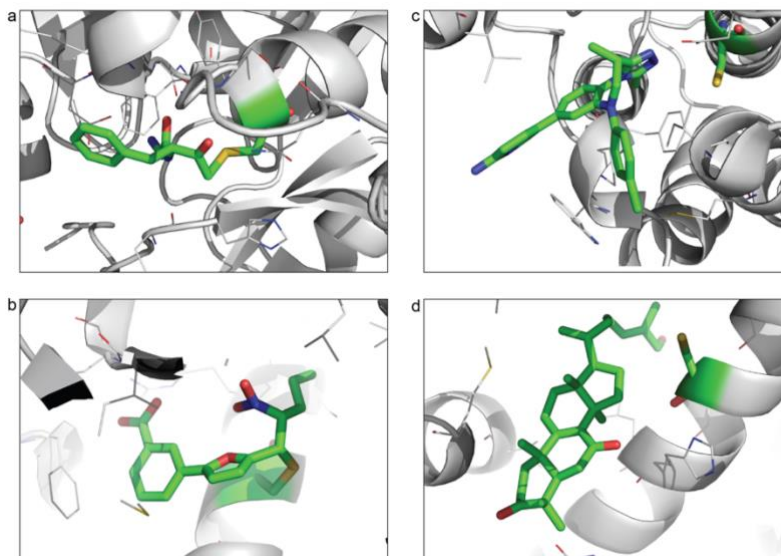


Figure 2-6: Cysteine residues near ligands

a and b) Cysteine residues ~1.8 Å from the nearest small molecule are likely covalently bound as in a) dipeptidyl peptidase 1 (PDB code: 2DJF) and b) PPARgamma (PDB code: 2ZK5). c) and d) Cysteines at a distance of ~3.8 Å as in c) BRD4 (PDB code: 4Z1S) or d) RORgamma (PDB code: 5NTN).

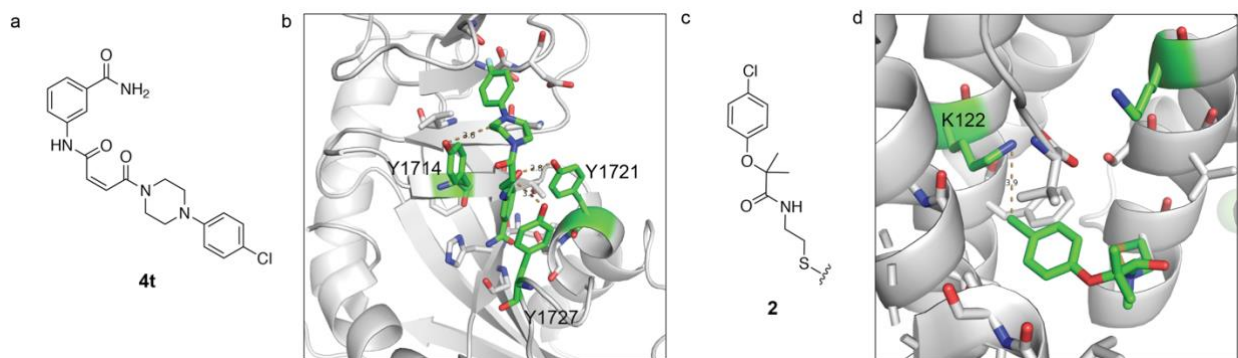


Figure 2-7: Examples of “low-hanging fruit” for the synthesis of covalent inhibitors

a) Structure of the dual PARP14/PARP1 inhibitor, **4t**. b) X-ray crystal structure of **4t** bound to PARP 14 (PDB code: 5NQE). Three tyrosine residues are within 4 Å of the small molecule. The catalytic tyrosine, Y1714, is present in all PARP family members. Y1721 and Y1727 are unique to PARP14 and could be covalently targeted to gain selectivity over PARP1. c) Structure of the disulfide tethering fragment **2**. d) X-ray crystal structure of **2** bound to the 14-3-3 σ N42C construct used in a tethering screen. The *para*-chloro group of **2** is within 4 Å of K122. Covalent modification of this residue could be a strategy to further elaborate this fragment hit.

catalyzes ADP-ribosylation and has been implicated in cancer survival pathways in both multiple myeloma and hepatocellular carcinoma.^{101,102} Compounds that selectively and potently inhibit PARP14 have applications as tool compounds to dissect its biological roles and validate PARP14 as a therapeutic target. Our analysis of the PDB identified multiple co-crystal structures of PARP14 with small-molecule inhibitors. Compound **4t** is the most potent (170 nM), but it equipotently inhibits PARP1¹⁰³ (Fig. 2-7a). Three tyrosine residues are within 5 Å of **4t** in the PARP14 crystal structure (Fig 2-7b). Tyr1714 is one of the catalytic residues and its covalent modification is unlikely to improve selectivity over PARP1 (Fig. 2-7b). However, both Tyr1721 and Tyr1727 are poorly conserved in other PARP family members,¹⁰⁴ and a covalent inhibitor that reacts with either of these tyrosine residues could be a potent and selective PARP14 inhibitor. The proximity of each tyrosine residue to the piperazine in **4t** will likely require significant compound redesign to incorporate an arylsulfonyl fluoride or aryl fluorosulfate.

The 14-3-3 family of proteins play a role in cell proliferation, cell survival and stress signaling by via protein-protein interactions at phospho-serine and phospho-threonine sites in their binding partners.^{105,106} Interrogating this rich biology with small molecule modulators of protein-protein interactions remains challenging due to a lack of potent and selective compounds. Our analysis identifies multiple potentially nucleophilic residues in the peptide binding cleft of 14-3-3 σ near small-molecule ligands, including a recently reported disulfide tethering hit that selectively stabilizes the inhibitory interaction between 14-3-3 σ (N42C mutant) and an estrogen receptor-derived phosphopeptide.¹⁰⁷ Lys122 lies within 4 Å of the *para*-chloro group in compound **2** (Fig 2-7c and d). Further elaboration of this fragment hit into a more potent covalent ligand could involve direct substitution of the *para*-chloro group with a fluorosulfate or a sulfonyl fluoride electrophile. The resulting compounds could be tested as negative modulators of the estrogen receptor for the treatment of breast cancer.

Conclusions and perspectives

This analysis of the PDB suggests that covalent targeting of cysteine residues provides access to ~20% of the proteome, at best. Other nucleophilic residues, such as tyrosine and lysine, are more difficult to target covalently due to their decreased nucleophilicity relative to cysteine. However, their prevalence in pockets and in small-molecule binding sites has the potential to increase the fraction of the proteome that can be targeted with covalent inhibitors. Recent proteome-wide efforts to identify ligandable lysines⁴¹ provide additional complementary evidence supporting the notion that a larger swath of the proteome can be targeted using covalent inhibitors. The work presented here motivates further study of proteome ligandability using an increased toolkit of electrophiles attached to noncovalent recognition scaffolds. Our work on one of the ligands/pockets shown in Fig. 2-4f -- PU-H71 bound to Hsp90 -- demonstrates that our PDB analysis can provide an excellent starting point for the design of a

lysine-targeted covalent inhibitor, even though the lysine is poorly reactive, solvated, and surface exposed (see Chapter 3).

Methods

General:

All scripts were written in Python 2.7.10. Venn diagrams were generated using Venny 2.0.¹⁰⁸

Bar graphs were created using Graphpad Prism 5.

Analysis of representative pockets

Supplementary Table 2 from Bhagavat et al. was used to calculate the frequency of each amino acid across all 6,761 pockets listed in their dataset.⁹⁶ To calculate the number of times cysteine, lysine or tyrosine occurred within the 6,761 pockets, the dataset was converted to a CSV file and used as an input for the script in Appendix B (AA_per_pocket.py). This script counts the number of times every single amino acid occurs within a pocket, and outputs the result as a CSV file. The results were then plotted in Prism.

PDB analysis

All X-ray structures (resolution ≤ 3 Å) containing a protein from a metazoan organism and a bound ligand were retrieved in PDB format from the Protein Data Bank using the RCSB download tool. The PDB files were processed using the python script in Appendix B (has_good_K_v3.3_CSV_Output.py). From each PDB file, the coordinates and the abbreviation for all atom entries labeled "HETATM" were retrieved. A total of 574 "ligands" corresponding to buffers, precipitants, ions, detergents and biological cofactors were excluded from this list. Additionally, the coordinates for all of the nucleophilic atoms of interest ("NZ" for lysine, "OH" for tyrosine, and "SG" for cysteine) were also retrieved and stored. The distance between each

ligand atom and each nucleophilic atom was calculated from the stored coordinates using the following equation:

$$distance = \sqrt{(x_{SM} - x_{nuc})^2 + (y_{SM} - y_{nuc})^2 + (z_{SM} - z_{nuc})^2}$$

The shortest distance between each ligand and the nucleophilic atom was stored. The list of all compound and nucleophile pairs meeting these criteria was exported as a CSV file. To correct errors in residue numbering in the PDB, the residues listed were then cross-referenced with the Uniprot entry using `pdbsws_res.txt`.¹⁰⁹

Chapter 3: Ligand pre-organization drives enantioselective covalent targeting of a surface-exposed lysine

Abstract:

Covalent modification of surface-exposed lysine residues is challenging due to the low reactivity of the epsilon-amino group, which is protonated and solvated at physiological pH. Optimization of covalent inhibitors generally focuses on improving reversible affinity (K_i), rather than increasing the intrinsic rate of covalent bond formation (k_{inact}). To date, the primary approach for optimizing k_{inact} has relied on increasing the intrinsic reactivity of the electrophile, which raises concerns due to increased off-target modification. These concerns are especially pronounced for lysine-targeted inhibitors due to the high prevalence of lysine residues in the proteome. In this chapter, we report the development of arylsulfonyl fluoride probes that react with a surface-exposed lysine in Hsp90. Through multiple high-resolution crystal structures of covalent and non-covalent ligand/Hsp90 complexes, we show that a chiral, conformationally constrained linker orients the sulfonyl fluoride toward the lysine, dramatically increasing k_{inact} without affecting the intrinsic reactivity or reversible binding affinity. We further demonstrate enantioselective covalent modification of cellular Hsp90, which results in a rapid heat shock response and concomitant degradation of the covalent ligand/Hsp90 complex.

Introduction

Covalent inhibitors have unique properties that make them especially useful as drugs and as research tools. Irreversible covalent modification of a protein results in drug-target residence times that match the lifetime of the target protein,^{3,4,110} and this is often independent of the drug clearance rate. Additionally, covalent inhibitors can discriminate between closely related paralogs by reacting with a poorly conserved nucleophilic residue near the ligand binding site.^{17,19,58,62,63,95} Selectivity for the intended residue is achieved by virtue of a two-step reaction mechanism in which reversible binding to the target precedes covalent modification. Both reversible affinity (K_i) and the rate of covalent bond formation (k_{inact}) within the initially formed noncovalent complex contribute to overall covalent inhibitor potency.¹¹¹ The most

common strategy for optimizing k_{inact} is to increase the intrinsic reactivity of the electrophile. However, this approach is suboptimal, as it increases the likelihood of off-target reactivity in a complex cellular environment. As a result, optimization of covalent inhibitors primarily relies on improving K_i , i.e., increasing the reversible affinity of the noncovalent recognition element.^{37,112} Because of these inherent challenges, the development of highly selective covalent inhibitors has focused primarily on proteins with a cysteine residue proximal to a ligand binding site. By contrast, there are fewer examples of selective covalent probes that react with non-cysteine nucleophiles.^{25,32,43,47,48,79,100}

Lysine is more prevalent than cysteine and is therefore an attractive target for covalent probe design. However, rapid and selective covalent modification of lysine is extremely challenging due to its physicochemical properties. Since the pKa of the lysine epsilon amine in solution is ~10.6, the amine is predominantly protonated and unreactive at physiological pH. Additionally, the four rotatable bonds in the lysine side chain allow for dozens to hundreds of low-energy conformers, therefore complicating the design of fast-acting, lysine-targeted inhibitors with high k_{inact} . Lastly, the high prevalence of lysine residues in the proteome increases the likelihood of off-target reactions. Overall, these attributes make it difficult to identify a lysine-targeted electrophile with the requisite balance of reactivity and stability. Arylsulfonyl fluorides overcome some of these challenges and have been incorporated into chemoproteomic probes for kinases and stabilizers of mutant transthyretin.^{22,24} However, lysine residues in these targets are likely pKa perturbed and therefore readily react with arylsulfonyl fluorides.^{22,113} There are few examples of arylsulfonyl fluorides that react with surface-exposed lysine residues, which are fully solvated, mostly protonated, flexible, and hence lack intrinsic reactivity.²³

In this chapter, we report the development of arylsulfonyl fluoride-based covalent inhibitors of Hsp90. These probes react with a surface-exposed lysine residue adjacent to the ATP-binding pocket. Using high-resolution x-ray crystal structures and kinetic characterization,

we show that ligand and linker conformational pre-organization leads to dramatically accelerated covalent modification of a surface-exposed lysine in Hsp90.

Results

We used the crystal structure of the Hsp90 N-terminal domain (NTD) bound to the ATP-competitive inhibitor, PU-H71,¹¹⁴ as a starting point to develop electrophilic probes that target Lys58, a surface-exposed lysine that lies outside the ATP pocket. The inhibitors consist of a modular design that includes (1) a non-covalent recognition scaffold based on the purine thioether of PU-H71, (2) a phenylsulfonyl fluoride electrophile, and (3) a variable linker to orient the phenylsulfonyl fluoride in close proximity to Lys58. The secondary amine of PU-H71 sits 6.2 Å from the epsilon-amine of Lys58 (Fig. 3-1a). Based on this distance and the inherent flexibility of both the propylene linker in PU-H71 and the side chain of Lys58, we reasoned that a phenylsulfonyl fluoride would be tolerated as a replacement for the isopropyl group, leading to the design and synthesis of sulfonyl fluoride **1** (Fig. 3-1b). Treatment of recombinant Hsp90 NTD with a saturating concentration of sulfonyl fluoride **1** (10 µM) resulted in covalent adduct formation as measured by LC-MS (Fig. 3-1c). Based on our kinetic analysis (K_i and k_{inact} determination, Fig. 3-1d and 3-2), we found that the reversible binding affinity of **1** ($K_i \sim 60$ nM) is comparable to the published dissociation constant (K_d) of PU-H71.¹¹⁵ Mutating Lys58 to an arginine abrogated covalent modification, despite the presence of 16 other lysine residues in the Hsp90 NTD (Fig 3-1c). To elucidate the binding pose, we determined the co-crystal structure of covalently modified Hsp90 NTD at a resolution of 1.6 Å. The purine thioether of **1** overlays perfectly with the previously determined co-crystal structure of PU-H71 bound to the Hsp90 NTD. In contrast to the noncovalent PU-H71 structure, the propylamine linker of **1** adopts a fully extended conformation, whereas the conformation of the Lys58 side chain is slightly kinked. Finally, continuous electron density between the epsilon-amine of Lys58 and the sulfonyl group of **1** provides unambiguous evidence for covalent bond formation (Fig 3-1e). We note that the

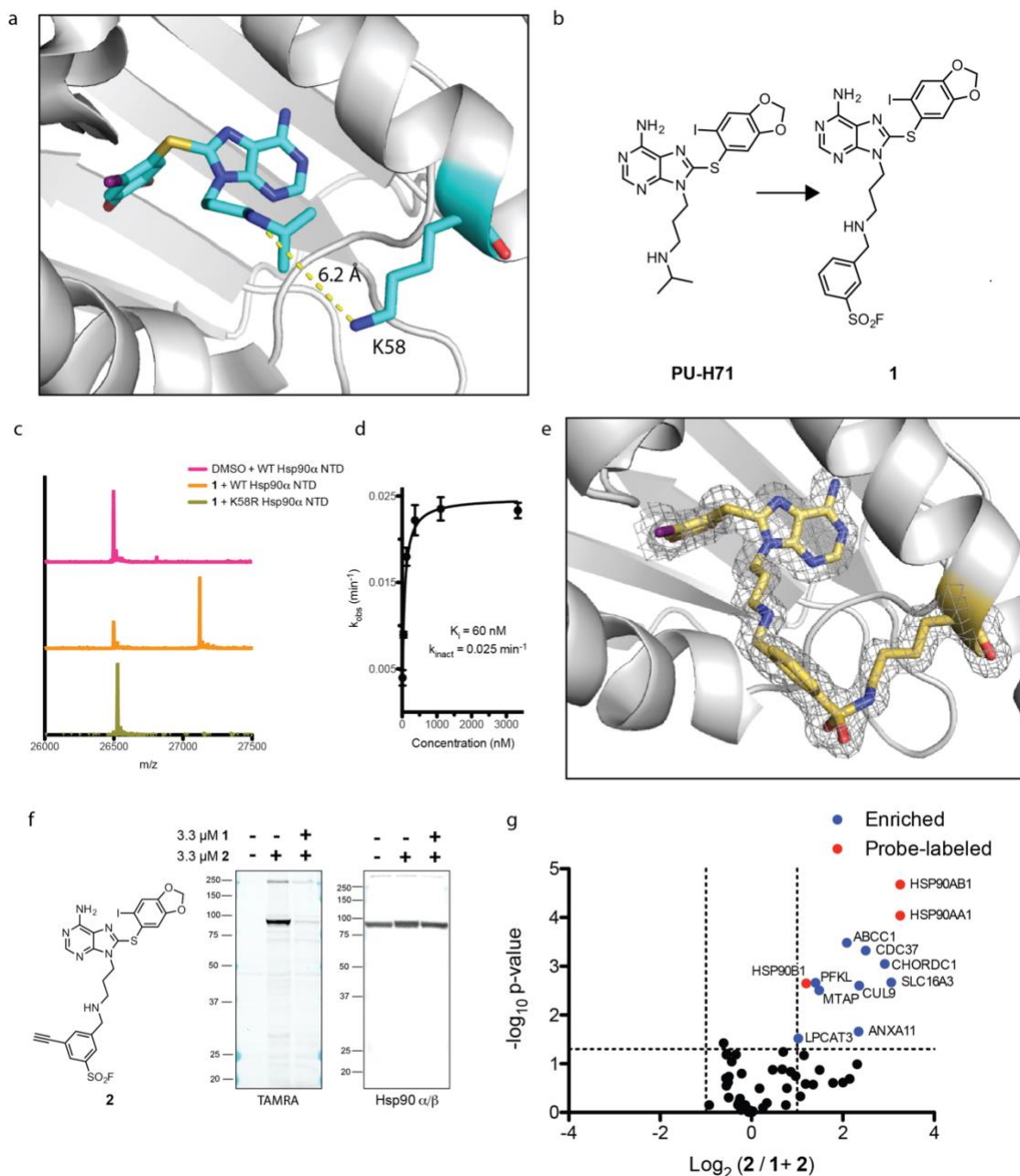


Figure 3-1: Characterization of sulfonyl fluoride-based covalent inhibitors of Hsp90.

a) The K58 ϵ -amine is within 6.2 Å of the PU-H71 secondary amine (PDB Code: 2FWZ). b) Compound **1** is a benzylsulfonyl fluoride-containing analog of PU-H71. c) **1** (10 μ M) forms a covalent adduct with the Hsp90 α NTD (1 μ M) after 1-h incubation at 37°C. No reaction is observed between **1** and Hsp90 α NTD K58R. d) Kinetic analysis of the reaction between **1** and Hsp90 α reveals a K_i of 60 nM, and k_{inact} of 0.025 min^{-1} . e) 1.6-Å resolution crystal structure of Hsp90 α -**1** adduct with overlaid $2F_o - F_c$ map shows covalent modification of K58. f) Occupancy probe **2** strongly labels a 90-kDa band in Skbr3 cells as shown by in-gel fluorescence. g) Streptavidin pulldown and MS identification of proteins labeled by **2** in Skbr3 cells treated with DMSO or **1**, followed by cell lysis and biotin-azide conjugation. Probe-labeled peptides were only identified for Hsp90 α and Hsp90 β /Grp94.

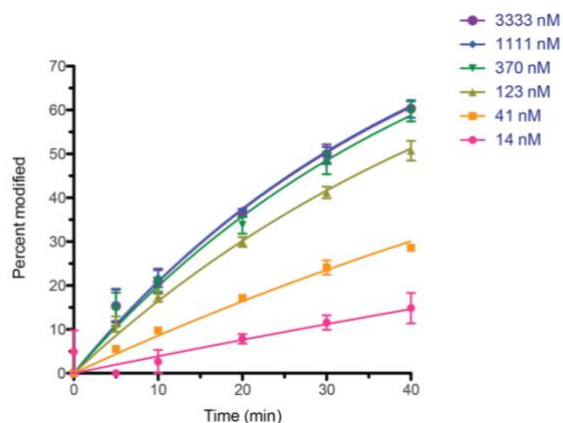


Figure 3-2: Concentration-dependent k_{obs} for the modification of Hsp90 by **1**

The concentration-dependent observed rate of modification of the Hsp90 α NTD by **1** was determined by LC-MS. Data were fit to a single-phase exponential association function to extract k_{obs} .

Hamachi lab has recently developed acyl sulfonamide probes that covalently modify Hsp90 on Lys58³⁶; the binding mode of these probes is unknown, however.

We next characterized the selectivity of **1** in living cells. We synthesized alkyne **2** as a clickable version of **1** suitable for cellular target engagement experiments. Skbr3 cells were treated with **2** (3.3 μ M, 90 min) and probe-labeled proteins were visualized by in-gel fluorescence after copper-catalyzed conjugation of TAMRA-azide (TAMRA = tetramethylrhodamine). This revealed a 90-kDa band as the major target. Labeling was strongly reduced by pretreating cells with an equivalent concentration of **1** (Fig 3-1f). To identify the cellular protein targets and modification sites of probe **2** in an unbiased manner, we used a hydrazine-cleavable biotin-azide for the click conjugation step, followed by streptavidin enrichment and on-bead trypsinization. Mass spectrometry analysis of the on-bead tryptic digest revealed the cytosolic Hsp90 paralogs, Hsp90 α and Hsp90 β , as the proteins most strongly competed by **1** (Fig. 3-1g). Known Hsp90-interacting proteins, such as the critical co-chaperone Cdc37, were also identified in this fraction.

Hydrazine-mediated cleavage of the immobilized biotin linker, followed by mass spectrometry analysis of the eluate, revealed a unique probe-modified peptide corresponding to

Hsp90 α (aa 47-60), as well as a second probe-modified peptide corresponding to Hsp90 β (aa 42-55) and/or Grp94 (aa 103-116) (Fig. 3-3). The MS2 spectra provide unambiguous evidence that both peptides were modified at the intended lysine residue (Lys58 based on Hsp90 α numbering). No other probe-labeled peptides were identified. Because Hsp90 β was competed much more strongly than Grp94 (based on label-free quantitation of the on-bead tryptic fractions, Fig. 3-1g), it is likely that sulfonyl fluoride **1** potently engages both Hsp90 α and Hsp90 β in cells, whereas it weakly engages Grp94. Overall, these findings demonstrate significant target engagement and selective modification of Hsp90 at Lys58 in living cells.

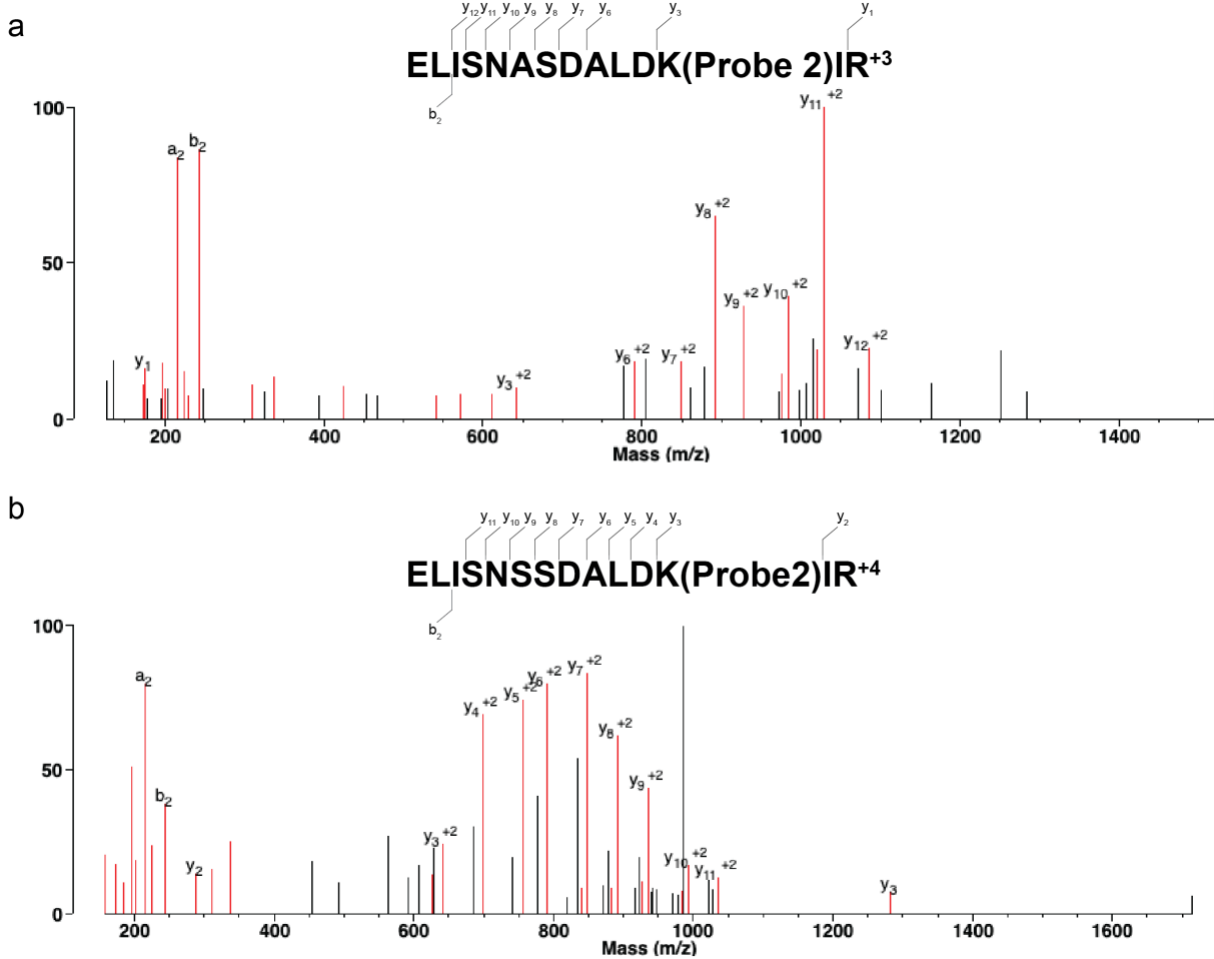


Figure 3-3 MS/MS fragmentation spectra of probe 2-labeled peptides.

a) Labeled peptide derived from Hsp90 β /Grp94, and b) Labeled peptide derived from Hsp90 α .

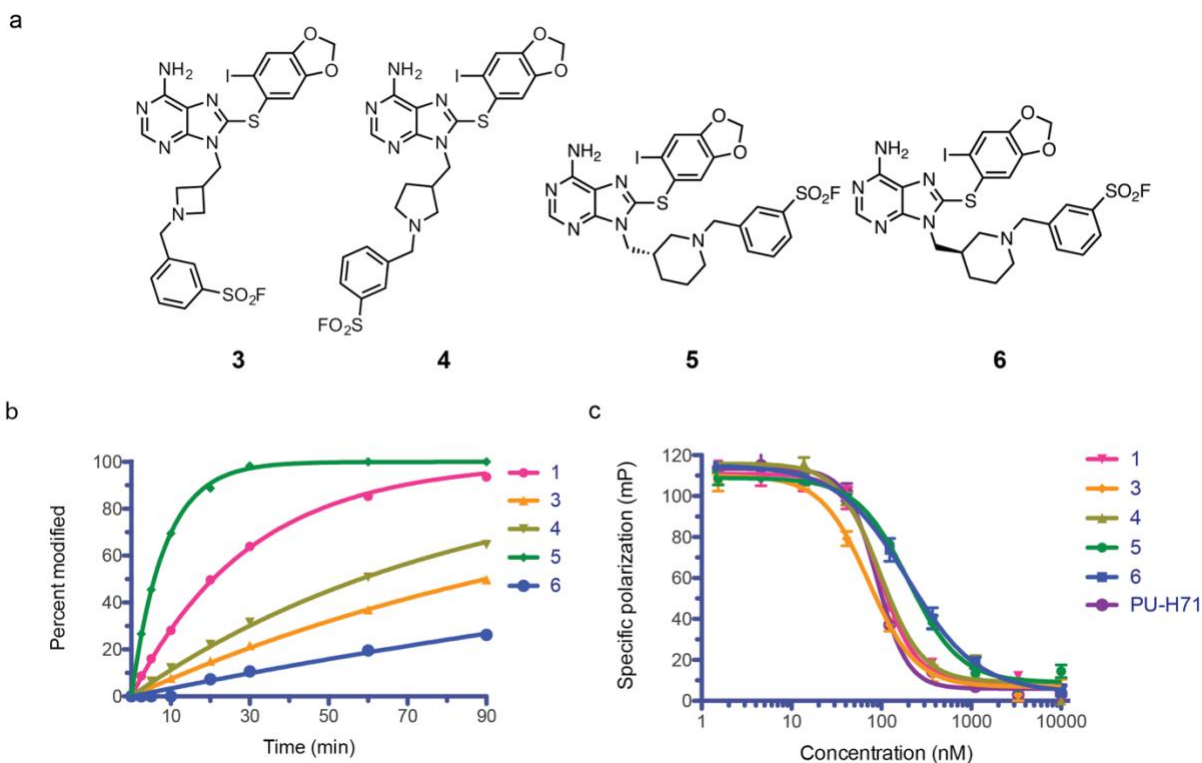


Figure 3-4: Constrained linkers result in enantioselective Hsp90 modification

a) Chemical structures of **3-6**. b) Distinct linkers in **1** and **3-6** ($10 \mu\text{M}$) significantly affect the rate of Hsp90 α NTD ($1 \mu\text{M}$) modification. Enantiomers **5** and **6** exhibit strikingly different rates of modification. c) The dissociation constant (K_d) of **1**, **3-6** and PU-H71 for K58R Hsp90 α NTD (30 nM) was measured by fluorescence polarization using the fluorescent trace, FITC-geldanamycin (5 nM). Enantiomers **5** and **6** bind Hsp90 with identical affinity, despite having drastically different reactivity with Hsp90.

Given that the maximum rate of Hsp90 modification by **1** corresponds to a $t_{1/2}$ of $\sim 30 \text{ min}$ (Fig. 3-1d), a key question concerns whether the arylsulfonyl fluoride is optimally oriented to facilitate efficient nucleophilic substitution by K58 in the context of an initially formed noncovalent complex. We tested the effect of constraining the propylamine linker within a 4-, 5-, or 6-membered ring to give azetidine **3**, pyrrolidine **4** (tested as a racemic mixture) and the enantiomerically pure piperidines **5** and **6** (Fig 3-4a). We measured the modification rates at a compound concentration of $10 \mu\text{M}$, which is 250-fold higher than the K_d of PU-H71 for Hsp90.¹¹⁵ Assuming the varied linkers do not perturb the reversible binding affinities, this simple measurement should reveal the relative rates of covalent bond formation between K58 and non-

covalently bound **1** and **3-6** (i.e., relative k_{inact} values). Strikingly, sulfonyl fluoride **5**, which has the (S)-methylpiperidine linker, exhibited the fastest modification rate ($t_{1/2} \sim 10$ min), while its enantiomer **6** was the slowest compound tested ($t_{1/2} > 90$ min) (Fig 3-4b).

To confirm that the observed modification kinetics truly reflect differences in k_{inact} and not the intrinsic binding affinity (K_i), we measured the equilibrium binding affinities (K_d) of **1**, and **3-6** using the K58R mutant of Hsp90 NTD, which lacks the targeted lysine. Using a well established fluorescence polarization assay with FITC-geldanamycin as the competitive fluorescent ligand¹¹⁶ (Fig 3-4c), we found that compounds **1** and **3-6** bind Hsp90 with similar affinities (K_d 100-200 nM). These data strongly suggest that the observed modification rates at 10 μM reflect differences in k_{inact} . To confirm the true value of K_i and k_{inact} for compound **5**, we measured the concentration and time-dependent modification of Hsp90 by **5**, and calculated a K_i of ~ 160 nM and a k_{inact} of 0.13 min^{-1} (Fig. 3-5), both of which are in good agreement with the fluorescence polarization and single timepoint measurement of reactivity (Fig. 3-4). Given that the intrinsic reactivities of sulfonyl fluorides **5** and **6** are identical toward (achiral) amines, we hypothesized

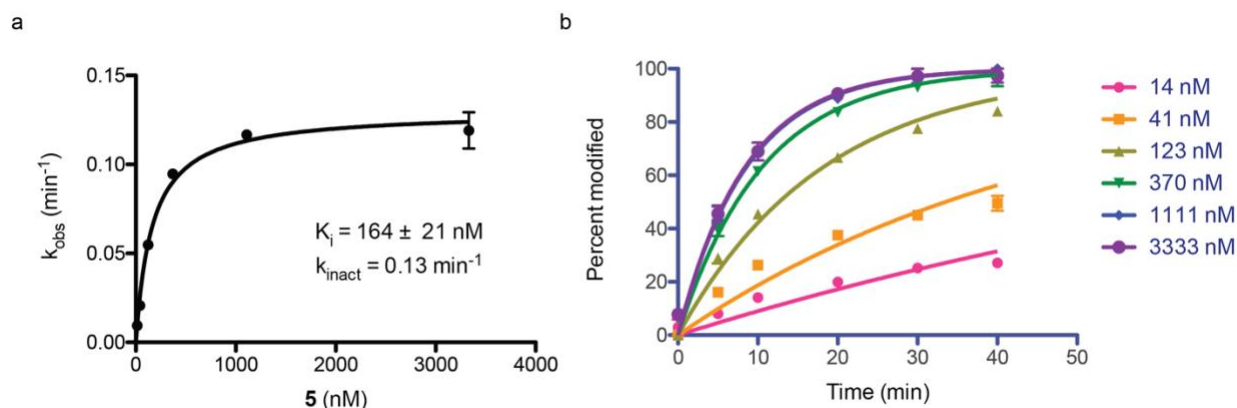


Figure 3-5: Kinetic analysis of compound 5

a) The observed rate of Hsp90 modification by **5** as a function of concentration was used to calculate the K_i (164 nM) and k_{inact} (0.13 min^{-1}) for the reaction. d) Time and concentration-dependent modification of Hsp90 by **5**. Data from each concentration were fit to a single exponential to extract k_{obs} .

that the (*S*)-methylpiperidine linker of **5**, but not the (*R*)-methylpiperidine linker of **6**, specifically orients the phenylsulfonyl fluoride electrophile toward Lys58 within the chiral environment of the Hsp90 binding site.

We used protein crystallography to provide structural insights into the dramatic difference in Hsp90 modification rates exhibited by **5** and **6**. The structure of **5** bound to the Hsp90 α NTD, determined at 1.8-Å resolution, demonstrates continuous electron density between the sulfonyl group and K58, consistent with covalent bond formation (Fig. 3-6a). We

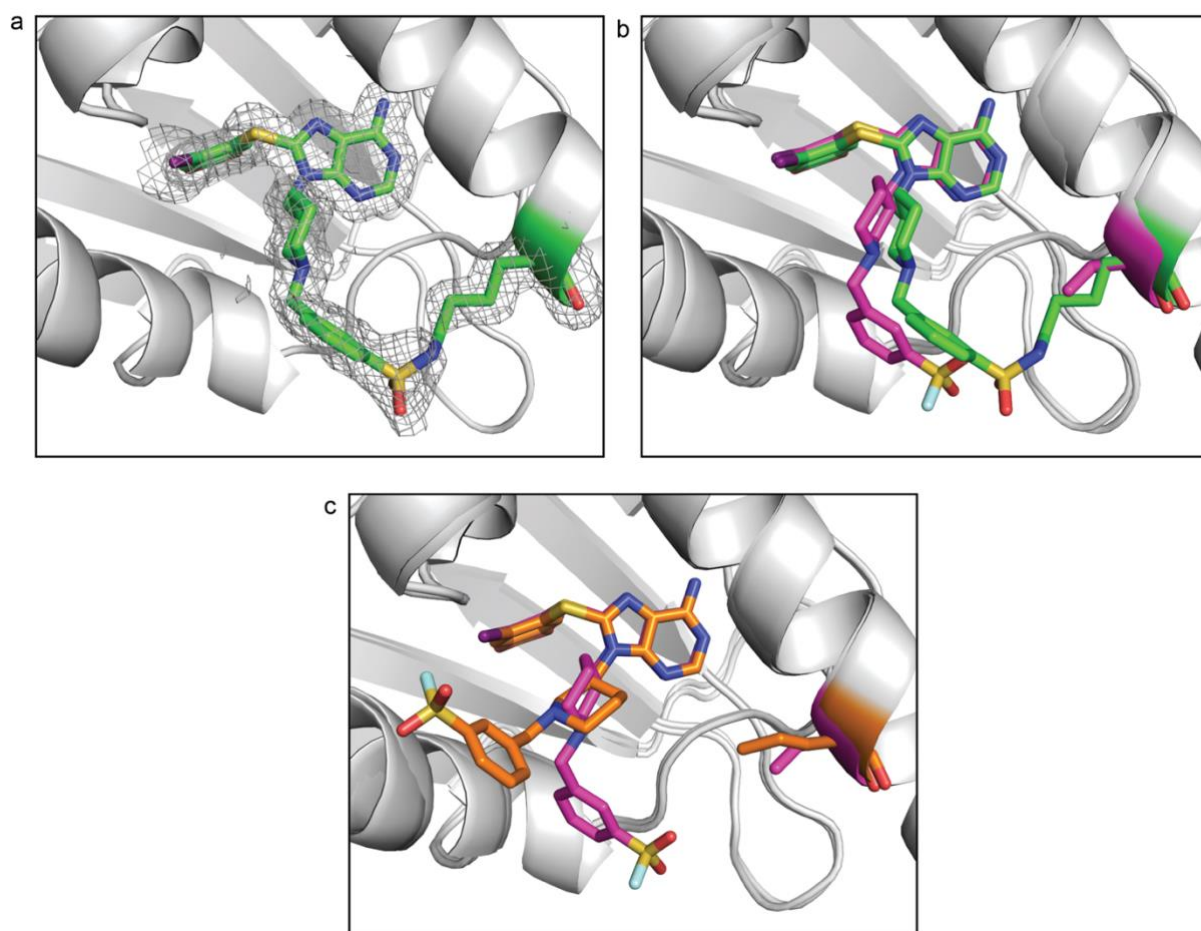


Figure 3-6: Crystal structure of **5 and **6** bound to Hsp90 α**

a) 1.8-Å crystal structure of **5** covalently bound to Hsp90 α K58 with overlay of $2F_o - F_c$ map. b) Overlay of structure in a) with **5** reversibly bound to Hsp90 α K58R NTD. **5** adopts a conformation poised to react with K58 upon binding to the protein. C) Overlay of crystal structures of **5** and **6** reversibly bound to Hsp90 α K58R NTD. The strikingly different orientation of the sulfonyl fluoride in **5** and **6** accounts for the observed differences in reactivity.

additionally solved the structure of the noncovalent complex of **5** bound to the Hsp90 K58R mutant, which adopts a pose nearly identical to the covalently bound product (Fig. 3-6b and 3-7). Strikingly, the linker in **6** adopts a radically different orientation upon binding to Hsp90, such that the arylsulfonyl fluoride substituent projects in the opposite direction from K58 (Fig. 3-6c and 3-7). We conclude that, upon binding to Hsp90, the (*S*)-methylpiperidine linker **5** adopts a low-energy conformation which places the sulfonyl fluoride within striking distance of the epsilon-amine of Lys58. By contrast, the (*R*)-methylpiperidine linker and attached arylsulfonyl fluoride of **6** adopt a reaction-competent conformation less frequently.

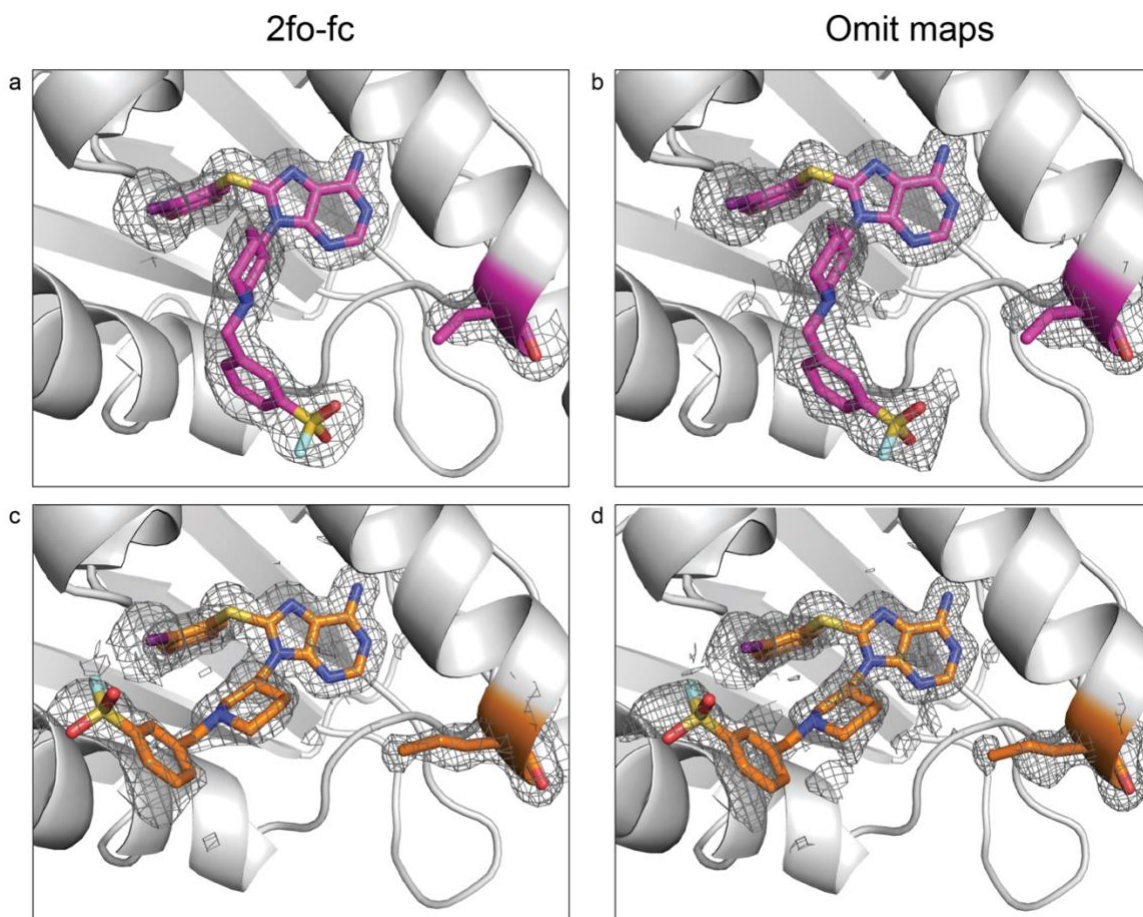


Figure 3-7: Electron density maps and X-ray structures of Hsp90 α K58R NTD bound to **5 and **6****

a) 2Fo-Fc map of **5** bound to Hsp90 α K58R NTD. b) 2Fo-Fc omit map of **5** bound to Hsp90 α K58R NTD. c) 2Fo-Fc map of **6** bound to Hsp90 α K58R NTD. d) 2Fo-Fc omit map of **6** bound to Hsp90 α K58R NTD.

We next compared the effects of **5** and **6** in Skbr3 cells. Compound **5** exhibited dose-dependent ($IC_{50} \sim 400$ nM) occupancy of endogenous cellular Hsp90, as revealed by reduced Hsp90 labeling by the competitive clickable probe **2**. We inferred from this experiment that cellular Hsp90 was covalently modified by **5**, because (1) clickable competitor **2** was added in cell lysates after removing excess **5**, and (2) the enantiomeric sulfonyl fluoride **6** failed to prevent labeling by probe **2** under these conditions (Fig. 3-8a). As expected for a covalent

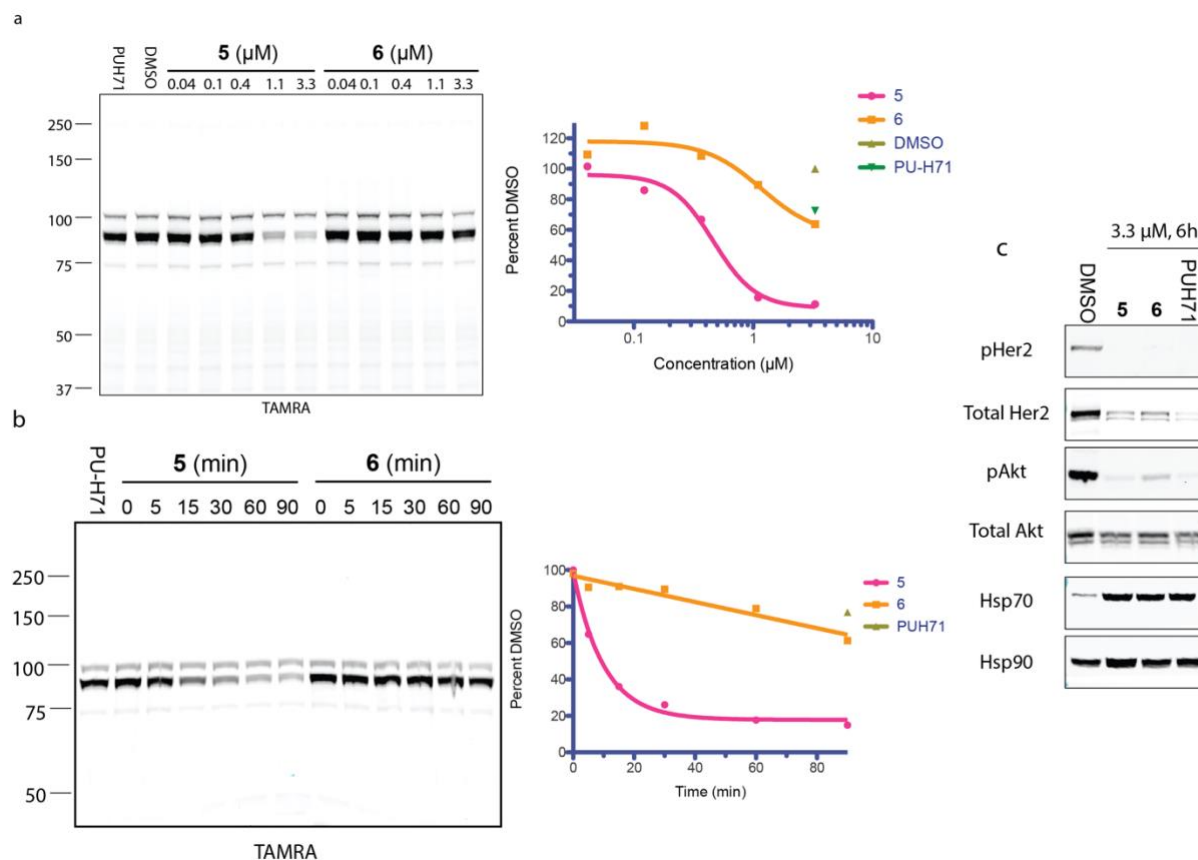


Figure 3-8: Optimized compound 5 engages Hsp90 in vivo in Skbr3 cells.

a) Dose-dependent occupancy of Hsp90 by **5** and **6** (90 min) was measured in Skbr3 cells. Cell lysates were treated with **2** (3.3 μ M, 60 min, 37°C) to measure occupancy by in-gel fluorescence. b) Time-dependent occupancy of Hsp90 by **5** and **6** (3.3 μ M) was measured in Skbr3 cells. Cell lysates were treated with **2** (3.3 μ M, 60 min, 37°C) to measure occupancy by in-gel fluorescence. c). Continuous treatment with **5**, **6**, and PU-H71 results in degradation of the Hsp90 client Her2, loss of Akt S473 phosphorylation, and activation of the heat shock response as measured by Hsp70 expression.

inhibitor, the extent of Hsp90 occupancy by **5** increased over time (Fig. 3-8b). Remarkably, the rate of endogenous Hsp90 occupancy by 3.3 μ M **5** in cells ($t_{1/2}$ ~10 min) was the same as the maximum rate of covalent modification (k_{inact}) observed with purified Hsp90 NTD. Continuous treatment of Skbr3 cells with compounds **5**, **6** and PU-H71 (3.3 μ M, 6h) activated the heat shock response (as shown by increased Hsp70) and downregulated known Hsp90 clients including Her2 and Akt (Fig. 3-8c). This is consistent with the fluorescence polarization data showing equipotent reversible binding of both enantiomers to Hsp90.

To explore the cellular phenotypes resulting from pulsed, covalent inhibition of Hsp90, we compared the effects of treating cells with a saturating concentration of compound **5**, **6**, and PU-H71 (3.3 μ M, 60 min), followed by rigorous compound washout. Under these conditions, compound **5** exhibited ~100% occupancy immediately after washout (Fig. 3-9a). By contrast, treatment of cells with **6** and PU-H71 did not result in Hsp90 occupancy, demonstrating washout of the reversibly-bound inhibitors. Hsp90 occupancy by **5**, quantified by competitive labeling with probe **2** (added to cell lysates), decreased in a time-dependent manner ($t_{1/2}$ ~12h), presumably due to degradation of probe-modified Hsp90 (see below) and re-synthesis of unoccupied Hsp90. Consistent with the occupancy results, we observed a sustained increase of Hsp70 in cells pulsed with **5**, but not in cells pulsed with **6** or PU-H71 (Fig. 3-9a). Qualitatively similar effects were observed with respect to downregulation of Her2 and Akt (Fig. 3-9b).

The time-dependent decrease in Hsp90 occupancy by **5** (Fig 3-9) could be due to resynthesis of uninhibited Hsp90, degradation of the inhibited Hsp90 or a combination of the two. To better understand the fate of irreversibly inhibited Hsp90, we treated intact cells with clickable probe **2** (3.3 μ M 90 min) and directly monitored the covalent Hsp90-**2** adduct over time by in-gel fluorescence after compound washout. The fluorescent signal from the covalent Hsp90-**2** adduct decreased steadily over 24 hours, with a half-life of 6-12 hours (Fig. 3-10a and c). Total Hsp90 levels did not change over this time period, likely due to resynthesis of Hsp90

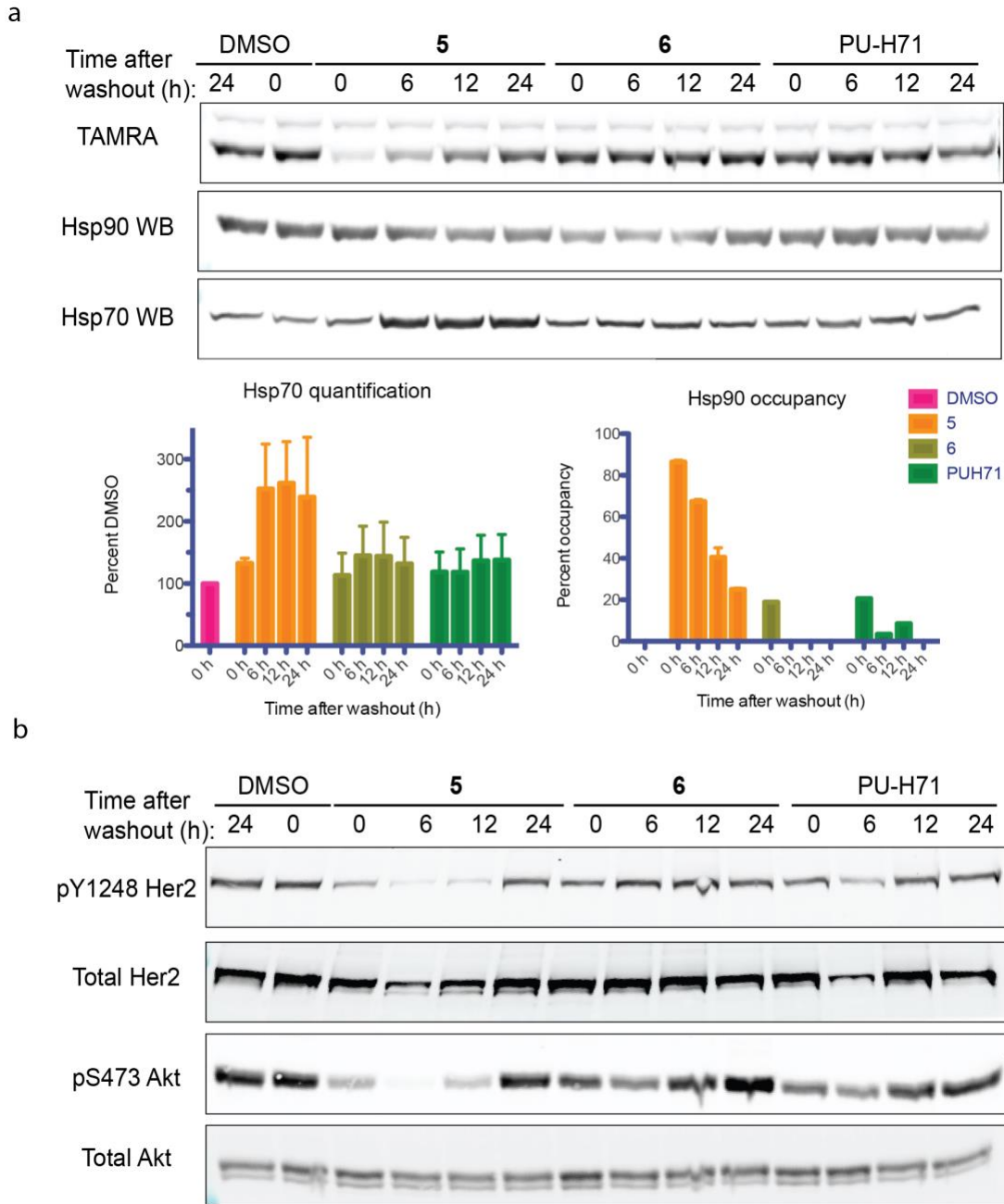
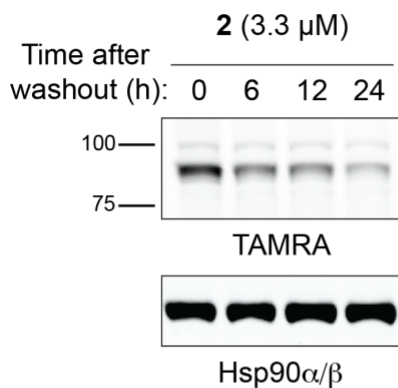


Figure 3-9: The effects of covalent Hsp90 modification after compound washout

Skbr3 cells were treated with **5**, **6**, or PU-H71 (3.3 μ M, 90 min) followed by compound washout. a) Only cells pulsed with **5** demonstrate compound occupancy after washout, as measured by in-gel fluorescence after treating cell lysates with probe **2**. Despite a decrease in Hsp90 occupancy over time as measured by in-gel fluorescence, Hsp70 levels remain elevated over the course of the experiment in cells pulsed with **5**. b) Cells pulsed with **5** decreased Her2 autophosphorylation and Akt phosphorylation, consistent with covalent modification of Hsp90.

concomitant with degradation of the covalent Hsp90-2 adduct. Strikingly, the observed degradation of the covalent Hsp90-2 adduct is much more rapid than constitutive degradation of unmodified Hsp90. In fact, we did not observe any Hsp90 turnover when new protein synthesis was inhibited by treating cells with 50 $\mu\text{g}/\text{mL}$ cycloheximide (Fig 3-10b and c). Small-molecule inhibition of Hsp90 is known to stabilize its interactions with certain clients in a biochemical reconstitution system.¹¹⁷ In cells, E3 ubiquitin ligases have been shown to associate with inhibited Hsp90/client complexes,¹¹⁸ presumably resulting in client ubiquitination and degradation. However, the mechanism by which this occurs remains unknown. Our results suggest that covalently inhibited Hsp90 is itself degraded, perhaps in the context of a terminally inactivated complex with its clients.

a



b

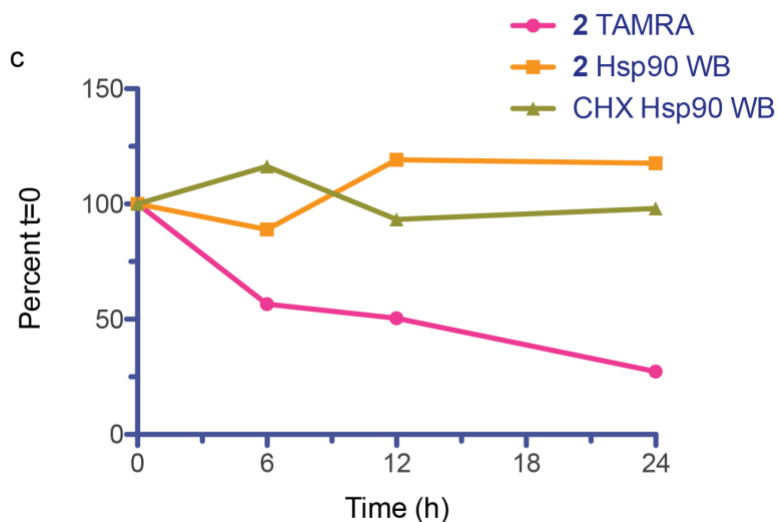
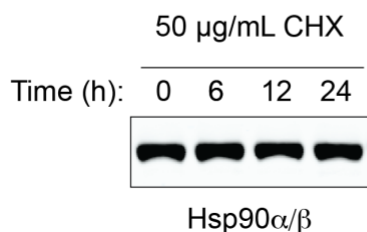


Figure 3-10: Direct measurement of inhibited Hsp90 turnover

a) Skbr3 cells were treated with 2 (3.3 μM , 90 mins) followed by compound washout. The abundance of probe-labeled Hsp90 was visualized by in-gel fluorescence after click conjugation of TAMRA-azide in cell lysates. Total Hsp90 was unchanged by western blot. b) Turnover of Hsp90 was measured in the presence of 50 $\mu\text{g}/\text{mL}$ cycloheximide over 24h. c) Quantification of the in-gel fluorescence and western blots in a and b. Probe-labeled Hsp90 exhibits faster degradation ($t_{1/2}$ ~6-12 h) than uninhibited Hsp90 (no turnover observed).

The synthesis of selective, lysine-reactive probes is challenging due to the epsilon-amine's low intrinsic reactivity and its high prevalence across the proteome. Here, we demonstrate that an aryl sulfonyl fluoride can efficiently react with a surface-exposed lysine residue. Moreover, we show that conformational constraints alone can have a dramatic effect on k_{inact} . Because the introduction of conformational constraints is expected to have little or no effect on the intrinsic reactivity of the sulfonyl fluoride, this strategy of maximizing k_{inact} is especially attractive from the standpoint of minimizing covalent modification with off-target nucleophiles. We anticipate that this approach will prove to be generally useful for the structure-based design of covalent inhibitors that target poorly reactive, solvent-exposed lysine residues.

Methods:

Protein Expression and Purification

The DNA sequence coding for amino acids 2-236 of human Hsp90aa1 was cloned from a full length Hsp90aa1 expression vector (gift from David Agard) into the pET28 plasmid containing a 6X His tag and a PreScission protease cleavage site. From this plasmid, the Hsp90aa1 K58R NTD was then prepared by site directed mutagenesis. Each plasmid was transformed into BL21(DE3) RIL *E. coli* and 6 L of bacterial culture in LB-kanamycin were grown for bacterial protein expression. Cultures were induced with 200 μ M IPTG at 18°C at an OD600 of 1.2 and cultures were grown overnight at 18°C while shaking at 200 RPM. Bacteria from each flask was pelleted (3000 x g, 30 min, 4°C) and then resuspended in lysis buffer (50 mM potassium phosphate buffer, pH 8, 500 mM KCl, 6 mM beta-mercaptoethanol, 10 mM imidazole, pH 8, 10% glycerol, 2 mM PMSF, 1X cOmplete protease inhibitors). Cells were lysed using an Emulciflex C3 microfluidizer, and lysates were cleared by centrifugation (35,000 x g, 30 mins, 4°C). Batch binding of the His-tagged protein was performed for 1h at 4°C using 1 mL of Qiagen Ni-NTA agarose resin per liter of culture. Beads were then transferred to a column in lysis buffer and washed with 20 column volumes of high salt buffer (30 mM HEPES, pH 8, 500 mM KCl, 6 mM BME), 20 columns volumes of low salt buffer (30 mM HEPES, pH8, 50 mM KCl, 6 mM BME), and finally eluted in 4 x 0.75 column volumes of elution buffer (50 mM HEPES, pH 8, 50 mM KCl, 6 mM BME, 300 mM imidazole, 10% glycerol). PreScission protease was used to remove the His-tag overnight. The protein was then purified on a HiTrapQ 5mL column using a buffer containing 50 mM HEPES pH 8, 6 mM BME, 5% glycerol and a gradient of 50 mM to 500 mM KCl. The eluted protein was then purified by size exclusion chromatography using an S75 Sephadex column and 50 mM HEPES, pH 8, 50 mM KCl, 1 mM DTT and 10% glycerol as the mobile phase.

Kinetics of purified Hsp90 α NTD modification

The kinetics of purified Hsp90 α NTD modification were measured by LC-MS in 20 mM HEPES pH 7.5, 50 mM KCl, 25 mM MgCl₂ (Labeling buffer). For determining k_{obs} under k_{inact} conditions, 1 μ M Hsp90 α NTD was incubated with 10 μ M compound (**1-6**) at 37°C in a heat block. At each time point, a 20 μ L aliquot was collected and added to 80 μ L of 2.5 % formic acid in a 96-well plate to quench the reaction. 2.5 μ L of this solution was injected into Waters Xevo G2-XS QToF LC-MS. For determination of K_i and k_{inact} , 50 nM Hsp90 α NTD was incubated with 0.14-3.3 μ M of compound **1** or **5**. At each time point, 80 μ L of this solution was collected and added to 20 μ L 10% formic acid (2% final) in a 96-well plate to quench the reaction. LC-MS runs were auto-processed using the Waters OpenLynx software. Briefly, a single peak eluting between 0.75 and 0.9 mins in the MS TIC chromatogram is selected. Background subtraction is performed on the resulting spectrum. Deconvolution of the charge envelope between 800 and 1600 m/z is performed using MaxEnt to determine the monoisotopic mass of the unlabeled and probe-labeled protein, using a target mass range of 25000 to 28000 Da, with a mass resolution of 1 Da. Quality control is performed using the OpenLynx Browser, and the resulting report is processed using the script in Appendix B to extract the intensity of the modified and unmodified protein. The percent modification is plotted as a function of time and fit to a single phase exponential association curve constrained to a plateau of 100, an initial value of 0.

Crystallization and structure determination of probe-modified Hsp90 α NTD

Hsp90 α NTD was first covalently-modified by **1**, **5**, and **6**, by incubating 2 μ M Hsp90 α NTD with 10 μ M compound in Labeling Buffer. Complete modification of the protein was confirmed by LC-MS before diafiltration in a 10K MWCO filter (Millipore Amicon Ultra) and buffer exchange into storage buffer containing 50 mM HEPES pH 8.0, 50 mM KCl, 1 mM DTT with 10% glycerol. The covalent probe-protein complex was concentrated to 20 mg/mL. Crystallization was performed

using the hanging drop method, using a 1:1 mixture of 10 mg/mL Hsp90 α -1 or Hsp90 α -5 adduct and a crystallization solution containing 20% PEG-3350, 0.2 M MgCl₂, 0.1 M sodium acetate, pH 5.6. Protein crystals formed within 24 to 72 hours at room temperature. Data collection was performed at the LBNL ALS beamline 8.3.1. Co-crystallization was performed as above, except the Hsp90 α K58R NTD was used. A solution of 10 μ M protein and 2 μ M compound was incubated at room temperature for 10 minutes prior to diafiltration, concentration and crystallization trials. The Hsp90 α K58R-5 complex was crystallized in 20% PEG-3350, 0.2 M MgCl₂, 0.1 M sodium acetate, pH 5.8. The Hsp90 α K58R-6 complex was crystallized in 20% PEG-3350, 0.2 M MgCl₂, 0.1M HEPES pH 7.5. The resulting X-ray diffraction data was processed using PHASER to perform molecular replacement using PDB Code: 2FWZ as a starting model. Structure refinement was performed using phenix.refine.

Biotin pull-down of probe-labeled Hsp90 from Skbr3 cells and MS sample prep

Skbr3 cells (gift from Moasser Lab) were grown in McCoy's 5A media (Gibco) supplemented with 10% fetal bovine serum (Axenia) and penicillin/streptomycin (Gibco). Cells were grown in a 500 cm² dish and pre-treated with DMSO or 3.3 μ M compound **1** for 90 mins at 37°C followed by a 90 min treatment with 3.3 μ M probe **2** at 37°C. Cells were washed twice with PBS and collected by scraping into 1 mL of Lysis buffer (0.1% NP-40, 100 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM PMSF, 1X cOmplete protease inhibitors without EDTA (Roche)). Protein concentration was normalized using bicinchoninic acid (BCA) protein quantification assay (Pierce). Click chemistry was performed by reacting 20.5 μ L of lysate with 4.25 μ L of TAMRA click master mix (Final click reactant concentrations of 1% SDS, 5 μ M TAMRA-picolyl azide, 1 mM TCEP, 100 μ M TBTA (from a 20 mM stock prepared in 4:1 *tert*-BuOH:DMSO), 1 mM CuSO₄, 2% DMSO) for 1 hour at RT. Samples were quenched with 5 μ L 6X SDS-PAGE loading dye and run on a 7.5% SDS-PAGE gel for 70 mins. The gel was briefly destained in transfer

buffer with a kimwipe and then imaged using a Typhoon Gel Imager (GE). Of the remaining sample, 4 mg of protein lysate (5 mg/mL) were cleared of endogenously biotinylated proteins using high capacity neutravidin agarose beads (Thermo #29204) overnight at 4 °C. Beads were removed by filtration (Pall #4650) and the filtrate was conjugated to a cleavable biotin azide (1% SDS, 100 µM TBTA (from a 20 mM stock prepared in 4:1 tert-BuOH:DMSO), 1 mM TCEP, 100 µM Biotin-DDE-picolyl azide (Click chemistry tools #1186), 1 mM CuSO₄) at room temperature for 90 mins. The reaction mixture was then added to 10 mL of acetone and incubated at -20°C overnight. The precipitated proteins were collected by centrifugation (3700 rpm, 4°C, 30 min), resuspended in cold methanol. Precipitated proteins were pelleted, dried, and redissolved in (1% SDS in PBS), and then diluted with (1% NP40 in PBS) to a final detergent concentration of (0.4% SDS, 0.6% NP40 in PBS) before desalting on a NAP-10 column to remove remaining excess biotin azide. The column eluate was incubated with high capacity Neutravidin beads at 4°C overnight. The beads were then washed with 1% NP-40, 0.1% SDS in PBS (3x10 min, RT), freshly prepared 6M urea in PBS (3x30min, 4°C) and PBS (3 X 10 min, RT), before a brief wash in digestion buffer (20 mM Tris, 2 mM CaCl₂, pH 8). Disulfide reduction was performed with 5 mM DTT at 56 °C for 30 minutes, followed by cysteine alkylation with 5 mM iodoacetamide for 15 min at RT in the dark. On-bead trypsinization was performed overnight at 37°C using 500 ng sequencing-grade trypsin (Promega #V5113). The resulting peptide solution was desalted using C18 OMIX tips (Agilent # A57003100), dried, and stored at -80°C. Probe-labeled peptides were recovered by extensively washing the beads to remove contaminating tryptic peptides consisting of PBS (5 x 5 min), 6M urea in PBS (5 x 5 min), water (5 x 5 min), 50% methanol (5 x 5 min), water (5 x 5 min) and finally cleaving the DDE linker with 2% aqueous hydrazine monohydrate (2 x 30 min). The probe-labeled peptide fraction was dried, redissolved in 200 µL 0.1% formic acid, and desalted using C18 OMIX tips. The eluted peptides were dried and stored at -80°C.

MS conditions for identification of tryptic peptides

Tryptic peptides were analyzed on a Fusion Lumos Tribrid (Thermo) connected to an ACQUITY M-Class UPLC system (Waters) and using an EASY-Spray 75 μm x 15 cm C18 column with 3 μm particle size (Thermo, ES800). Peptides were loaded onto a column warmed to 45°C, and equilibrated with 5% solvent B (0.1% formic acid in acetonitrile) for 3 mins at 600 nL/min. The peptides were then eluted with a flow rate of 300 nL/min in 5% solvent B for 3 min, followed by 5-30% B for 72 min, 50% B for 2 mins, and 5% B for 6 mins. Peptide mass to charge ratio was measured in the orbitrap at a resolution of 120,000, a scan range of 375-1500 m/z, with an AGC target of 40,000, a maximum injection time of 50 ms and with internal calibration enabled. Ions with a peptide-like isotopic distribution (MIPS set to "peptide"), that exceed an intensity threshold of 20,000 and that contained a charge between 2 and 7 were selected for HCD fragmentation. Peptides were selected for HCD fragmentation using a 1.6 m/z isolation window and an HCD collision energy of 30%. Fragment ions were measured in the orbitrap at a resolution of 30,000 and were collected with an AGC target of 50,000 and a maximum injection time of 100 ms. Peptides selected for fragmentation one time were then dynamically excluded from further selection for fragmentation for the following 30 seconds using a 10 ppm window. The maximum duty cycle was set to 3 s.

Analysis of MS data

Raw files were converted into peaklists using in-house software called PAVA. Database search was performed using Protein Prospector using the human SwissProt database downloaded on 4/8/2019 and concatenated with a corresponding randomized decoy database. Only tryptic peptides were considered with at most one missed cleavage. Carbamidomethylation of cysteine was included as a constant modification, while methionine oxidation, acetylation of the N-terminus, N-terminal glutamine to pyroglutamate conversion, and loss of the N-terminal methionine were included as variable modifications. When searching for probe-labeled peptides,

an additional variable modification (C33 H31 I N12 O5 S2) at uncleaved lysine residues and at tyrosine residues was included to account for the additional mass of the probe and triazole adduct. Precursor tolerance was set to 20 ppm, while fragment tolerance was set to 30 ppm. The maximum peptide false discovery rate was set to 1%. The raw data and the database search results, in biblispec format, were imported into Skyline version 4.2 in MS1 filtering mode with a precursor tolerance of 5 ppm, inclusion of three precursor isotope peaks, and extraction of MS chromatograms within 5 minutes of identification by MS/MS. Data were then analyzed with the MS Stats group comparison module, without normalization, to generate protein level fold changes and p-values between probe and competitor treated samples.

Fluorescence polarization

The protocol for the measurement of dissociation constants of compounds **1-6** and PU-H71 for Hsp90 α K58R NTD was adapted from ¹¹⁶. 30 nM Hsp90 α K58R NTD was incubated with 5 nM geldanamycin-FITC tracer (Invivogen, ant-fgl-1), and 1.5 nM to 10 μ M compound in Felts' buffer (20 mM HEPES pH 7.3, 50 mM KCl, 5 mM MgCl₂, 20 mM Na₂MoO₄, 0.01% NP-40 and 0.1 mg/ml bovine gamma globulin in 0.9% NaCl) in a non-binding, black 96-well plate (Greiner bio-one #655900) No protein, no geldanamycin-FITC and no compound controls were included on each plate. Samples were incubated overnight at 4°C on a rocker and fluorescence polarization was measured using a Tecan Spark plate reader. Background subtracted polarization in millipolarization (mP) units was calculated with the following formula: $1000 * ((\text{Perp}_{\text{sample}} - \text{Perp}_{\text{No probe}}) - (\text{Parallel}_{\text{sample}} - \text{Parallel}_{\text{No probe}})) / ((\text{Perp}_{\text{sample}} - \text{Perp}_{\text{No probe}}) + (\text{Parallel}_{\text{sample}} - \text{Parallel}_{\text{No probe}}))$. Specific polarization was calculated by subtracting the free probe (no protein) polarization from each sample.

Cell culture conditions and media for dose response and time course

The time-dependent labeling of **5** and **6** were measured by treating Skbr3 cells with 3.3 μM of each compound for the indicated amount of time. Dose-dependent labeling with **5** and **6** was performed by treating Skbr3 cells with 41 nM to 3.3 μM of compound for 90 min at 37°C. Excess compound was then removed by replacing the media with fresh, complete media for 4 x 10 min at 37°C. Cells were then washed twice with warm PBS and plates were stored at -80°C. Lysis buffer was added to each well and cells were collected by scraping. Protein amount was normalized by BCA. For occupancy measurements, lysates were incubated with 3.3 μM probe **2** at 37°C for 1 hour using a Bio-Rad thermocycler. Click chemistry was performed for 1 hour at RT (final click reactant concentrations of 1% SDS, 5 μM TAMRA-picolyl azide (Click chemistry tools, #1254), 1 mM TCEP, 100 μM TBTA (from a 20 mM stock prepared in 4:1 *tert*-BuOH:DMSO), 1 mM CuSO₄, 2% DMSO) and quenched with 6X SDS loading dye.

Western blot (antibodies, concentrations, duration of incubations)

SDS-page gels were transferred to 0.45 μM nitrocellulose membranes (Bio-Rad, 162-0115) using a Bio-Rad criterion transfer apparatus (Part # 1704070) and using Towbin Buffer for 75 V, 38 min at 4°C. Membranes were briefly stained with 0.1% Ponceau S solution in 5% acetic acid. Membranes were blocked for 1h at RT with 2% BSA in TBST and incubated with primary antibody overnight at 4°C (Hsp90 α/β , Stressmarq SMC-135D, 1:1000; Hsp70, Santa Cruz Biotech #sc-66048, 1:1000; Her2, Cell Signaling Technologies (CST), #2165S, 1:1000; pHer2, CST #2274S, 1:1000; Akt, CST #4685S, 1:1000; pAkt S473, CST #2920S, 1:1000; Tubulin, Sigma-Aldrich #T6199, 1:2500).

Synthetic methods:

8-mercaptoadenine: To a 250 mL round bottom flask was added 4,5,6-triaminopyrimidine sulfate hydrate (22.4 mmol), water (84 mL), ethanol (42 mL), sodium bicarbonate (112 mmol)

and carbon disulfide (224 mmol). The flask was equipped with a condenser and heated to 60°C under argon for 4 days. The reaction was then cooled to RT and carbon disulfide was removed by rotary evaporation. The resulting brown solution was filtered and 4 mL of acetic acid was added to produce an off-white precipitate. This was collected in a Buchner funnel and dried. Yield: 3.6g, 97%. ¹H-NMR (400 MHz; DMSO-d₆): δ 8.05 (s, 1H), 6.76 (d, J = 0.2 Hz, 2H).

5,6-diiodobenzo[d][1,3]dioxole: To a 1 L round bottom flask was added 1,3-benzodioxole (81.9 mmol), N-iodosuccinimide (278 mmol), acetonitrile (500 mL) and trifluoroacetic acid (245.7 mmol, 18.8 mL). The reaction was set to stir at room temperature and was monitored by NMR. The reaction was complete after 3 days. Solvent was removed and the residue dissolved in dichloromethane. The solution was washed four times with saturated sodium thiosulfate, two times with 1M sodium hydroxide, and once with brine. The organics were dried with sodium sulfate, concentrated and dried overnight. Product was used without further purification. Yield: 27.8g, 91%. ¹H-NMR (400 MHz; CDCl₃): δ 7.31 (s, 2H), 5.98 (s, 2H).

8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine: To a flame-dried 250 mL RBF was added anhydrous DMF (120 mL), 8-mercaptoadenine (18 mmol), 5,6-diiodobenzo[d][1,3]dioxole (54 mmol), copper iodide (3.6 mmol), and neocuproine hydrate (3.6 mmol). The solution was purged with argon for 1 h before adding sodium *tert*-butoxide (27 mmol) and heating at 110°C for 16 h. The solution was cooled, adsorbed onto Celite and purified by normal phase chromatography in 0-20% methanol. Yield: 2g, 27%. NMR: ¹H-NMR (400 MHz; DMSO-d₆): δ 8.21 (s, 1H), 7.76 (s, 2H), 7.55 (s, 1H), 7.15 (s, 1H), 6.12 (s, 2H).

3-((tert-butoxycarbonyl)amino)propyl methanesulfonate: Dichloromethane (3 mL), 3-(Boc-amino)-1-propanol (2.85 mmol), and triethylamine (6.6 mmol) was added to a 20 mL scintillation vial, which was set to stir on ice. Mesyl chloride (3.4 mmol) was added dropwise and the reaction was left to stir on ice for 2 hours. The reaction was monitored by TLC and potassium permanganate staining. When complete, it was diluted with saturated ammonium chloride, and extracted 3 times with dichloromethane. Organics were pooled, washed with brine, dried over sodium sulfate and concentrated. The resulting product was used without further purification. Yield: 722 mg, 98%. ¹H-NMR (400 MHz; CDCl₃): δ 4.72 (s, 1H), 4.29 (t, J = 6.0 Hz, 2H), 3.29-3.24 (m, 2H), 3.03 (s, 3H), 1.94 (quintet, J = 6.2 Hz, 2H), 1.44 (s, 10H).

tert-butyl 3-(((methylsulfonyl)oxy)methyl)azetidine-1-carboxylate: Dichloromethane (2.5 mL), tert-butyl 3-(hydroxymethyl)azetidine-1-carboxylate (2.15 mmol), and triethylamine (4.9 mmol) was added to a 20 mL scintillation vial, which was set to stir on ice. Mesyl chloride (2.6 mmol) was added dropwise and the reaction was left to stir on ice for 2 hours. The reaction was diluted with saturated ammonium chloride and extracted 3 times with dichloromethane. Organics were pooled, washed with brine, dried over sodium sulfate and concentrated. The resulting product was used without further purification. Yield: 93%. ¹H-NMR (400 MHz; CDCl₃): δ 5.30 (s,), 4.35 (d, J = 6.8 Hz, 2H), 4.05 (dd, J = 9.1, 8.4 Hz, 2H), 3.72 (dd, J = 9.1, 5.2 Hz, 2H), 3.14 (s,), 3.05 (s, 3H), 1.44 (s, 10H).

tert-butyl 3-(((methylsulfonyl)oxy)methyl)pyrrolidine-1-carboxylate: Dichloromethane (3 mL), tert-butyl 3-(hydroxymethyl)pyrrolidine-1-carboxylate (2.5 mmol), and triethylamine (5.7 mmol) was added to a 20 mL scintillation vial, which was set to stir on ice. Mesyl chloride (3 mmol) was added dropwise and the reaction was left to stir on ice for 2 hours. The reaction was diluted with saturated ammonium chloride and extracted 3 times with dichloromethane. Organics were pooled, washed with brine, dried over sodium sulfate and concentrated. The

resulting product was used without further purification. Yield: 89%. ¹H-NMR (400 MHz; CDCl₃): δ 4.22-4.14 (m, 3H), 3.56-3.53 (m, 2H), 3.37-3.35 (m, 2H), 3.14 (s, 2H), 3.03 (s, 3H), 2.64-2.62 (m, 2H), 2.06-2.03 (m, 1H), 1.46 (s, 10H).

(S)-tert-butyl 3-(((methylsulfonyl)oxy)methyl)piperidine-1-carboxylate: Dichloromethane (8 mL), tert-Butyl (3S)-3-(hydroxymethyl)piperidine-1-carboxylate (4.6 mmol), and triethylamine (10.7 mmol) was added to a 20 mL scintillation vial, which was set to stir on ice. Mesyl chloride (5.6 mmol) was added dropwise and the reaction was left to stir on ice for 15 hours. The reaction was diluted with saturated ammonium chloride and extracted 3 times with dichloromethane. Organics were pooled, washed with brine, dried over sodium sulfate and concentrated. The resulting product was used without further purification. Yield: 92%. ¹H-NMR (400 MHz; CDCl₃): δ 4.11-4.02 (m, 2H), 3.93-3.90 (m, 1H), 3.78 (d, *J* = 13.2 Hz, 1H), 3.00 (s, 3H), 2.90 (ddd, *J* = 13.4, 10.3, 3.2 Hz, 1H), 2.82-2.74 (m, 1H), 1.95-1.90 (m, 1H), 1.82-1.78 (m, 1H), 1.64 (td, *J* = 8.9, 4.2 Hz, 1H), 1.43 (s, 10H), 1.33 (dd, *J* = 33.5, 10.5 Hz, 2H).

(R)-tert-butyl 3-(((methylsulfonyl)oxy)methyl)piperidine-1-carboxylate: Dichloromethane (4 mL), tert-Butyl (3R)-3-(hydroxymethyl)piperidine-1-carboxylate (2.3 mmol), and triethylamine (5.3 mmol) was added to a 20 mL scintillation vial, which was set to stir on ice. Mesyl chloride (2.8 mmol) was added dropwise and the reaction was left to stir on ice for 15 hours. The reaction was diluted with saturated ammonium chloride and extracted 3 times with dichloromethane. Organics were pooled, washed with brine, dried over sodium sulfate and concentrated. The resulting product was used without further purification. Yield: 89%. ¹H-NMR (400 MHz; CDCl₃): δ 4.14-4.05 (m, 2H), 3.97-3.92 (m, 1H), 3.83-3.78 (m, 1H), 3.02 (s, 3H), 2.92 (ddd, *J* = 13.4, 10.3, 3.2 Hz, 1H), 2.80-2.78 (m, 1H), 1.97-1.94 (m, 1H), 1.85-1.80 (m, 1H), 1.69-1.64 (m, 1H), 1.46 (s, 9H), 1.36-1.30 (m, 2H).

9-(3-aminopropyl)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine: 8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine (1.2 mmol) was added to a 20 mL vial with dimethylformamide (4 mL). Cesium carbonate (1.2 mmol) was added and the reaction was set to stir at 80°C. Separately, 3-((tert-butoxycarbonyl)amino)propyl methanesulfonate (2.4 mmol) was dissolved in dimethylformamide (4 mL) and added dropwise to the reaction vial. The reaction was left to stir for 30 mins and was then cooled. DMF was removed under vacuum and the residue was re-dissolved in dichloromethane (4 mL) and set to stir on ice. Trifluoroacetic acid (4 mL) was then added dropwise and the reaction was left to stir. When complete, the reaction was concentrated, re-dissolved in dichloromethane and purified by normal phase chromatography using a 24g Gold Redisep column and a gradient of 30-50% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 145 mg, 25%. ¹H-NMR (400 MHz; MeOD): δ 8.16 (s, 1H), 7.45 (s, 1H), 7.16 (s, 1H), 6.07 (s, 2H), 4.32 (t, *J* = 7.0 Hz, 2H), 2.66 (t, *J* = 7.0 Hz, 2H), 2.01 (quintet, *J* = 7.0 Hz, 2H).

9-(azetidin-3-ylmethyl)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine: 8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine (0.35 mmol) was added to a 20 mL vial with dimethylformamide (1.25 mL). Cesium carbonate (0.35 mmol) was added and the reaction was set to stir at 80°C. Separately, tert-butyl 3-(((methylsulfonyl)oxy)methyl)azetidine-1-carboxylate (0.7 mmol) was dissolved in dimethylformamide (1.25 mL) and added dropwise to the reaction vial. The reaction was left to stir for 30 mins and was then cooled. DMF was removed under vacuum and the residue was re-dissolved in dichloromethane and purified by normal phase chromatography using a 4g Redisep column and a gradient of 4-6% methanol in dichloromethane. Fractions containing the product were concentrated, re-dissolved in dichloromethane (4 mL) set to stir on ice. Trifluoroacetic acid (4 mL) was then added dropwise

and the reaction was left to stir. When complete, the reaction was concentrated, re-dissolved in dichloromethane and purified by normal phase chromatography using a gradient of 30-100% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 34 mg, 20%. ¹H-NMR (400 MHz; MeOD): δ 8.18 (s, 1H), 7.49 (s, 1H), 7.18 (s, 1H), 6.10 (s, 2H), 4.50 (d, *J* = 7.2 Hz, 2H), 3.75 (t, *J* = 7.9 Hz, 4H), 3.37 (s, 1H).

8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9-(pyrrolidin-3-ylmethyl)-9H-purin-6-amine: 8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine (0.35 mmol) was added to a 20 mL vial with dimethylformamide (1.25 mL). Cesium carbonate (0.35 mmol) was added and the reaction was set to stir at 80°C. Separately, tert-butyl 3-(((methylsulfonyl)oxy)methyl)pyrrolidine-1-carboxylate (0.7 mmol) was dissolved in dimethylformamide (1.25 mL) and added dropwise to the reaction vial. The reaction was left to stir for 30 mins and was then cooled. DMF was removed under vacuum and the residue was re-dissolved in dichloromethane and purified by normal phase chromatography using a 4g Redisep column and a gradient of 0-10% methanol in dichloromethane. Fractions containing the product were concentrated, re-dissolved in dichloromethane (4 mL) set to stir on ice. Trifluoroacetic acid (4 mL) was then added dropwise and the reaction was left to stir. When complete, the reaction was concentrated, re-dissolved in dichloromethane and purified by normal phase chromatography using a 12g Redisep Gold column and a gradient of 0-100% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 25 mg, 14%. ¹H-NMR (400 MHz; MeOD): δ 8.16 (s, 1H), 7.46 (s, 1H), 7.17 (s, 1H), 6.07 (s, 2H), 4.25 (d, *J* = 7.5 Hz, 2H), 3.10 (dt, *J* = 5.5, 2.8 Hz, 1H), 2.97-2.76 (m, 4H), 1.92 (t, *J* = 6.5 Hz, 1H), 1.69-1.65 (m, 1H).

(S)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9-(piperidin-3-ylmethyl)-9H-purin-6-amine: 8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine (0.6 mmol) was added to a 20 mL vial with dimethylformamide (1.5 mL). Cesium carbonate (0.6 mmol) was added and the reaction was set to stir at 80°C. Separately, (S)-tert-butyl 3-(((methylsulfonyl)oxy)methyl)piperidine-1-carboxylate (1.2 mmol) was dissolved in dimethylformamide (1.5 mL) and added dropwise to the reaction vial. The reaction was left to stir for 30 mins and was then cooled. DMF was removed under vacuum and the residue was re-dissolved in dichloromethane and purified by normal phase chromatography using a 4g Rediseq column and a gradient of 0-10% methanol in dichloromethane. Fractions containing the product were concentrated, re-dissolved in dichloromethane (4 mL) set to stir on ice. Trifluoroacetic acid (4 mL) was then added dropwise and the reaction was left to stir. When complete, the reaction was concentrated, re-dissolved in dichloromethane and purified by normal phase chromatography using a 12g Rediseq Gold column and a gradient of 0-100% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 9%. ¹H-NMR (400 MHz; MeOD): δ 8.15 (s, 1H), 7.46 (s, 1H), 7.17 (s, 1H), 6.07 (s, 2H), 4.12-4.09 (m, 2H), 2.98-2.89 (m, 2H), 2.60-2.45 (m, 2H), 2.24 (td, J = 7.3, 3.7 Hz, 1H), 1.77-1.72 (m, 2H), 1.50-1.46 (m, 1H), 1.36-1.27 (m, 1H).

(R)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9-(piperidin-3-ylmethyl)-9H-purin-6-amine: 8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine (0.6 mmol) was added to a 20 mL vial with dimethylformamide (1.5 mL). Cesium carbonate (0.6 mmol) was added and the reaction was set to stir at 80°C. Separately, (R)-tert-butyl 3-(((methylsulfonyl)oxy)methyl)piperidine-1-carboxylate (1.2 mmol) was dissolved in dimethylformamide (1.5 mL) and added dropwise to the reaction vial. The reaction was left to stir for 30 mins and was then cooled. DMF was removed under vacuum and the residue was re-dissolved in dichloromethane and methanol, adsorbed onto silica and purified by normal phase chromatography using a 4g Rediseq column and a gradient of 0-10% methanol in dichloromethane. Fractions containing the product were

concentrated, re-dissolved in dichloromethane (3 mL) set to stir on ice. Trifluoroacetic acid (3 mL) was then added dropwise and the reaction was left to stir. When complete, the reaction was concentrated, re-dissolved in dichloromethane methanol, adsorbed onto Celite and purified by normal phase chromatography using a 12g Redisepp Gold column and a gradient of 20-50% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 45 mg, 15%. ¹H-NMR (400 MHz; MeOD): δ 8.15 (s, 1H), 7.46 (s, 1H), 7.17 (s, 1H), 6.07 (s, 2H), 4.16-4.04 (m, 2H), 2.98-2.88 (m, 2H), 2.61-2.54 (m, 1H), 2.44 (ddd, *J* = 29.8, 12.2, 10.7 Hz, 1H), 2.25-2.20 (m, 1H), 1.80-1.71 (m, 2H), 1.55-1.45 (m, 2H).

3-(((3-(6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-

yl)propyl)amino)methyl)benzene-1-sulfonyl fluoride (1): 9-(3-aminopropyl)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine (0.026 mmol) was added to a vial with dimethylformamide (100 μL). A solution of triethylamine (0.05 mmol) in DMF (100 μL) was added, followed by a solution of 1-fluorosulfonyl-3-bromomethyl benzene (0.023 mmol) in DMF (100 μL). The reaction was left to stir at RT for 30 min. Solvent was removed under vacuum and the product was purified by normal phase chromatography using 0-50% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 6.9 mg, 45%. ¹H-NMR (400 MHz; DMSO-d₆): δ 8.14 (s, 1H), 8.07 (d, *J* = 0.9 Hz, 1H), 7.99-7.97 (m, 1H), 7.86 (d, *J* = 7.8 Hz, 1H), 7.71 (t, *J* = 7.9 Hz, 1H), 7.46 (s, 1H), 7.35 (d, *J* = 0.3 Hz, 2H), 6.82 (s, 1H), 6.06 (s, 2H), 4.22 (t, *J* = 7.2 Hz, 2H), 3.77 (s, 2H), 2.48 (d, *J* = 6.9 Hz, 2H), 1.91-1.83 (m, 2H). Molecular formula:

C₂₂H₂₀FIN₆O₄S₂. Expected M+H⁺: 643.0094. Found: 643.0107

3-(((3-((6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-yl)methyl)azetid-1-

yl)methyl)benzene-1-sulfonyl fluoride (3): 9-(azetid-3-ylmethyl)-8-((6-

iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine (0.03 mmol) was added to a 4 mL reaction

vial with 100 μ L dimethylformamide. Triethylamine (0.06 mmol) in DMF (100 μ L) was then added, followed by 1-fluorosulfonyl-3-bromomethyl benzene (0.027 mmol) in DMF (100 μ L). The reaction stirred at RT for 30 mins. Solvent was removed under vacuum and the residue was dissolved in dichloromethane and purified by normal phase chromatography in 0-50% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 3.7 mg, 20%. $^1\text{H-NMR}$ (400 MHz; DMSO- d_6): δ 8.15 (s, 1H), 8.01-7.97 (m, 2H), 7.83-7.81 (m, 1H), 7.72 (t, J = 7.8 Hz, 1H), 7.48 (s, 1H), 7.38 (s, 2H), 6.84 (s, 1H), 6.05 (s, 2H), 4.38 (d, J = 7.1 Hz, 2H), 3.68 (s, 2H), 3.24-3.20 (m, 2H), 3.04-3.01 (m, 2H), 2.89-2.87 (m, 1H). Molecular formula: $\text{C}_{23}\text{H}_{20}\text{FIN}_6\text{O}_4\text{S}_2$
Expected $\text{M}+\text{H}^+$: 655.0094. Found: 655.0094

3-((3-((6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-yl)methyl)pyrrolidin-1-

yl)methyl)benzene-1-sulfonyl fluoride (4): 8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9-

(pyrrolidin-3-ylmethyl)-9H-purin-6-amine (0.03 mmol) was added to a 4 mL reaction vial with 100 μ L dimethylformamide. Triethylamine (0.06 mmol) in DMF (100 μ L) was then added, followed by 1-fluorosulfonyl-3-bromomethyl benzene (0.027 mmol) in DMF (100 μ L). The reaction stirred at RT for 30 mins. Solvent was removed under vacuum and the residue was dissolved in dichloromethane and purified by normal phase chromatography in 0-50% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 8.7 mg, 48%. $^1\text{H-NMR}$ (400 MHz; DMSO- d_6): δ 8.12 (s, 1H), 8.01 (d, J = 7.6 Hz, 2H), 7.90 (d, J = 7.8 Hz, 1H), 7.76-7.72 (m, 1H), 7.48 (s, 1H), 7.36 (s, 2H), 6.84 (s, 1H), 6.06 (s, 2H), 5.75 (s,), 4.13 (dd, J = 10.8, 7.8 Hz, 2H), 3.78-3.63 (m, 3H), 2.78-2.75 (m, 1H), 2.65 (d, J = 5.2 Hz, 1H), 2.41 (dd, J = 11.8, 4.1 Hz, 3H), 1.85-1.83 (m, 1H), 1.57-1.55 (m, 1H). Molecular formula: $\text{C}_{24}\text{H}_{22}\text{FIN}_6\text{O}_4\text{S}_2$. Expected $\text{M}+\text{H}^+$: 669.0251.
Found: 669.0236.

(S)-3-((3-((6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-yl)methyl)piperidin-1-yl)methyl)benzene-1-sulfonyl fluoride (5): (S)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9-(piperidin-3-ylmethyl)-9H-purin-6-amine (0.03 mmol) was added to a 4 mL reaction vial with 100 μ L dimethylformamide. Triethylamine (0.06 mmol) in DMF (100 μ L) was then added, followed by 1-fluorosulfonyl-3-bromomethyl benzene (0.03 mmol) in DMF (100 μ L). The reaction stirred at RT for 30 mins. Solvent was removed under vacuum and the residue was dissolved in dichloromethane and purified by normal phase chromatography in 0-50% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 14 mg, 70%. $^1\text{H-NMR}$ (400 MHz; DMSO- d_6): δ 8.14 (s, 1H), 8.00-7.97 (m, 2H), 7.83-7.81 (m, 1H), 7.72 (t, $J = 7.7$ Hz, 1H), 7.48 (s, 1H), 7.38-7.37 (m, 2H), 6.84 (s, 1H), 6.06-6.05 (m, 2H), 4.06-4.03 (m, 2H), 3.55 (d, $J = 3.2$ Hz, 2H), 2.56-2.54 (m, 2H), 2.15 (s, 1H), 2.03-1.91 (m, 2H), 1.64-1.62 (m, 1H), 1.51-1.47 (m, 1H), 1.36-1.35 (m, 1H), 1.05-1.02 (m, 1H). Molecular formula: $\text{C}_{25}\text{H}_{24}\text{FIN}_6\text{O}_4\text{S}_2$ Expected $\text{M}+\text{H}^+$: 683.0407. Found: 683.0417.

(R)-3-((3-((6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-yl)methyl)piperidin-1-yl)methyl)benzene-1-sulfonyl fluoride (6): (R)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9-(piperidin-3-ylmethyl)-9H-purin-6-amine (0.03 mmol) was added to a 4 mL reaction vial with 100 μ L dimethylformamide. Triethylamine (0.06 mmol) in DMF (100 μ L) was then added, followed by 1-fluorosulfonyl-3-bromomethyl benzene (0.03 mmol) in DMF (100 μ L). The reaction stirred at RT for 30 mins. Solvent was removed under vacuum and the residue was dissolved in dichloromethane and purified by normal phase chromatography in 0-50% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 11.3 mg, 56%. $^1\text{H-NMR}$ (400 MHz; DMSO- d_6): δ 8.14 (d, $J = 0.4$ Hz, 1H), 8.01-7.98 (m, 2H), 7.83 (dd, $J = 7.6, 0.4$ Hz, 1H), 7.74-7.70 (m, 1H), 7.49 (t, $J = 0.6$ Hz, 1H), 7.40-7.37 (m, 2H), 6.85 (t, $J = 0.6$ Hz, 1H), 6.06 (d, $J = 3.7$ Hz, 2H), 4.09-4.02 (m, 2H), 3.56 (d, $J = 3.0$ Hz, 2H), 2.61-2.53 (m, 2H), 2.16 (s, 1H), 2.04-1.91 (m, 2H),

1.67-1.62 (m, 1H), 1.53-1.47 (m, 1H), 1.37-1.33 (m, 1H), 1.06-1.02 (m, 1H). Molecular formula: $C_{25}H_{24}F_6O_4S_2$ Expected $M+H^+$: 683.0407. Found: 683.0417.

3-fluorosulfonyl benzoic acid: To a 250 mL round bottom flask was added 3-chlorosulfonyl benzoic acid (45.5 mmol) and 1,4-dioxane (56 mL). 2M aqueous potassium hydrogen difluoride (57 mL, 113 mmol) was then added, and the reaction was set to stir vigorously overnight. The reaction was added to a separatory funnel and partitioned between ethyl acetate and water. The organics were washed twice with water, once with brine and dried over sodium sulfate. The product was used without further purification. Yield: 8.5g, 92%. NMR: 1H -NMR (400 MHz; DMSO- d_6): δ 8.46 (d, $J = 8.4$ Hz, 2H), 8.40-8.38 (m, 1H), 7.94 (t, $J = 7.8$ Hz, 1H).

3-(fluorosulfonyl) methyl benzoate: 3-fluorosulfonyl benzoic acid (6.6 mmol) was added to a 100 mL round bottom flask with methanol (10 mL). Thionyl chloride (24 mL) was then added. The flask was equipped with a condenser and the reaction was set to stir at 75°C for 2 h. The reaction was cooled and excess thionyl chloride was removed under vacuum. Methanol was then added to the solution and excess methanol was removed under vacuum. The resulting product was used without further purification. Yield: 1.4g, 95%. NMR: 1H -NMR (400 MHz; CDCl $_3$): δ 8.68 (s, 1H), 8.44 (d, $J = 7.8$ Hz, 1H), 8.20 (d, $J = 8.0$ Hz, 1H), 7.75 (t, $J = 7.9$ Hz, 1H), 4.00 (s, 3H).

3-bromo-5-(fluorosulfonyl) methylbenzoate: 3-fluorosulfonyl methyl benzoate (34.4 mmol) was added to a 250 mL round bottom flask. Neat sulfuric acid (40 mL) was added to the flask and the solution was set to stir. Dibromoisocyanuric acid (17.2 mmol) was separately dissolved in neat sulfuric acid (40 mL) and this was added to the reaction flask. The reaction was monitored by TLC in 30% ethyl acetate in hexanes. When complete, the reaction was added

dropwise to 50 mL of ice. This solution was extracted five times with dichloromethane, washed with brine once and dried over sodium sulfate. The resulting product was used without further purification. Yield: 5.9g, 61%. ¹H-NMR (400 MHz; CDCl₃): δ 8.57 (t, *J* = 1.3 Hz, 1H), 8.54 (d, *J* = 1.3 Hz, 1H), 8.30 (t, *J* = 1.7 Hz, 1H), 4.00 (s, 3H).

3-(fluorosulfonyl)-5-((trimethylsilyl)ethynyl)methylbenzoate: 3-bromo-5-(fluorosulfonyl)methylbenzoate (13.5 mmol), and copper iodide (1.35 mmol) was added to a solution of tetrahydrofuran (35 mL), triethylamine (67.6 mmol, 9.5 mL), and trimethylsilylacetylene (47.3 mmol, 6.5 mL) in a 20 mL vial. This was set to stir on ice for 20 min.

Bis(triphenylphosphine)palladium(II) dichloride (2.7 mmol) was then added and the reaction was left to stir for 2 hours. The reaction was partitioned between dichloromethane and water. The aqueous layer was extracted three times with dichloromethane. The organic fractions were pooled, washed once with brine and dried over sodium sulfate. The crude mixture was adsorbed onto silica and purified in 0-10% ethyl acetate in hexanes. Yield: 3.35 g, 83%. ¹H-NMR (400 MHz; CDCl₃): δ 8.58 (s, 1H), 8.47 (s, 1H), 8.25 (s, 1H), 4.01 (s, 3H), 1.58 (s, 1H), 0.30 (s, 9H).

3-(hydroxymethyl)-5-((trimethylsilyl)ethynyl)benzene-1-sulfonyl fluoride: 3-(fluorosulfonyl)-5-((trimethylsilyl)ethynyl)methylbenzoate (8 mmol) was dissolved in dichloromethane (25 mL) and set to stir under argon in an acetonitrile-dry ice bath (-45°C). 1M DIBAL in THF (40 mmol) was added to the reaction flask by syringe pump over 1 hour, and the reaction was then left to stir for an additional 2 hr. The reaction was warmed to RT, quenched by addition of saturated ammonium chloride and water, and filtered through a fritted funnel. The filter cake was washed twice with dichloromethane. Organics were concentrated under vacuum, and the product was purified by normal phase chromatography using a 12g Rediseq column using a gradient of 0-

20% ethyl acetate in hexanes. Yield: 1.2 g, 53%. ¹H-NMR (400 MHz; CDCl₃): δ 8.02 (s, 1H), 7.97 (d, *J* = 0.3 Hz, 1H), 7.84 (s, 1H), 4.83 (d, *J* = 6.0 Hz, 2H), 2.07 (s, 1H), 0.30-0.28 (m, 9H).

3-ethynyl-5-(hydroxymethyl)benzene-1-sulfonyl fluoride: 3-(hydroxymethyl)-5-

((trimethylsilyl)ethynyl)benzene-1-sulfonyl fluoride (4.2 mmol) was dissolved in acetonitrile (7 mL). 2M aqueous potassium hydrogen difluoride (12.6 mmol, 6.3 mL) was added to the flask and set to stir vigorously at room temperature for 30 min. The organic layer was removed under vacuum and the aqueous layer was partitioned between ethyl acetate and water. The organic layer was washed twice with water, once with brine and dried over sodium sulfate before concentrating. The product was used without further purification. Yield: 835 mg, 93%. ¹H-NMR (400 MHz; CDCl₃): δ 8.02 (s, 1H), 7.99 (s, 1H), 7.85 (s, 1H), 4.82 (d, *J* = 5.9 Hz, 2H), 3.25 (s, 1H).

3-(bromomethyl)-5-ethynylbenzene-1-sulfonyl fluoride: 3-ethynyl-5-

(hydroxymethyl)benzene-1-sulfonyl fluoride (3.3 mmol), triphenylphosphine (3.9 mmol) and carbon tetrabromide (6.5 mmol) was dissolved in dichloromethane (20 mL) and set to stir for one hour at RT. The reaction was adsorbed onto silica and purified on a 4g GOLD Redisep column in 0-20% ethyl acetate in hexanes. Yield: 735 mg, 81%. ¹H-NMR (400 MHz; CDCl₃): δ 8.03 (t, *J* = 1.6 Hz, 1H), 7.98 (t, *J* = 1.7 Hz, 1H), 7.87 (s, 1H), 4.49 (s, 2H), 3.29 (s, 1H).

3-(((3-(6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-

yl)propyl)amino)methyl)-5-ethynylbenzene-1-sulfonyl fluoride (2): 9-(3-aminopropyl)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine (0.04 mmol) was added to a reaction vial

with dimethylformamide (100 μ L). Triethylamine (0.09 mmol) in DMF (100 μ L) was then added, followed by a solution of 3-(bromomethyl)-5-ethynylbenzene-1-sulfonyl fluoride (0.038 mmol) in DMF (100 μ L). The reaction was left to stir at RT for 30 min. DMF was removed under vacuum and the residue was re-dissolved in dichloromethane and purified in 0-50% (20% methanol, 2% ammonium hydroxide in dichloromethane). Yield: 8.6 mg, 34%. $^1\text{H-NMR}$ (400 MHz; DMSO-d_6): δ 8.14 (s, 1H), 8.11 (s, 1H), 8.04 (s, 1H), 7.97 (s, 1H), 7.46 (s, 1H), 7.36-7.36 (m, 2H), 6.82 (s, 1H), 6.06 (s, 2H), 4.54 (s, 1H), 4.23-4.20 (m, 2H), 3.79 (d, $J = 0.3$ Hz, 2H), 1.91-1.85 (m, 2H). Molecular formula: $\text{C}_{24}\text{H}_{20}\text{FIN}_6\text{O}_4\text{S}_2$. Expected $\text{M}+\text{H}^+$: 667.0094. Found: 667.0079.

Chapter 4: Development and characterization of lysine-reactive aryl aldehyde kinase probes by mass spectrometry

The work in this chapter was performed in collaboration with Dr. Tangpo Yang.

Abstract

Protein kinases comprise a conserved enzyme family whose members play a central role in signal transduction and cellular growth. Kinase inhibitors that function in an ATP-competitive manner have applications as both drugs and research tools. However, determining the selectivity of a kinase inhibitor, particularly in an in vivo context, remains challenging. Here, we report the synthesis of reversible covalent, aldehyde-based kinase inhibitors. Using these probes and a quantitative chemoproteomic pipeline (click chemistry with biotin azide, streptavidin pulldown, and label-free quantitative mass spectrometry), we have enriched over 50% of the endogenous human kinome across cell types and experiments. Despite promiscuously binding to hundreds of endogenous cellular kinases under equilibrium conditions, two salicylaldehyde-probes exhibit remarkable kinetic selectivity -- prolonged kinase residence time after probe washout -- for a small subset of the identified kinases. Lastly, we demonstrate that the clickable salicylaldehyde-based probe, YTP-2-137, exhibits the requisite balance of metabolic stability and on-target reactivity to covalently modify 80 kinases in mice treated with the probe.

Introduction

Protein kinases play a central role in signal transduction and the regulation of cell growth. They have emerged as an important class of drug targets, primarily in the context of targeted therapies for the treatment of cancer. However, most kinase inhibitors are ATP-competitive and must discriminate among hundreds of structurally related kinases. When kinase inhibitors are used as drugs, off-target inhibition can cause side effects and potentially limit efficacy. Additionally, when kinase inhibitors are used as tools, off-target inhibition can confound interpretation of experimental results. Accurate characterization of kinase inhibitor selectivity is therefore a critical part of both drug and tool development.

The most widely used methods for determining kinase inhibitor selectivity have several limitations. Assays that measure enzymatic activity with purified kinases or direct binding to endogenous kinases in cell lysates employ artificial conditions that often fail to reflect a compound's affinity under native conditions in living, intact cells.^{39,40,59,119} We have previously reported a broadly promiscuous kinase probe, XO44, which is cell permeable and reacts irreversibly with the catalytic lysine in over 130 kinases in living cells.²⁴ With this probe and a competitive kinase inhibitor of interest, measurements of kinase occupancy can be performed in a physiologically relevant cellular context. However, the sulfonyl fluoride probe's irreversible mechanism of action likely results in an underestimate of occupancy by a competitive inhibitor, particularly with reversible inhibitors that exhibit short residence times (fast dissociation rates, e.g., $t_{1/2} < 5$ min). Furthermore, sulfonyl fluorides are unlikely to be suitable for in vivo measurements of kinase occupancy due to their known chemical and metabolic instability.¹²⁰ Finally, sulfonyl fluorides are intrinsically reactive and may permanently covalently react with off-target proteins, resulting in toxicity. Hence, there remains a need for an improved kinase occupancy probe with a reversible covalent mode of inhibition, which can also function under demanding conditions in vivo.

The aryl aldehyde is an amine-reactive electrophile that can form a reversible imine adduct with the lysine epsilon-amine or with a protein's free N-terminal amine. Two late-stage clinical drug candidates contain an aryl aldehyde that reacts reversibly with the N-terminal amine in hemoglobin^{43,84} or with a cysteine residue in FGFR4,¹²¹ suggesting that aryl aldehydes may generally have the requisite balance of stability and reactivity for in vivo applications. A reversible covalent, promiscuous kinase probe with in vivo stability could directly address the shortcomings of existing kinase profiling technologies based on sulfonyl fluoride probes. In this work, we report the development of aryl aldehyde-containing promiscuous kinase probes that react with the catalytic lysine in a reversible covalent manner. Using these probes and a quantitative chemoproteomic pipeline, we demonstrate biochemical enrichment of over half the

human kinome. In addition, we demonstrate direct covalent engagement of dozens of kinases from mice dosed with a clickable, aldehyde-based kinase probe.

Results

Using a promiscuous kinase-binding pyrimidine 3-aminopyrazole scaffold, we synthesized aryl aldehyde-containing reversible covalent kinase probes that use a piperazine linker to orient the aldehyde toward the catalytic lysine, similar to our previous sulfonyl fluoride probe, XO44 (Fig. 4-1a). We treated Jurkat cells with the benzaldehyde-containing probe, YTP-2-147 (2 μ M, 30 min), and trapped putative hydrolytically labile imine adducts with sodium cyanoborohydride added directly to the cell lysis buffer. Probe-labeled proteins were then visualized by in-gel fluorescence following copper-catalyzed conjugation of a TAMRA-azide (TAMRA = tetramethylrhodamine). Qualitatively, the overall banding pattern and fluorescence intensity of YTP-2-147-labeled proteins is similar to the pattern observed with the sulfonyl fluoride-containing probe, XO44, suggesting that the two compounds largely share the same targets (Fig. 4-1b). Compound washout of XO44-treated samples showed no change in the fluorescence intensity of labeled proteins, as expected for an irreversible probe. However, samples treated with YTP-2-147 exhibited a dramatic reduction in labeling intensity within 10 min of compound washout (Fig. 4-1b). This result demonstrates that the benzaldehyde adduct is rapidly reversible and that YTP-2-147 exhibits a short residence time for most of its targets.

Several examples of *ortho*-substituted benzaldehyde-containing probes have been shown to have greater potency than their unsubstituted counterparts.^{43,45,47} *Ortho*-hydroxybenzaldehydes (also known as salicylaldehydes) can form imines that are kinetically stabilized by an intramolecular hydrogen bond.^{85,86,122} We therefore hypothesized that salicylaldehyde-based kinase

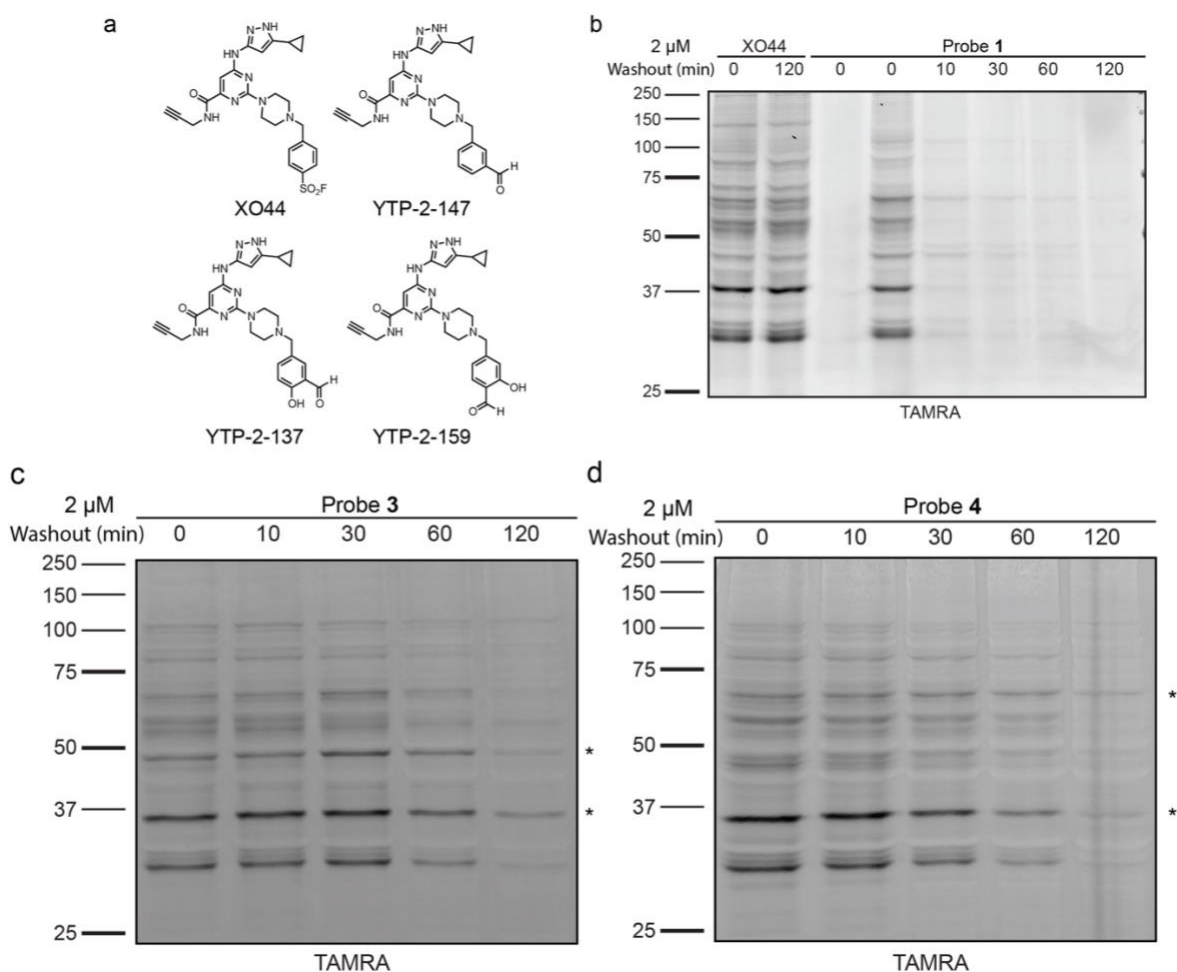


Figure 4-1: Monitoring aldehyde washout by in-gel fluorescence

a) Structures of the arylsulfonyl fluoride (XO44), benzaldehyde (YTP-2-147) and salicylaldehyde (YTP-2-137 and YTP-2-159) probes. b) Jurkat cells were treated with 2 μ M XO44 or YTP-2-147 for 30 min followed by compound washout. Probe-labeled proteins were then visualized by in-gel fluorescence after cell lysis, cyanoborohydride reduction, and TAMRA-azide conjugation. c and d) Jurkat cells were treated with 2 μ M YTP-2-137 or YTP-2-159 for 30 min, followed by compound washout. Probe-labeled proteins were visualized by in-gel fluorescence after cell lysis, borohydride reduction, and TAMRA-azide conjugation. Bands that appear to be more resistant to compound washout are indicated with an asterisk. All data and compounds shown in this figure are courtesy of Dr. Tangpo Yang.

probes would exhibit longer residence times than the benzaldehyde YTP-2-147. Jurkat cells treated with the salicylaldehydes YTP-2-137 and YTP-2-159 produced a labeling pattern by in-gel fluorescence comparable to XO44 and YTP-2-147 (Fig. 4-1). However, the fluorescence intensity of most of the salicylaldehyde-modified proteins (TAMRA-labeled bands) was unchanged 10 minutes after compound washout and slowly declined over the following two-hour

period. These results are consistent with dramatically increased residence times (decreased off-rates) for both salicylaldehyde probes. Additionally, select bands between 35 and 75 kDa exhibited a slower rate of decay in fluorescence intensity, which appeared to be independent of the magnitude of the initial fluorescence labeling intensity (Fig. 4-1c and d). This suggests that YTP-2-137 and YTP-2-159 exhibit a long residence time (> 1 hour) for a subset of targets.

We developed a chemoproteomic pipeline to identify the targets, and in particular the long-residence time targets, of each probe. Jurkat cells were treated with YTP-2-137, YTP-2-147 or YTP-2-159 (2 μ M, 30 min), rinsed briefly with PBS, and then incubated with compound-free media. Samples were collected at 0, 1, 3 and 6 h after compound washout and were lysed in buffer containing sodium cyanoborohydride (for YTP-2-147) or sodium borohydride (for YTP-2-137 and YTP-2-159). Probe-labeled proteins were enriched by streptavidin bead pulldown following copper-catalyzed conjugation of biotin-azide. Enriched proteins were washed extensively, and on-bead trypsinization of the retained proteins was performed, followed by peptide identification by LC-MS/MS. Across all three compounds, a total of 186 kinases were identified immediately after compound washout, of which 114 kinases are shared among the three compounds (Fig. 4-2a). Median kinase intensity for all three compounds is between three and ten-fold greater than non-kinase intensity, and 8-12-fold greater than the kinase intensity in the DMSO control (Fig. 4-2b). Together, these data demonstrate that YTP-2-137, YTP-2-147 and YTP-159 are covalent probes that can selectively enrich their kinase targets from living cells.

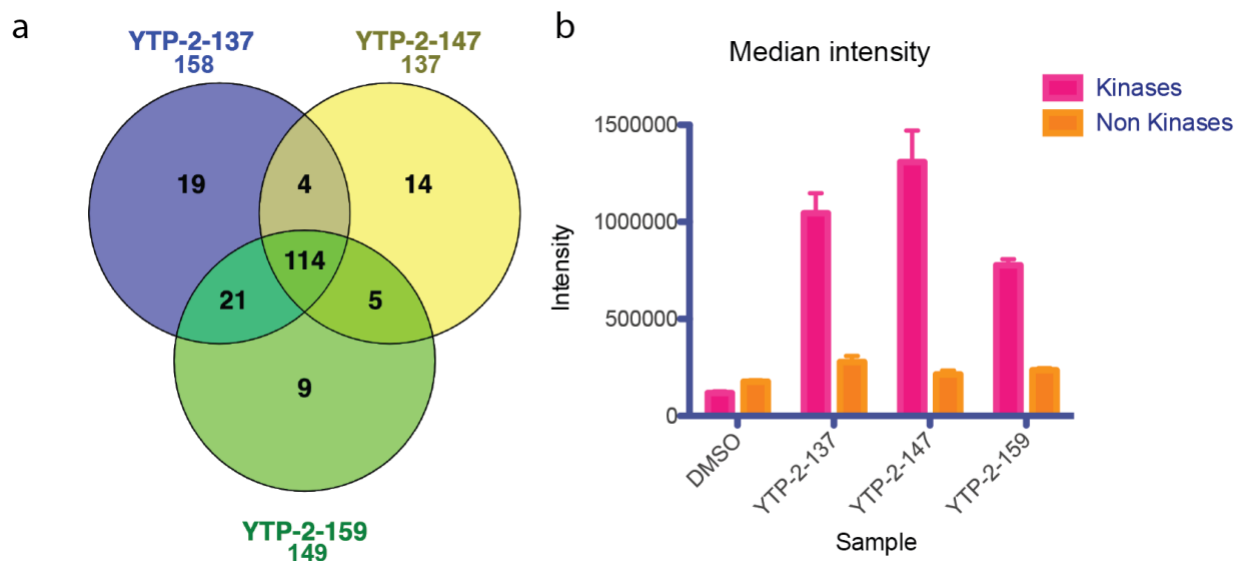


Figure 4-2: Kinase enrichment from Jurkat cells treated with aryl aldehyde probes

Jurkat cells were treated with YTP-2-137, YTP-2-147 or YTP-2-159 (2 μ M, 30 min). Probe-labeled proteins were identified after cell lysis, (cyano)borohydride reduction, biotin-azide conjugation, streptavidin pulldown, on-bead trypsinization, and finally, LC-MS/MS analysis of the resulting peptides. a) Overlap of kinases identified using YTP-2-137, YTP-2-147 and YTP-2-159. Only kinases that exhibit 10-fold enrichment over DMSO are included. b) Median intensity of kinases and non-kinases identified using the indicated compounds.

Using the washout samples, we next quantified the residence times for YTP-2-137, YTP-2-147 and YTP-159 at their bound kinases. Of the kinases enriched by YTP-2-147, 98% exhibited $\geq 90\%$ reduction in intensity after a 1-hour washout; $\sim 65\%$ of these kinases were undetectable after washout (Fig. 4-3a), consistent with the in-gel fluorescence results showing extensive compound washout within 10 minutes (Fig. 4-1b). This is in contrast to the 18 and 9 kinases, respectively, enriched with $\geq 50\%$ of their initial intensity after a 1-hour washout of the salicylaldehyde compounds YTP-2-137 and YTP-2-159. Even 6 hours after compound washout, a total of 5 kinases were enriched with $\geq 50\%$ of their initial intensity by YTP-2-137 (Aurora A, Aurora B, Mast3, and Sgk3) and YTP-2-159 (Mast2, Mast3, and Sgk3). Although both salicylaldehyde probes are promiscuous kinase inhibitors under equilibrium conditions, they become relatively selective kinase inhibitors after washout, which facilitates kinetic discrimination among kinase targets on the basis of residence time. Since the structurally

related benzaldehyde, YTP-2-147, exhibits a dissociation half-life of less than 20 minutes for all of its targets, we hypothesize that the observed kinetic discrimination is based primarily on the slow hydrolysis of the imine formed by the salicylaldehyde probes. We speculate that kinase targets with the longest compound residence times form imines with low solvent accessibility

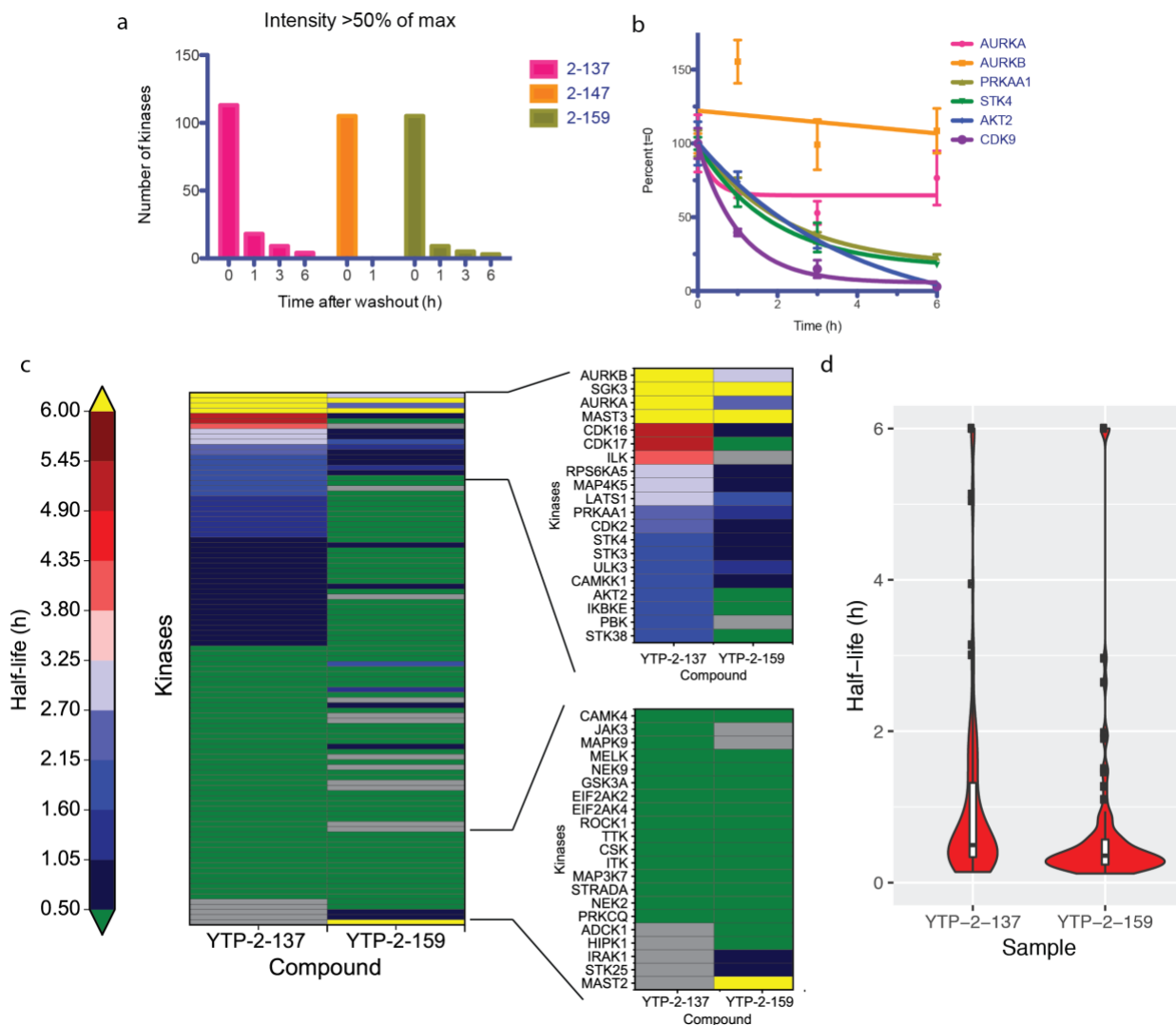


Figure 4-3: LC-MS/MS measurement of kinase probe half-life

a) Number of kinases remaining at each time point with greater than 50% of their initial intensity. Only kinases that are enriched 10-fold over DMSO, with at least three unique peptides are shown. b) The time-dependent intensity of select kinases is shown. The intensity was fit to a single-phase exponential decay function, from which the half-life of compound dissociation was determined. c) Heat map showing the dissociation half-life of YTP-2-137 and YTP-2-159 from their kinase targets. d) Violin blots with overlaid box and whisker plots showing the distribution of half-lives for each compound.

and/or are able to orient the catalytic lysine such that internal strain of the resulting imine is minimized and intramolecular hydrogen bonding between the imine nitrogen and the *ortho*-hydroxyl is maximized.

The 5 kinases that bind most tightly to YTP-2-137 and YTP-2-159 are all serine/threonine kinases with disparate roles in cell biology. Aurora A and Aurora B are both essential kinases that play a role in cell cycle progression. In addition, Aurora A is thought to play a critical role in neuroblastoma progression through its interactions with MYCN.^{123–126} Recent work demonstrates that simultaneous inhibition of Aurora and the PI3K/mTOR pathway may overcome acquired or intrinsic resistance to PI3K inhibitors in preclinical breast cancer models.¹²⁷ Mast2 (Microtubule Associated Serine/Threonine kinase 2) inhibits neurite outgrowth and neuronal survival, perhaps through its interactions with PTEN^{128–130} However, chromosomal translocations resulting in gene fusions *ARID1A* and *MAST2* have been identified in the MDA-MD-468 breast cancer cell line, while fusions with *GPBP1L1* and *MAST2* have been identified in patient-derived tumor samples.¹³¹ Whether the kinase catalytic activity of *MAST2* is required for these fusion proteins to act as drivers in cancer is unknown. Selective *MAST2* inhibitors could be used to probe this function or developed into compounds that promote the degradation of this fusion protein. Mast3 regulates the response to changes in cyclic AMP in neurons,^{132,133} while Sgk3 acts downstream of VPS34 and Class I PI3Ks.^{134,135} Selective inhibitors are not known for either kinase. We propose that long-residence time, salicylaldehyde-based inhibitors could be developed to selectively inhibit each of these kinases. Such tools would be valuable elucidating the biological roles and therapeutic potential of these targets.

We sought to more precisely quantify the dissociation half-life of YTP-2-137 and YTP-2-159 for each of kinase by fitting time-dependent enrichment intensities (from label-free MS) to a single-exponential decay model. Since this model does not take into consideration all of the factors that account for ligand binding and dissociation kinetics in cells, such as rapid ligand

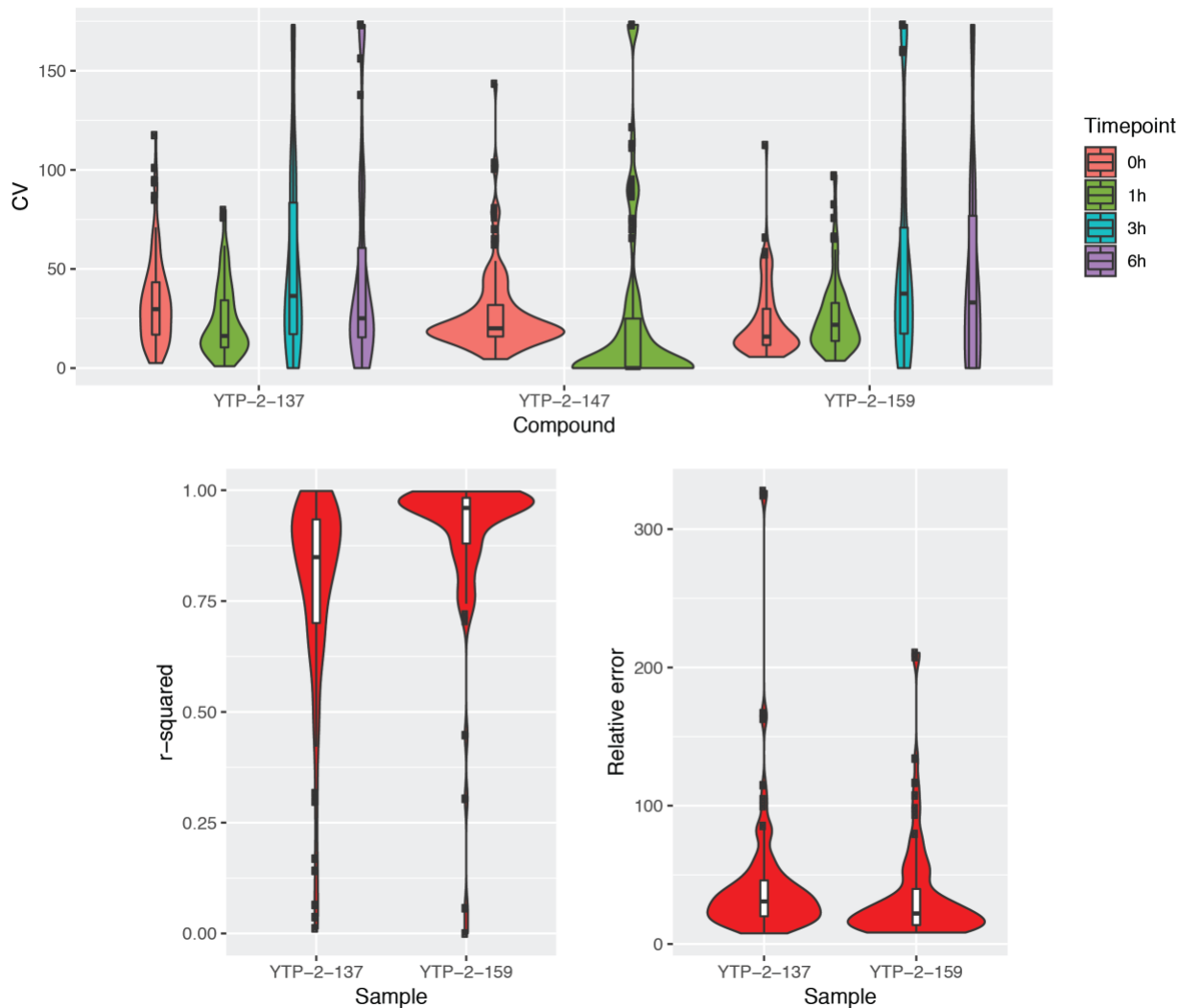


Figure 4-4: Reproducibility of replicates and goodness of fit.

A) Violin plots depicting distribution of coefficient of variation among the biological triplicate measurements for each kinase. Box and whisker plot is included within each violin plot and depicts the median with a bold line. b and c) Violin plots depicting coefficient of determination (r^2) (b) and the relative error (c) for the fit of a single exponential decay function to kinase intensity after washout.

rebinding, slow membrane permeability, and ligand accumulation within cells, the resulting apparent half-lives likely do not reflect true biochemical dissociation rates in most cases.

Furthermore, compound dissociation rates for a given kinase could be heterogeneous (and therefore appear to be multi-modal) due to the existence of biochemically distinct kinase subpopulations, e.g., governed by distinct post-translational modifications and protein-protein or protein-lipid interactions in the cell. The observed decay in enriched kinase intensity upon

compound washout could reflect a superposition of off-rates for each of these states.

Nevertheless, this apparent half-life does allow relative ordering of kinases based on the extent to which they release each probe upon washout. Kinases that do not exhibit significant probe dissociation after 6 h post-washout, such as Aurora A and B, fit the model poorly, whereas kinase/probe interactions with half-lives of less than 3 h show better overall agreement with the model (Fig. 4-3b and 4-4). The median apparent half-life was 0.5 h ($n = 113$ kinases) and 0.36 h ($n = 105$ kinases) for YTP-2-137 and YTP-2-159, respectively. The distribution of half-lives for YTP-2-137 skewed longer than YTP-2-159 (Fig. 4-3c and d). We conclude that YTP-2-137 and YTP-2-159 are reversible covalent kinase probes with strikingly long residence time for a subset of kinase targets.

Given the encouraging results with YTP-2-137 in cultured cells, we were motivated to test its suitability as a kinase probe in a living animal. In collaboration with John Kath and coworkers at Pfizer, mice were treated intraperitoneally with vehicle or 10 mg/kg YTP-2-137, leading to a maximum plasma concentration of ~50 nM YTP-2-137, 3 h after injection. Lungs from each mouse were harvested at this time point, flash frozen and subsequently processed as described above with an additional step for tissue homogenization. LC-MS/MS analysis revealed specific enrichment of 80 kinases, including 22 that had not previously been identified in Jurkat cells (Fig. 4-5a). Furthermore, the median kinase intensity was at least 3-fold greater than the non-kinase intensity from each mouse, and approximately 10-fold greater than the median kinase intensity in any of the vehicle-treated mice (Fig. 4-5b). To estimate the extent of target engagement *in vivo*, we performed a probe labeling experiment in mouse lung lysates with 2 μ M YTP-2-137—conditions that we estimate will result in ~100% occupancy of most kinase targets. Under these conditions, 223 kinases were identified from mouse lung lysates, including 79 of the 80 kinases identified from *in vivo* treatment with YTP-2-137. This suggests that the *in vivo* experiment was likely performed at subsaturating concentrations. Altogether, across multiple chemoproteomic experiments performed with human Jurkat T-cell and mouse

lung proteomes, a total of 261 kinases (~50% of the kinome) have been targeted, enriched, and identified using the salicylaldehyde probe, YTP-2-137.

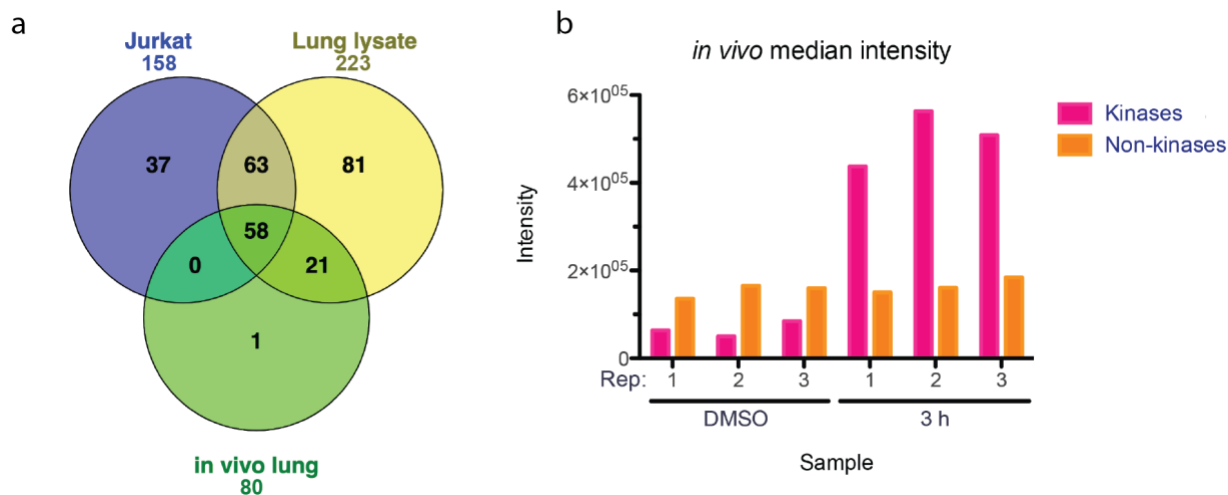


Figure 4-5: In vivo treatment with YTP-2-137

a) Venn diagram comparing kinases identified by YTP-2-137 pulldown from Jurkat cells, mouse lung lysates treated with YTP-2-137, or lung lysates from mice treated with YTP-2-137. The kinases enriched from treated mice comprise a subset of the kinase enriched from compound-treated lung lysates. More unique kinases were identified from the lung samples in this experiment. b) Median intensity of kinases and non-kinases enriched from each mouse treated with either vehicle or YTP-2-137.

In this study, we have developed and characterized the first salicylaldehyde-based reversible covalent kinase probes, to our knowledge. Using these clickable probes and a chemoproteomic pipeline, we have identified approximately half of the human kinome across experimental conditions and cell types, including the enrichment of 89 kinases from lung tissue from mice dosed with YTP-2-137. Despite promiscuous kinase engagement, we have uncovered striking kinetic discrimination -- based on distinct probe/kinase residence times -- for a large subset of kinase targets. This "residence time-based selectivity", achieved by the judicious choice of a reversible electrophile, represents a novel approach for achieving kinase inhibitor selectivity, which we propose can be improved by appending a salicylaldehyde to a more selective noncovalent-recognition scaffold. Although a reversible covalent strategy has been employed by our lab and others to target cysteines in kinase active sites, not all kinases

have a cysteine. By contrast, all kinases have at least one lysine, and in many cases, 2 or more lysines proximal to the active site. Finally, the in vivo activity of YTP-2-137 demonstrates the feasibility of using clickable probes to quantify kinase inhibitor occupancy in vivo.

Methods:

Treatment and washout of YTP-2-137, YTP-2-147 and YTP-2-159

Cell culture experiments for this chapter were performed by Dr. Tangpo Yang. Jurkat cells were grown in RPMI (Gibco) supplemented with 10% FBS (Axenia) and penicillin/streptomycin (Gibco). Jurkat cells (9.6×10^7 per sample, 80 mL) were treated with 2 μ M of YTP-2-137, YTP-2-147 or YTP-2-159 in biological triplicate for 30 mins at 37 °C. Cells were then collected by centrifugation, washed twice with warm PBS and re-plated in fresh, compound-free RPMI. Samples were collected at 0, 1, 3 and 6 h after this washout. Each sample was then collected by centrifugation, washed twice with cold PBS and lysed in sodium borohydride (YTP-2-137 and YTP-2-159) or sodium cyanoborohydride-containing (YTP-2-147) lysis buffer (100 mM HEPES, 150 mM NaCl, 0.1% NP-40 pH 7.5, 2 mM PMSF, 2X EDTA-free cOmplete protease inhibitors, and 5 mM sodium borohydride or 25 mM sodium cyanoborohydride). Lysates were incubated on ice for 60 mins before clarifying lysates by centrifugation. Protein quantification was performed using the Pierce bicinchoninic acid (BCA) protein quantification assay.

MS sample prep

Lysate processing for mass spectrometry analysis was performed by Dr. Tangpo Yang. 7.5 mg of reduced cell lysate at 1.5 mg/mL was depleted of endogenously biotinylated proteins by overnight incubation with agarose streptavidin beads (Thermo #20349) at 4°C. Beads were removed by filtration (Pall #4650) and probe-labeled proteins were functionalized with a biotin-azide by copper-catalyzed click chemistry (1% SDS, 100 μ M TBTA (from a 20 mM stock prepared in 4:1 tert-BuOH:DMSO), 1 mM TCEP, 100 μ M Biotin-DMTP-picolyl azide, 1 mM CuSO₄). The reaction mixture was incubated at room temperature for 1.5 h, followed by addition of 40 μ L of high capacity Neutravidin agarose beads (Thermo #29204) to each sample. Samples were incubated overnight at 4°C. The following day, the supernatant was removed, and beads were washed with 1% NP-40, 0.1% SDS (3 x 10 min, RT), 6 M urea in PBS (3 x 30

min, 4°C) and cold PBS (3 x 10 min, RT). Beads were then rinsed twice in digestion buffer (20 mM Tris, 2 mM CaCl₂, pH 8.0). On-beads reduction (30 min, 56°C) of disulfide bonds was performed with 5 mM DTT in digestion buffer, followed by addition of iodoacetamide to a final concentration of 10 mM to alkylate free cysteine residues (15 min, RT, in the dark). Each sample was then incubated with 500 ng trypsin overnight at 37 °C. Trypsinization was quenched with formic acid (2% final concentration). Supernatant was collected and peptides were desalted with OMIX tips, dried and stored at -80°C.

YTP-2-137 labeling in mouse lung lysates

Mouse lung samples were prepared by Dr. Tangpo Yang. Frozen mouse lungs were a gift from the Ptacek lab. 8 frozen mouse lungs were lysed on ice using a Dounce homogenizer containing 18 mL of cold lysis buffer (100 mM HEPES, 150 mM NaCl, 0.1% NP-40 pH 7.5, 2 mM PMSF, 2X EDTA-free cOmplete protease inhibitors). Lysates were clarified by centrifugation. Protein amount was quantified by BCA and biotinylated proteins were depleted from 20 mg of lysate using 1.3 mL of streptavidin agarose beads (Thermo #20349) by incubating for 1 hour incubation at 4°C. YTP-2-137 was added to a final concentration of 2 µM and samples were incubated at RT for 30 mins. Samples were then treated with 25 mM sodium borohydride at RT for 30 mins. Click chemistry conjugation of a biotin azide was then performed (1% SDS, 100 µM BTAA, 1 mM TCEP, 100 µM Biotin-DMTP-picolyl azide, 1 mM CuSO₄) for 1.5 h at RT. Acetone precipitation was then performed overnight at -20 °C. The precipitated proteins were collected by centrifugation (3700 rpm, 4°C, 30 min), and resuspended in cold methanol. Precipitated proteins were pelleted, dried, and redissolved in (1% SDS in PBS), and then diluted with (1% NP40 in PBS) to a final detergent concentration of (0.4% SDS, 0.6% NP40 in PBS) before desalting on a NAP-10 column to remove remaining excess biotin azide. The column eluate was incubated with streptavidin magnetic beads (Thermo #88817) at 4°C

overnight. The beads were then washed with 1% NP-40, 0.1% SDS in PBS (3x10 min, RT), freshly prepared 6M urea in PBS (3x30min, 4°C) and PBS (3 X 10 min, RT), before a brief wash in digestion buffer (20 mM Tris, 2 mM CaCl₂, pH 8). Disulfide reduction was performed with 5 mM DTT at 56 °C for 30 minutes, followed by cysteine alkylation with 10 mM iodoacetamide for 15 min at RT in the dark. On-bead trypsinization was performed overnight at 37°C using 500 ng sequencing-grade trypsin. The resulting peptide solution was desalted using C18 OMIX tips (Agilent #A57003100), dried, and stored at -80°C.

YTP-2-137 labeling in mice

Frozen mouse lungs from mice treated intraperitoneally with 10 mg/kg YTP-2-137 were received from collaborators at Pfizer. Samples were processed by Dr. Tangpo Yang. Lungs were homogenized as described above. Reduction was performed on the lysate using 10 mM sodium borohydride for 30 min at 4°C. Samples were then processed as described above in the Methods section titled “YTP-2-137 labeling in mouse lung lysates.”

MS conditions for identification of tryptic peptides

Tryptic peptides were analyzed on a Fusion Lumos Tribrid (Thermo) connected to an ACQUITY M-Class UPLC system (Waters) and using an EASY-Spray 75 µm x 15 cm C18 column with 3 µm particle size (Thermo, ES800). Peptides were loaded onto a column warmed to 45°C, and equilibrated with 5% B (0.1% formic acid in acetonitrile) for 3 mins at 600 nL/min and they were then eluted with a flow rate of 300 nL/min and using 5% solvent B for 3 min, followed by 5-30% B for 72 min, 50% B for 2 mins, and 5% B for 6 mins. Ion mass to charge ratio was measured in the orbitrap at a resolution of 120,000, a scan range of 375-1500 m/z, with an AGC target of 40,000, a maximum injection time of 50 ms and with internal calibration enabled. Ions with a peptide-like isotopic distribution (MIPS set to “peptide”) that exceeded an intensity threshold of 20,000 and contained a charge between 2 and 7 were selected for HCD fragmentation. MS2

spectra of HCD fragmented peptides were collected using a 1.6 m/z isolation window and an HCD collision energy of 30%. Fragment ions were measured in the orbitrap at a resolution of 30,000 and were collected with an AGC target of 50,000 and a maximum injection time of 100 ms. Peptides selected for fragmentation one time were then dynamically excluded from further selection for fragmentation for the following 30 seconds using a 10 ppm window. The maximum duty cycle was set to 3 s.

Database matching and peptide quantification.

Raw files were converted into peaklists using in-house software called PAVA. Database search was performed using Batch-Tag within Protein Prospector v 5.23.0 and were searched against a human SwissProt database downloaded on 4/8/2019 and concatenated with a corresponding randomized decoy database. Only tryptic peptides were considered with at most one missed cleavage. Carbamidomethylation of cysteine was included as a constant modification, while methionine oxidation, acetylation of the N-terminus, N-terminal glutamine to pyroglutamate conversion, and loss of the N-terminal methionine were included as variable modifications. Precursor tolerance was set to 20 ppm, while fragment tolerance was set to 30 ppm. Following database matching, quantification of peptide intensity was performed using Search Compare within Protein Prospector. The maximum peptide false discovery rate was set to 1%. Ion chromatograms for each peptide were extracted and intensity calculated using a retention time window of ± 20 seconds. The resulting data was exported as a tab-delimited text file.

Protein quantification and apparent half-life calculation.

Protein intensity was calculated by taking the sum of the peptide intensities for all peptides that uniquely match a single protein, using the script PP_LFQ_v5.py (included in Appendix B). The results from this script are reformatted into an annotated table using format-and-combine-protein-value_v2.py (included in Appendix B). A kinase was defined as enriched by a probe if it

was identified with at least one peptide in one drug-treated sample and it had a protein intensity greater than 10-fold over the DMSO control. A kinase was used for quantification of drug apparent half-life if it was identified with at least three unique peptides and had a protein intensity greater than 10-fold over the DMSO control. The apparent half-life of a given probe for each kinase was calculated by fitting the data from biological triplicates to a single exponential decay function using the script `batch_exp_fit_triplicates.py` (included in Appendix B) which implements the `lmfit` python curve fitting module. Heatmaps were generated in python using the `matplotlib` library. Violin plots summarizing the resulting fits were generated using the `ggplot2` library in R.

References

1. Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chemie Int. Ed. Ed.* **55**, 13408–13421 (2016).
2. Jones, L. H. Reactive Chemical Probes: Beyond the Kinase Cysteinome. *Angew. Chemie Int. Ed.* **57**, 9220–9223 (2018).
3. Copeland, R. A., Pompliano, D. L. & Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* **5**, 730–739 (2006).
4. Copeland, R. A. The drug-target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discov.* **15**, 87–95 (2016).
5. Moellering, R. E. & Cravatt, B. F. How chemoproteomics can enable drug discovery and development. *Chem. Biol.* **19**, 11–22 (2012).
6. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548–51 (2013).
7. Lito, P., Solomon, M., Li, L. S., Hansen, R. & Rosen, N. Allele-specific inhibitors inactivate mutant KRAS G12C by a trapping mechanism. *Science* **351**, 604–608 (2016).
8. Van Neck, T. *et al.* Inhibition of the CRM1-mediated nucleocytoplasmic transport by N-azolylacrylates: Structure-activity relationship and mechanism of action. *Bioorganic Med. Chem.* **16**, 9487–9497 (2008).
9. Neggers, J. E. *et al.* Identifying drug-target selectivity of small-molecule CRM1/XPO1 inhibitors by CRISPR/Cas9 genome editing. *Chem. Biol.* **22**, 107–116 (2015).
10. Patricelli, M. P. *et al.* Selective inhibition of oncogenic KRAS output with small molecules targeting the inactive state. *Cancer Discov.* **6**, 316–329 (2016).

11. Walker, C. J. *et al.* Preclinical and clinical efficacy of XPO1/CRM1 inhibition by the karyopherin inhibitor KPT-330 in Ph+ leukemias. *Blood* **122**, 3034–3044 (2013).
12. Lapalombella, R. *et al.* Selective inhibitors of nuclear export show that CRM1 / XPO1 is a target in chronic lymphocytic leukemia. *Blood* **120**, 4621–4634 (2012).
13. Miller, R. M. & Taunton, J. Targeting Protein Kinases with Selective and Semipromiscuous Covalent Inhibitors. *Methods Enzymol.* **548**, 93–116 (2014).
14. Liu, Q. *et al.* Developing Irreversible Inhibitors of the Protein Kinase Cysteine. *Chem. Biol.* **20**, 146–159 (2013).
15. Singh, J., Petter, R. C. & Kluge, A. F. Targeted covalent drugs of the kinase family. *Curr. Opin. Chem. Biol.* **14**, 475–480 (2010).
16. Barf, T. & Kaptein, A. Irreversible protein kinase inhibitors: Balancing the benefits and risks. *J. Med. Chem.* **55**, 6243–6262 (2012).
17. Li, D. *et al.* BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. *Oncogene* **27**, 4702–4711 (2008).
18. Cross, D. A. E. *et al.* AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discov.* **4**, 1046–61 (2014).
19. Honigberg, L. A. *et al.* The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. *Proc. Natl. Acad. Sci.* **107**, 13075–80 (2010).
20. Rabindran, S. K. *et al.* Antitumor activity of HKI-272, an orally active, irreversible inhibitor of the HER-2 tyrosine kinase. *Cancer Res.* **64**, 3958–3965 (2004).
21. Byrd, J. C. *et al.* Acabrutinib (ACP-196) in Relapsed Chronic Lymphocytic Leukemia. *N.*

- Engl. J. Med.* **374**, 323–332 (2016).
22. Grimster, N. P. *et al.* Aromatic sulfonyl fluorides covalently kinetically stabilize transthyretin to prevent amyloidogenesis while affording a fluorescent conjugate. *J. Am. Chem. Soc.* **135**, 5656–68 (2013).
 23. Hoppmann, C. & Wang, L. Proximity-enabled bioreactivity to generate covalent peptide inhibitors of p53–Mdm4. *Chem. Commun.* **52**, 5140–5143 (2016).
 24. Zhao, Q. *et al.* Broad-spectrum kinase profiling in live cells with lysine-targeted sulfonyl fluoride probes. *J. Am. Chem. Soc.* **139**, 680–685 (2017).
 25. Gushwa, N. N., Kang, S., Chen, J. & Taunton, J. Selective targeting of distinct active site nucleophiles by irreversible SRC-family kinase inhibitors. *J. Am. Chem. Soc.* **134**, 20214–7 (2012).
 26. Baranczak, A. *et al.* A Fluorogenic Aryl Fluorosulfate for Intraorganellar Transthyretin Imaging in Living Cells and in *Caenorhabditis elegans*. *J. Am. Chem. Soc.* **137**, 7404–7414 (2015).
 27. Mortenson, D. E. *et al.* ‘Inverse Drug Discovery’ Strategy to Identify Proteins That Are Targeted by Latent Electrophiles As Exemplified by Aryl Fluorosulfates. *J. Am. Chem. Soc.* **140**, 200–210 (2018).
 28. Ihle, N. T. *et al.* Molecular pharmacology and antitumor activity of PX-866, a novel inhibitor of phosphoinositide-3-kinase signaling. *Mol. Cancer Ther.* **3**, 763–772 (2004).
 29. Anscombe, E. *et al.* Identification and Characterization of an Irreversible Inhibitor of CDK2. *Chem. Biol.* **22**, 1–6 (2015).
 30. Manglik, A. *et al.* Crystal structure of the μ -opioid receptor bound to a morphinan

- antagonist. *Nature* **485**, 321–6 (2012).
31. Suh, E. H. *et al.* Stilbene vinyl sulfonamides as fluorogenic sensors of and traceless covalent kinetic stabilizers of transthyretin that prevent amyloidogenesis. *J. Am. Chem. Soc.* **135**, 17869–17880 (2013).
 32. Pettinger, J. *et al.* An Irreversible Inhibitor of HSP72 that Unexpectedly Targets Lysine-56. *Angew. Chemie Int. Ed.* **56**, 3536–3540 (2017).
 33. Mohamed, M. S., Larson, D. L., Takemori, A. E. & Portoghese, P. S. Activity of N-Methyl-alpha- and -beta-funaltrexamine at Opioid Receptors. *J. Med. Chem.* **29**, 1551–1553 (1986).
 34. Shannon, D. A. *et al.* Investigating the proteome reactivity and selectivity of aryl halides. *J. Am. Chem. Soc.* **136**, 3330–3333 (2014).
 35. Choi, S., Connelly, S., Reixach, N., Wilson, I. A. & Kelly, J. W. Chemoselective small molecules that covalently modify one lysine in a non-enzyme protein in plasma. *Nat. Chem. Biol.* **6**, 133–139 (2010).
 36. Tamura, T. *et al.* Rapid labelling and covalent inhibition of intracellular native proteins using ligand-directed N-Acyl-N-Alkyl sulfonamide. *Nat. Commun.* **9**, 1–12 (2018).
 37. Dalton, S. E. *et al.* Selectively Targeting the Kinome-Conserved Lysine of PI3K δ as a General Approach to Covalent Kinase Inhibition. *J. Am. Chem. Soc.* **140**, 932–939 (2018).
 38. Yasueda, Y. *et al.* A Set of Organelle-Localizable Reactive Molecules for Mitochondrial Chemical Proteomics in Living Cells and Brain Tissues. *J. Am. Chem. Soc.* **138**, 7592–7602 (2016).
 39. Patricelli, M. P. *et al.* Functional interrogation of the kinome using nucleotide acyl

- phosphates. *Biochemistry* **46**, 350–358 (2007).
40. Patricelli, M. P. *et al.* In situ kinase profiling reveals functionally relevant properties of native kinases. *Chem. Biol.* **18**, 699–710 (2011).
 41. Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181–1190 (2017).
 42. Ward, C. C., Kleinman, J. I. & Nomura, D. K. NHS-Esters As Versatile Reactivity-Based Probes for Mapping Proteome-Wide Ligandable Hotspots. *ACS Chem. Biol.* **12**, 1478–1483 (2017).
 43. Metcalf, B. *et al.* Discovery of GBT440, an Orally Bioavailable R-State Stabilizer of Sick Cell Hemoglobin. *ACS Med. Chem. Lett.* **8**, 321–326 (2017).
 44. Cross, B. C. S. *et al.* The molecular basis for selective inhibition of unconventional mRNA splicing by an IRE1-binding small molecule. *Proc. Natl. Acad. Sci.* **109**, E869–E878 (2012).
 45. Sanches, M. *et al.* Structure and mechanism of action of the hydroxy-aryl-aldehyde class of IRE1 endoribonuclease inhibitors. *Nat. Commun.* **5**, 1–16 (2014).
 46. Cal, P. M. S. D. *et al.* Iminoboronates: A new strategy for reversible protein modification. *J. Am. Chem. Soc.* **134**, 10299–10305 (2012).
 47. Akçay, G. *et al.* Inhibition of Mcl-1 through covalent modification of a noncatalytic lysine side chain. *Nat. Chem. Biol.* **12**, 931–936 (2016).
 48. Pettinger, J., Jones, K. & Cheeseman, M. D. Lysine-targeting covalent inhibitors. *Angewandte Chemie International Edition* **56**, 15200–15209 (2017).
 49. Aksnes, H., Drazic, A., Marie, M. & Arnesen, T. First Things First: Vital Protein Marks by N-Terminal Acetyltransferases. *Trends Biochem. Sci.* **41**, 746–760 (2016).

50. Vu, L. P. *et al.* The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat. Med.* **23**, 1369–1376 (2017).
51. Wang, X. *et al.* Structural basis of N6-adenosine methylation by the METTL3-METTL14 complex. *Nature* **534**, 575–578 (2016).
52. Isom, D. G., Castañeda, C. A., Cannon, B. R. & García-Moreno E, B. Large shifts in pKa values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci.* **108**, 5260–5265 (2011).
53. Baeza, J., Smallegan, M. J. & Denu, J. M. Site-specific reactivity of nonenzymatic lysine acetylation. *ACS Chem. Biol.* **10**, 122–128 (2015).
54. Schneekloth, J. S. *et al.* Chemical Genetic Control of Protein Levels: Selective in Vivo Targeted Degradation. *J. Am. Chem. Soc.* **126**, 3748–3754 (2004).
55. Sakamoto, K. M. *et al.* Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl. Acad. Sci.* **98**, 8554–9 (2001).
56. Winter, G. E. *et al.* Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **348**, 1376–1381 (2015).
57. Lai, A. C. & Crews, C. M. Induced protein degradation: an emerging drug discovery paradigm. *Nat. Rev. Drug Discov.* **16**, 101–114 (2016).
58. Smith, G. A., Uchida, K., Weiss, A. & Taunton, J. Essential biphasic role for JAK3 catalytic activity in IL-2 receptor signaling. *Nat. Chem. Biol.* **12**, 373–379 (2016).
59. Bantscheff, M. *et al.* Quantitative chemical proteomics reveals mechanisms of action of

- clinical ABL kinase inhibitors. *Nat. Biotechnol.* **25**, 1035–1044 (2007).
60. Savitski, M. F. M. M. F. M. *et al.* Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome. *Science* **346**, 1255784 (2014).
 61. Xie, T. *et al.* Pharmacological targeting of the pseudokinase Her3. *Nat. Chem. Biol.* **10**, 1006–1012 (2014).
 62. Zhang, T. *et al.* Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat. Chem. Biol.* **12**, 876–884 (2016).
 63. Kwiatkowski, N. *et al.* Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616–620 (2014).
 64. Tan, L. *et al.* Discovery of type II inhibitors of $\text{tgf}\beta$ -activated kinase 1 (TAK1) and mitogen-activated protein kinase kinase kinase kinase 2 (MAP4K2). *J. Med. Chem.* **58**, 183–196 (2015).
 65. Matthews, J. M. *et al.* Pathophysiological significance and therapeutic targeting of germinal center kinase in diffuse large B-cell lymphoma. *Blood* **128**, 239–248 (2016).
 66. Yang, Q. *et al.* Pharmacological inhibition of BMK1 suppresses tumor growth through promyelocytic leukemia protein. *Cancer Cell* **18**, 258–267 (2010).
 67. Ronan, B. *et al.* A highly potent and selective Vps34 inhibitor alters vesicle trafficking and autophagy. *Nat. Chem. Biol.* **10**, 1013–1019 (2014).
 68. Xiao, Y., Guo, L. & Wang, Y. A Targeted Quantitative Proteomics Strategy for Global Kinome Profiling of Cancer Cells and Tissues. *Mol. Cell. Proteomics* **13**, 1065–1075 (2014).
 69. Worboys, J. D., Sinclair, J., Yuan, Y. & Jørgensen, C. Systematic evaluation of quantotypic peptides for targeted analysis of the human kinome. *Nat. Methods* **11**, 1041–1044

- (2014).
70. McAllister, F. E. *et al.* Mass spectrometry based method to increase throughput for kinome analyses using ATP probes. *Anal. Chem.* **85**, 4666–4674 (2013).
 71. Nordin, B. E. *et al.* ATP Acyl phosphate reactivity reveals native conformations of Hsp90 paralogs and inhibitor target engagement. *Biochemistry* **54**, 3024–3036 (2015).
 72. Gardner Swain, C. & Scott, C. B. Rates of Solvolysis of Some Alkyl Fluorides and Chlorides. *J. Am. Chem. Soc.* **75**, 246–248 (1953).
 73. Dong, J., Krasnova, L., Finn, M. G. & Sharpless, K. B. Sulfur(VI) Fluoride Exchange (SuFEx): Another Good Reaction for Click Chemistry. *Angew. Chemie Int. Ed.* **53**, 9430–9448 (2014).
 74. Narayanan, A. & Jones, L. H. Sulfonyl fluorides as privileged warheads in chemical biology. *Chem. Sci.* **6**, 2650–2659 (2015).
 75. Ciuffarin, E., Senatore, L. & Isola, M. Nucleophilic substitution at four coordinate sulphur. Mobility of the leaving group. *J. Chem. Soc. Perkin Trans. 2* 468–471 (1972).
 76. Aberlin, M. E. & Bunton, C. A. The Spontaneous Hydrolysis of Sulfonyl Fluorides. *J. Org. Chem.* **35**, 1825–1828 (1970).
 77. Mukherjee, H. *et al.* A study of the reactivity of S(VI)-F containing warheads with nucleophilic amino-acid side chains under physiological conditions. *Org. Biomol. Chem.* **15**, 9685–9695 (2017).
 78. Zoller, M. J. & Taylor, S. S. Affinity Labeling of the Nucleotide Binding Site of the Catalytic Subunit of CAMP-dependent Protein Kinase Site Using p-fluorosulfonyl-[¹⁴C]benzoyl 5'Adenosine. *J. Biol. Chem.* **254**, 8363–8368 (1979).

79. Hett, E. C. *et al.* Rational targeting of active-site tyrosine residues using sulfonyl fluoride probes. *ACS Chem. Biol.* **10**, 1094–1098 (2015).
80. Serafimova, I. M. *et al.* Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nat. Chem. Biol.* **8**, 471–476 (2012).
81. Miller, R. M., Paavilainen, V. O., Krishnan, S., Serafimova, I. M. & Taunton, J. Electrophilic fragment-based design of reversible covalent kinase inhibitors. *J. Am. Chem. Soc.* **135**, 5298–5301 (2013).
82. Krishnan, S. *et al.* Design of reversible, cysteine-targeted Michael acceptors guided by kinetic and computational analysis. *J. Am. Chem. Soc.* **136**, 12624–12630 (2014).
83. Bradshaw, J. M. *et al.* Prolonged and tunable residence time using reversible covalent kinase inhibitors. *Nat. Chem. Biol.* **11**, 525–531 (2015).
84. Ferrone, F. A. *et al.* GBT440 increases haemoglobin oxygen affinity, reduces sickling and prolongs RBC half-life in a murine model of sickle cell disease. *Br. J. Haematol.* **175**, 141–153 (2016).
85. Bruyneel, W., Charette, J. J. & De Hoffmann, E. Kinetics of Hydrolysis of Hydroxy and Methoxy Derivatives of N-Benzylidene-2-aminopropane. *J. Am. Chem. Soc.* **88**, 3808–3813 (1966).
86. Kovaříček, P. & Lehn, J. M. Merging constitutional and motional covalent dynamics in reversible imine formation and exchange processes. *J. Am. Chem. Soc.* **134**, 9446–9455 (2012).
87. Wireko, F. C. & Abraham, D. J. X-ray diffraction study of the binding of the antisickling agent 12C79 to human hemoglobin. *Proc. Natl. Acad. Sci.* **88**, 2209–2211 (1991).

88. Safo, M. K. *et al.* Structural basis for the potent antisickling effect of a novel class of five-membered heterocyclic aldehydic compounds. *J. Med. Chem.* **47**, 4665–4676 (2004).
89. Abdulmalik, O. *et al.* Crystallographic analysis of human hemoglobin elucidates the structural basis of the potent and dual antisickling activity of pyridyl derivatives of vanillin. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **D67**, 920–928 (2011).
90. Papandreou, I. *et al.* Identification of an Ire1alpha endonuclease specific inhibitor with cytotoxic activity against human multiple myeloma. *Blood* **117**, 1311–1314 (2011).
91. Volkman, K. *et al.* Potent and selective inhibitors of the inositol-requiring enzyme 1 endoribonuclease. *J. Biol. Chem.* **286**, 12743–12755 (2011).
92. Tomasio, S. M., Harding, H. P., Ron, D., Cross, B. C. S. & Bond, P. J. Selective inhibition of the unfolded protein response: targeting catalytic sites for Schiff base modification. *Mol. Biosyst.* **9**, 2408–16 (2013).
93. Gutiérrez-Moreno, N. J., Medrano, F. & Yatsimirsky, A. K. Schiff base formation and recognition of amino sugars, aminoglycosides and biological polyamines by 2-formyl phenylboronic acid in aqueous solution. *Org. Biomol. Chem.* **10**, 6960–6972 (2012).
94. Niphakis, M. J. & Cravatt, B. F. Enzyme inhibitor discovery by activity-based protein profiling. *Annu. Rev. Biochem.* **83**, 341–77 (2014).
95. Cohen, M. S., Zhang, C., Shokat, K. M. & Taunton, J. Structural Bioinformatics-Based Design of Selective, Irreversible Kinase Inhibitors. *Science (80-.)*. **308**, 1318–1321 (2005).
96. Bhagavat, R., Sankar, S., Srinivasan, N. & Chandra, N. An Augmented Pocketome: Detection and Analysis of Small-Molecule Binding Pockets in Proteins of Known 3D Structure. *Structure* **26**, 499-512.e2 (2018).

97. Hernandez, M., Ghersi, D. & Sanchez, R. SITEHOUND-web: A server for ligand binding site identification in protein structures. *Nucleic Acids Res.* **37**, 413–416 (2009).
98. Kalidas, Y. & Chandra, N. PocketDepth: A new depth based algorithm for identification of ligand binding sites in proteins. *J. Struct. Biol.* **161**, 31–42 (2008).
99. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 1–11 (2009).
100. Cuesta, A. & Taunton, J. Lysine-Targeted Inhibitors and Chemoproteomic Probes. *Annu. Rev. Biochem.* **88**, 1–17 (2019).
101. Barbarulo, A. *et al.* Poly(ADP-ribose) polymerase family member 14 (PARP14) is a novel effector of the JNK2-dependent pro-survival signal in multiple myeloma. *Oncogene* **32**, 4231–4242 (2013).
102. Iansante, V. *et al.* PARP14 promotes the Warburg effect in hepatocellular carcinoma by inhibiting JNK1-dependent PKM2 phosphorylation and activation. *Nat. Commun.* **6**, (2015).
103. Upton, K. *et al.* Design and synthesis of potent inhibitors of the mono(ADP-ribose)transferase, PARP14. *Bioorganic Med. Chem. Lett.* **27**, 2907–2911 (2017).
104. Otto, H. *et al.* In silico characterization of the family of PARP-like poly(ADP-ribose)transferases (pARTs). *BMC Genomics* **6**, 1–23 (2005).
105. Stevers, L. M. *et al.* Modulators of 14-3-3 Protein-Protein Interactions. *J. Med. Chem.* **61**, 3755–3778 (2018).
106. Morrison, D. K. The 14-3-3 proteins: integrators of diverse signaling cues that impact cell fate and cancer development. *Trends Cell Biol.* **19**, 16–23 (2009).

107. Sijbesma, E. *et al.* Site-Directed Fragment-Based Screening for the Discovery of Protein-Protein Interaction Stabilizers. *J. Am. Chem. Soc.* **141**, 3524–3531 (2019).
108. Oliveros, J. C. Venny. An interactive tool for comparing lists with Venn's diagrams. (2015). Available at: <http://bioinfogp.cnb.csic.es/tools/venny/>. (Accessed: 7th June 2019)
109. Martin, A. C. R. Mapping PDB chains to UniProtKB entries. **21**, 4297–4301 (2017).
110. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
111. Strelow, J. M. A Perspective on the Kinetics of Covalent and Irreversible Inhibition. *J. Biomol. Screen.* 1087057116671509 (2016). doi:10.1177/1087057116671509
112. Cheng, H. *et al.* Discovery of 1-((3R,4R)-3-[5-Chloro-2-(1-methyl-1H-pyrazol-4-ylamino)-7H-pyrrolo[2,3-d]pyrimidin-4-ylloxymethyl]-4-methoxy-pyrrolidin-1-yl)propenone (PF-06459988), A Potent, WT Sparing, Irreversible Inhibitor of T790M-Containing EGFR Mutants. *J. Med. Chem.* acs.jmedchem.5b01633 (2016). doi:10.1021/acs.jmedchem.5b01633
113. Liu, R., Yue, Z., Tsai, C. C. & Shen, J. Assessing Lysine and Cysteine Reactivities for Designing Targeted Covalent Kinase Inhibitors. *J. Am. Chem. Soc.* **141**, 6553–6560 (2019).
114. Immormino, R. M., Kang, Y., Chiosis, G. & Gewirth, D. T. Structural and quantum chemical studies of 8-aryl-sulfanyl adenine class Hsp90 inhibitors. *J. Med. Chem.* **49**, 4953–4960 (2006).
115. Taldone, T. *et al.* Experimental and structural testing module to analyze paralogue-specificity and affinity in the Hsp90 inhibitors series. *J. Med. Chem.* **56**, 6803–6818 (2013).

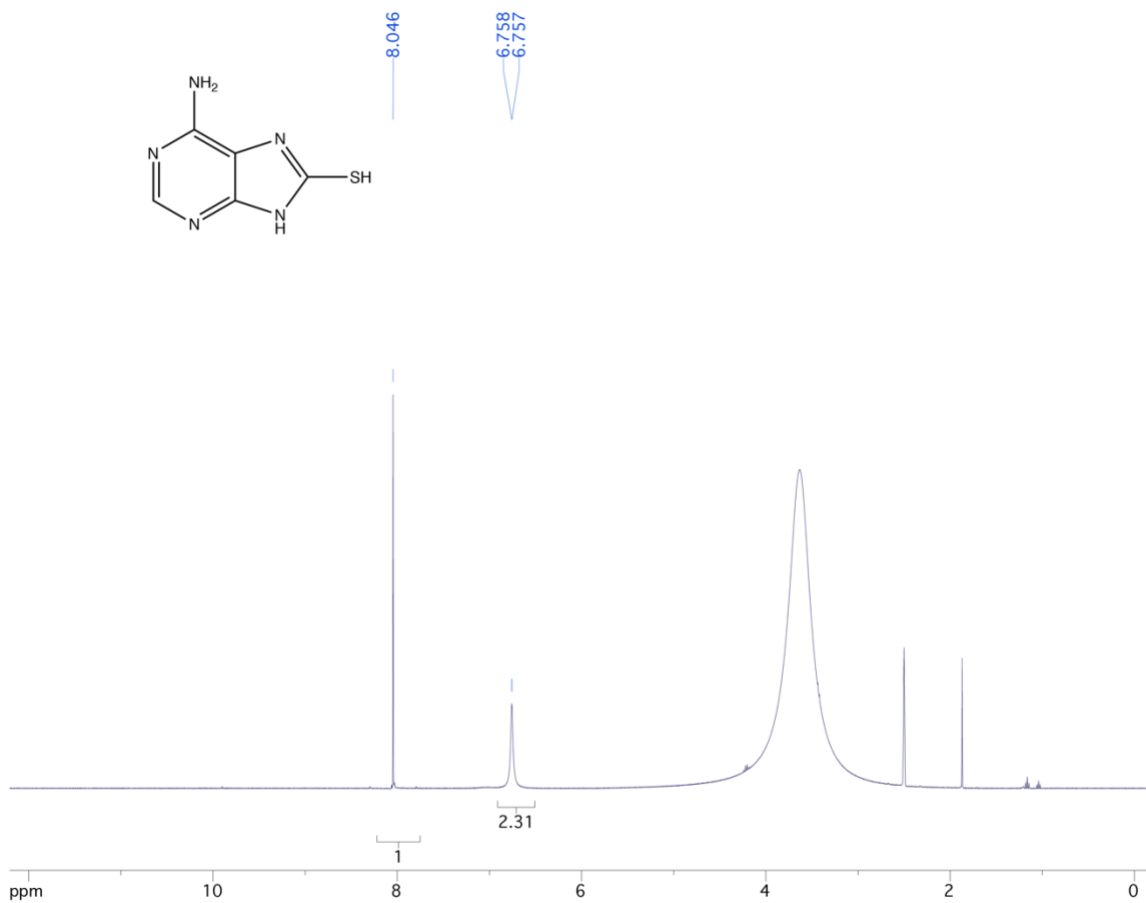
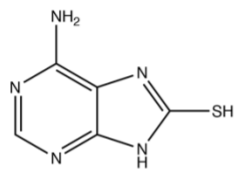
116. Kim, J. *et al.* Development of a fluorescence polarization assay for the molecular chaperone Hsp90. *J. Biomol. Screen. Off. J. Soc. Biomol. Screen.* **9**, 375–381 (2004).
117. Kirschke, E., Goswami, D., Southworth, D., Griffin, P. R. & Agard, D. A. Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. *Cell* **157**, 1685–1697 (2014).
118. Samant, R. S., Clarke, P. A. & Workman, P. E3 ubiquitin ligase Cullin-5 modulates multiple molecular and cellular responses to heat shock protein 90 inhibition in human cancer cells. *Proc. Natl. Acad. Sci.* **111**, 6834–6839 (2014).
119. Fabian, M. A. *et al.* A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **23**, 329–336 (2005).
120. Gambini, L. *et al.* Covalent Inhibitors of Protein-Protein Interactions Targeting Lysine, Tyrosine, or Histidine Residues. *J. Med. Chem.* acs.jmedchem.9b00561 (2019).
doi:10.1021/acs.jmedchem.9b00561
121. Knoepfel, T. *et al.* 2-Formylpyridyl Ureas as Highly Selective Reversible-Covalent Inhibitors of Fibroblast Growth Factor Receptor 4. *ACS Med. Chem. Lett.* **9**, 215–220 (2018).
122. Godoy-Alcántar, C., Yatsimirsky, A. K. & Lehn, J. M. Structure-stability correlations for imine formation in aqueous solution. *J. Phys. Org. Chem.* **18**, 979–985 (2005).
123. Willems, E. *et al.* The functional diversity of Aurora kinases: A comprehensive review. *Cell Div.* **13**, 1–17 (2018).
124. Gustafson, W. C. *et al.* Drugging MYCN through an Allosteric Transition in Aurora Kinase A. *Cancer Cell* **26**, 414–427 (2014).

125. Otto, T. *et al.* Stabilization of N-Myc Is a Critical Function of Aurora A in Human Neuroblastoma. *Cancer Cell* **15**, 67–78 (2009).
126. Richards, M. W. *et al.* Structural basis of N-Myc binding by Aurora-A and its destabilization by kinase inhibitors. *Proc. Natl. Acad. Sci.* **113**, 13726–13731 (2016).
127. Donnella, H. J. *et al.* Kinome rewiring reveals AURKA limits PI3K-pathway inhibitor efficacy in breast cancer. *Nat. Chem. Biol.* **14**, 768–777 (2018).
128. Loh, S. H. Y., Francescut, L., Lingor, P., Bähr, M. & Nicotera, P. Identification of new kinase clusters required for neurite outgrowth and retraction by a loss-of-function RNA interference screen. *Cell Death Differ.* **15**, 283–298 (2008).
129. Terrien, E. *et al.* Interference with the PTEN-MAST2 Interaction by a Viral Protein Leads to Cellular Relocalization of PTEN. *Sci. Signal.* **5**, ra58–ra58 (2012).
130. Delhommel, F. *et al.* Deciphering the unconventional peptide binding to the PDZ domain of MAST2. *Biochem. J.* **469**, 159–168 (2015).
131. Robinson, D. R. *et al.* Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nat. Med.* **17**, 1646–1651 (2011).
132. Andrade, E. C. *et al.* ARPP-16 Is a Striatal-Enriched Inhibitor of Protein Phosphatase 2A Regulated by Microtubule-Associated Serine/Threonine Kinase 3 (Mast 3 Kinase). *J. Neurosci.* **37**, 2709–2722 (2017).
133. Musante, V. *et al.* Reciprocal regulation of ARPP-16 by PKA and MAST3 kinases provides a cAMP-regulated switch in protein phosphatase 2A inhibition. *Elife* **6**, 1–32 (2017).
134. Bago, R. *et al.* The hVps34-SGK3 pathway alleviates sustained PI3K/Akt inhibition by stimulating mTORC1 and tumour growth. *EMBO J.* **35**, 2263–2263 (2016).

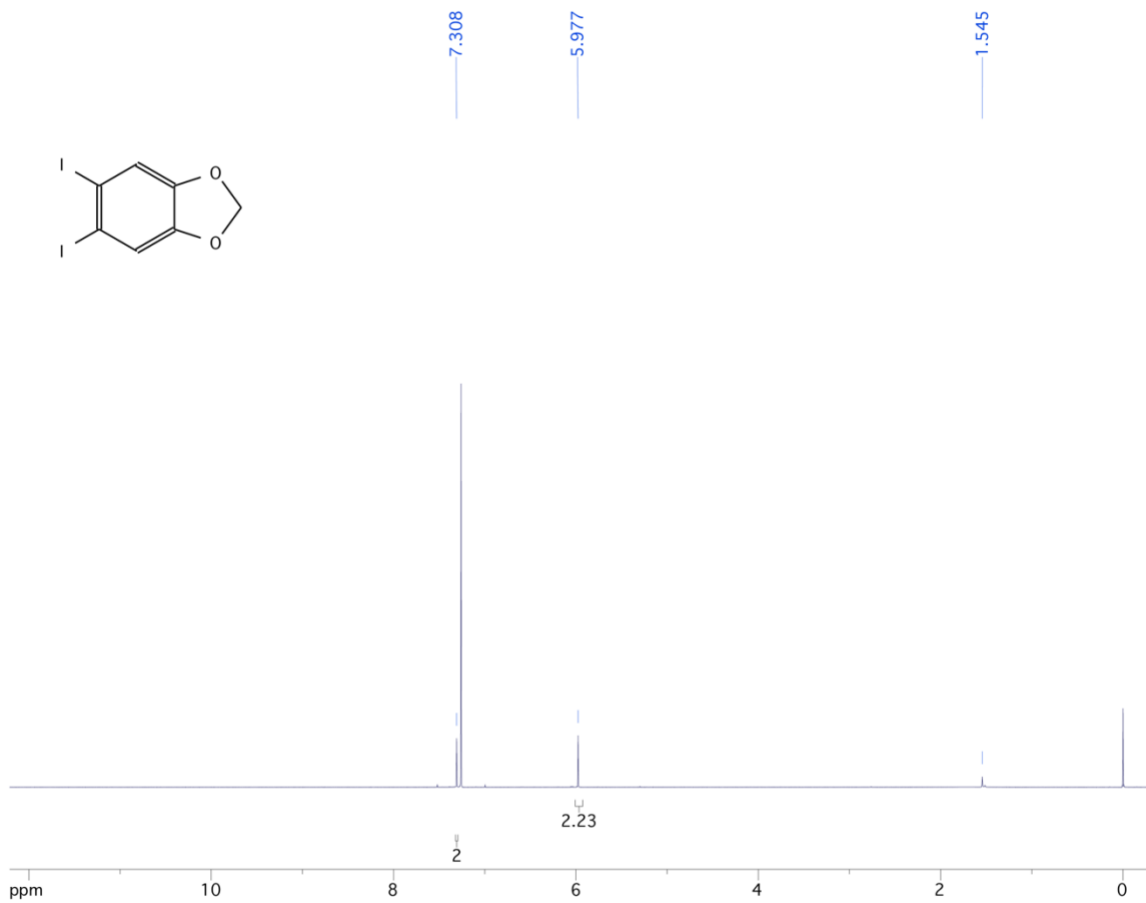
135. Malik, N. *et al.* Mechanism of activation of SGK3 by growth factors via the Class 1 and Class 3 PI3Ks. *Biochem. J.* **475**, 117–135 (2018).

Appendix A: **NMR spectra**

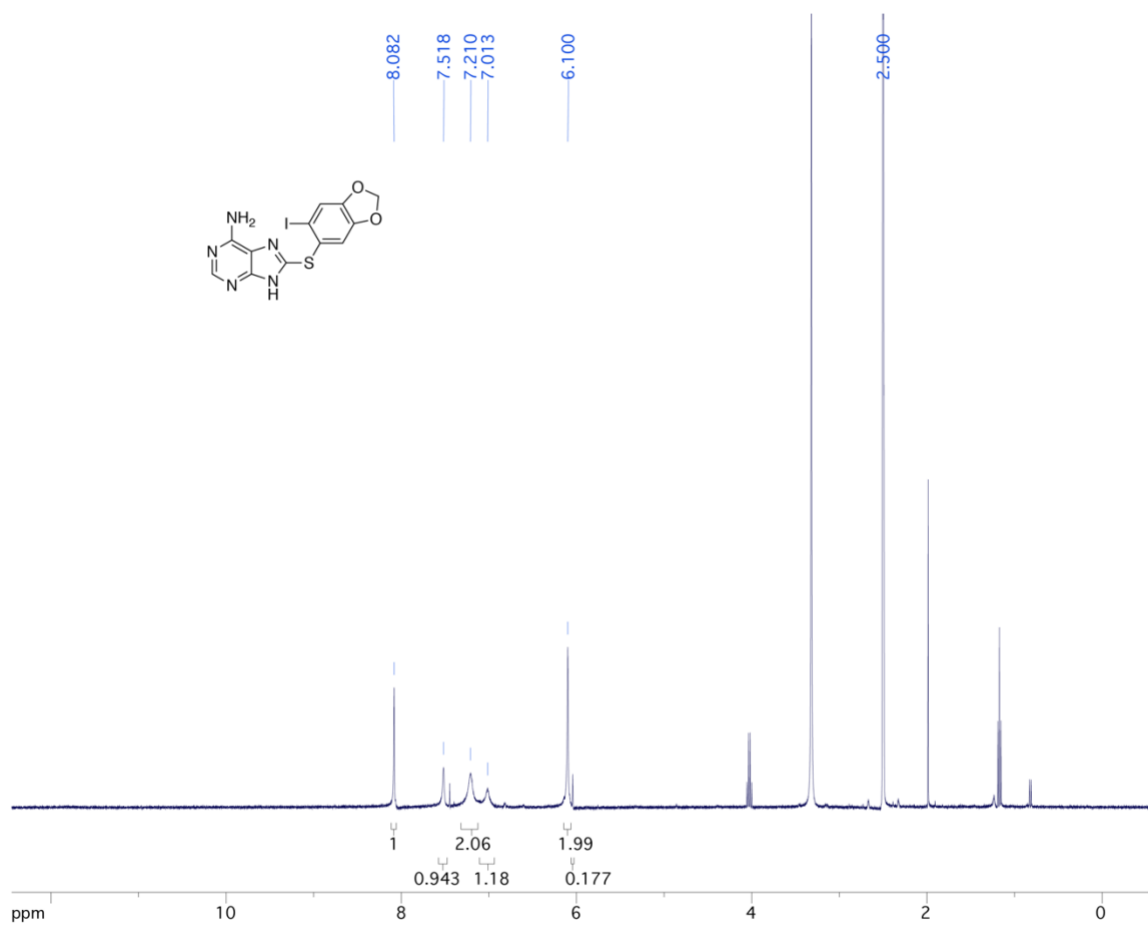
8-mercaptoadenine



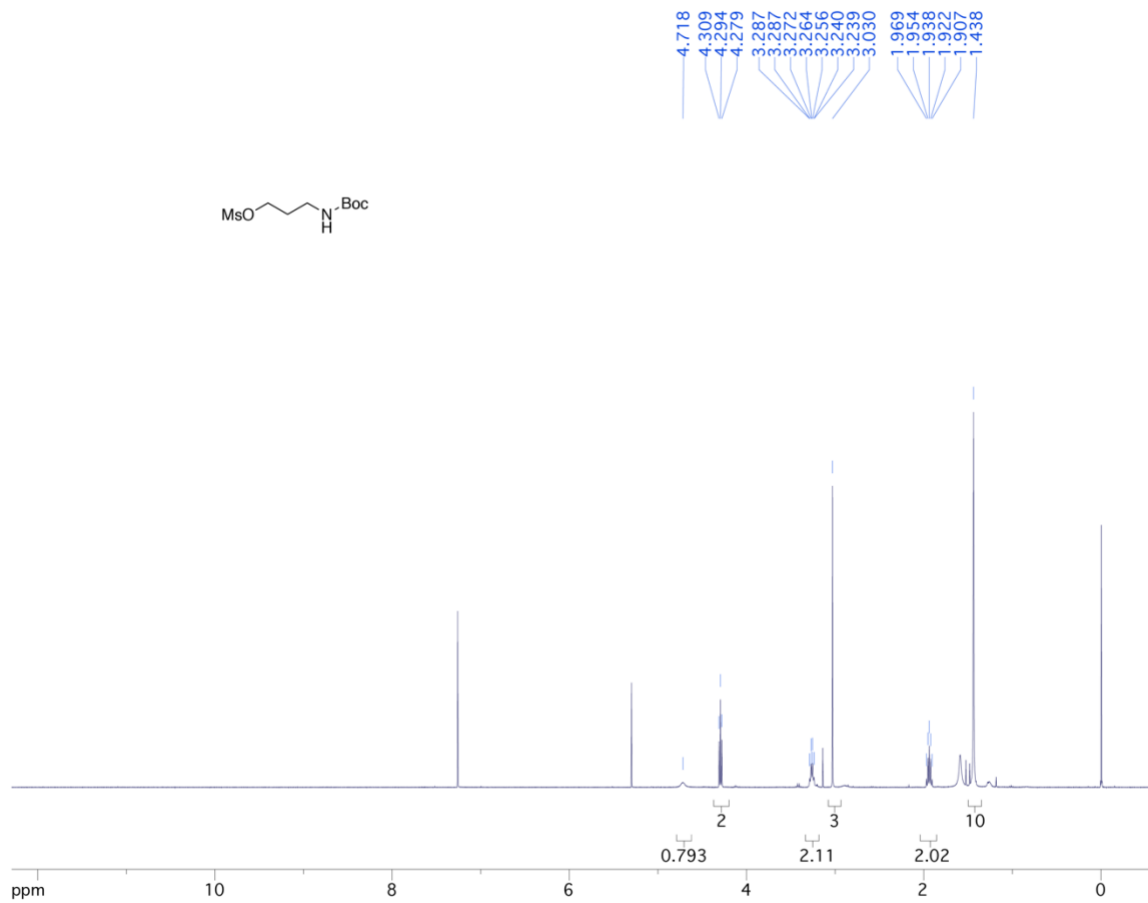
5,6-diiodobenzo[d][1,3]dioxole:



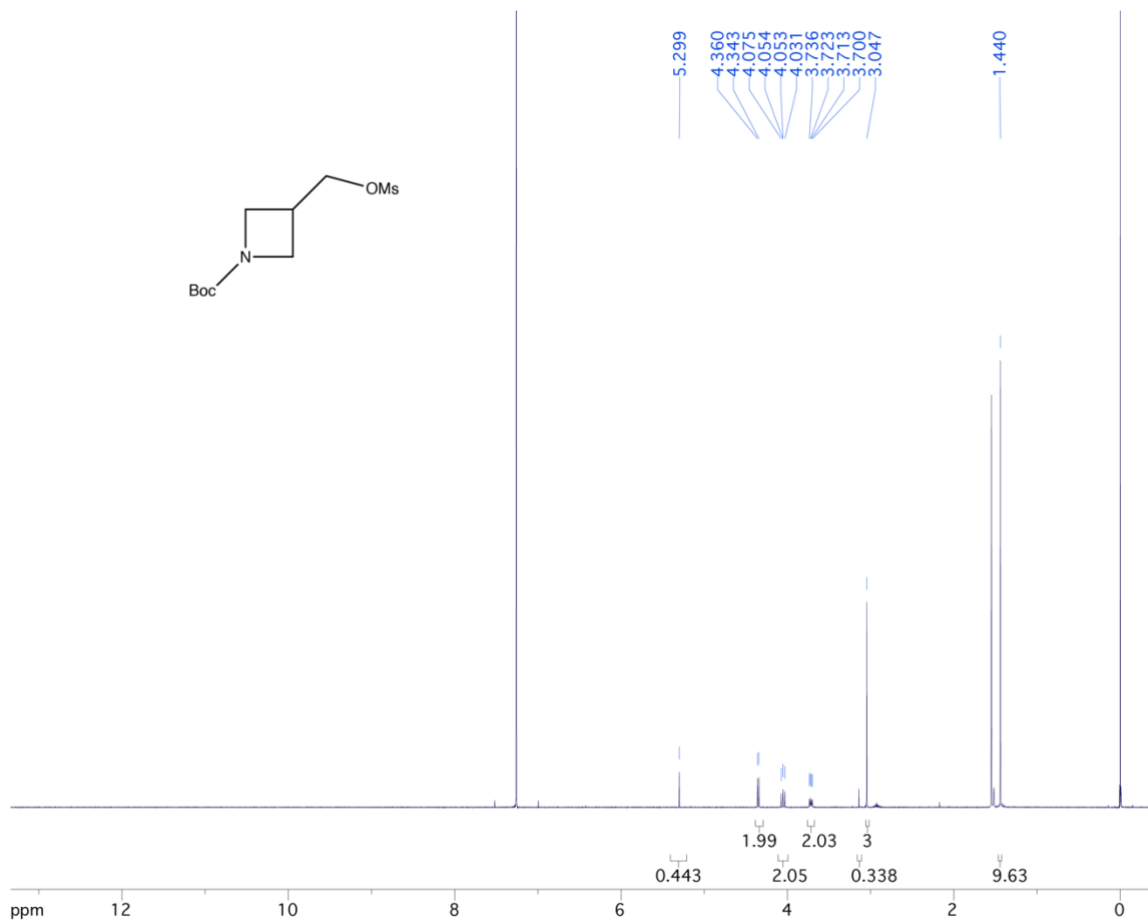
8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine



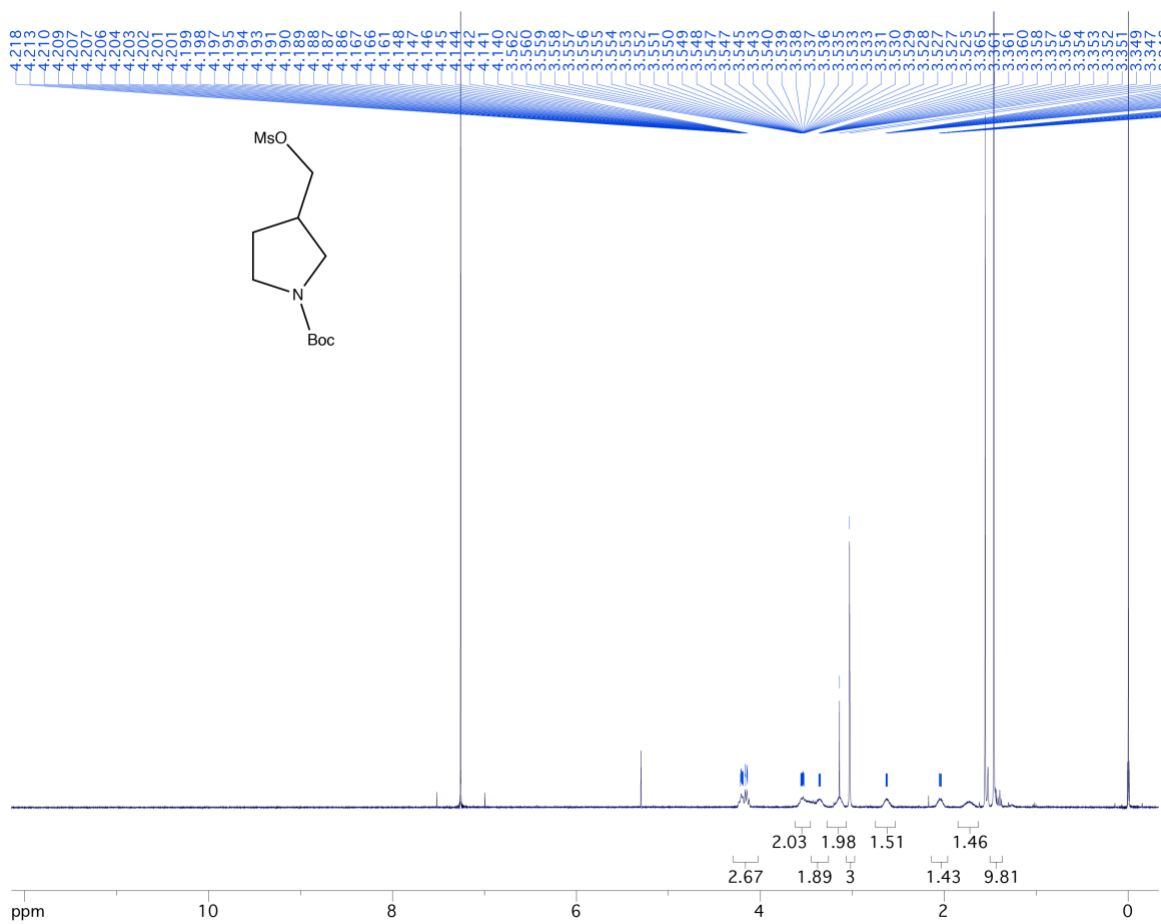
3-((tert-butoxycarbonyl)amino)propyl methanesulfonate:



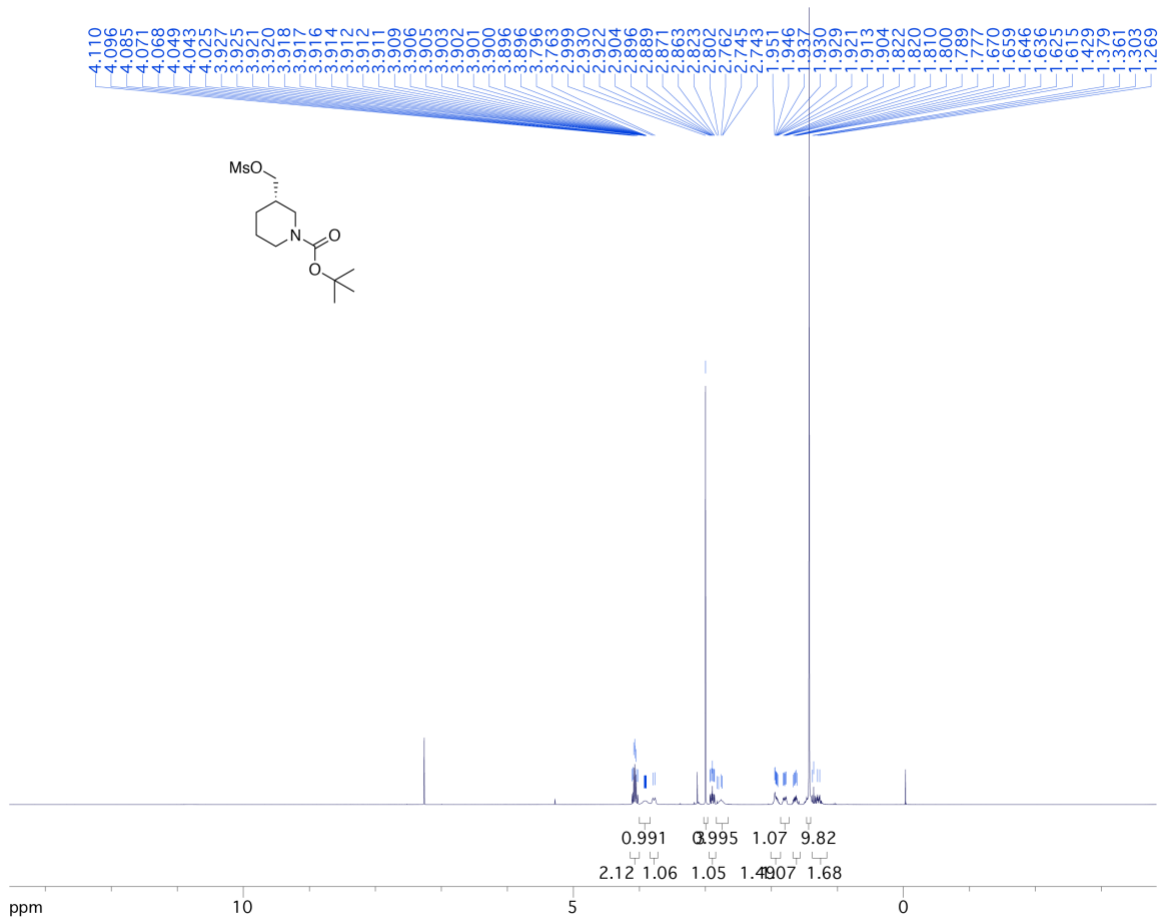
tert-butyl 3-(((methylsulfonyl)oxy)methyl)azetidine-1-carboxylate



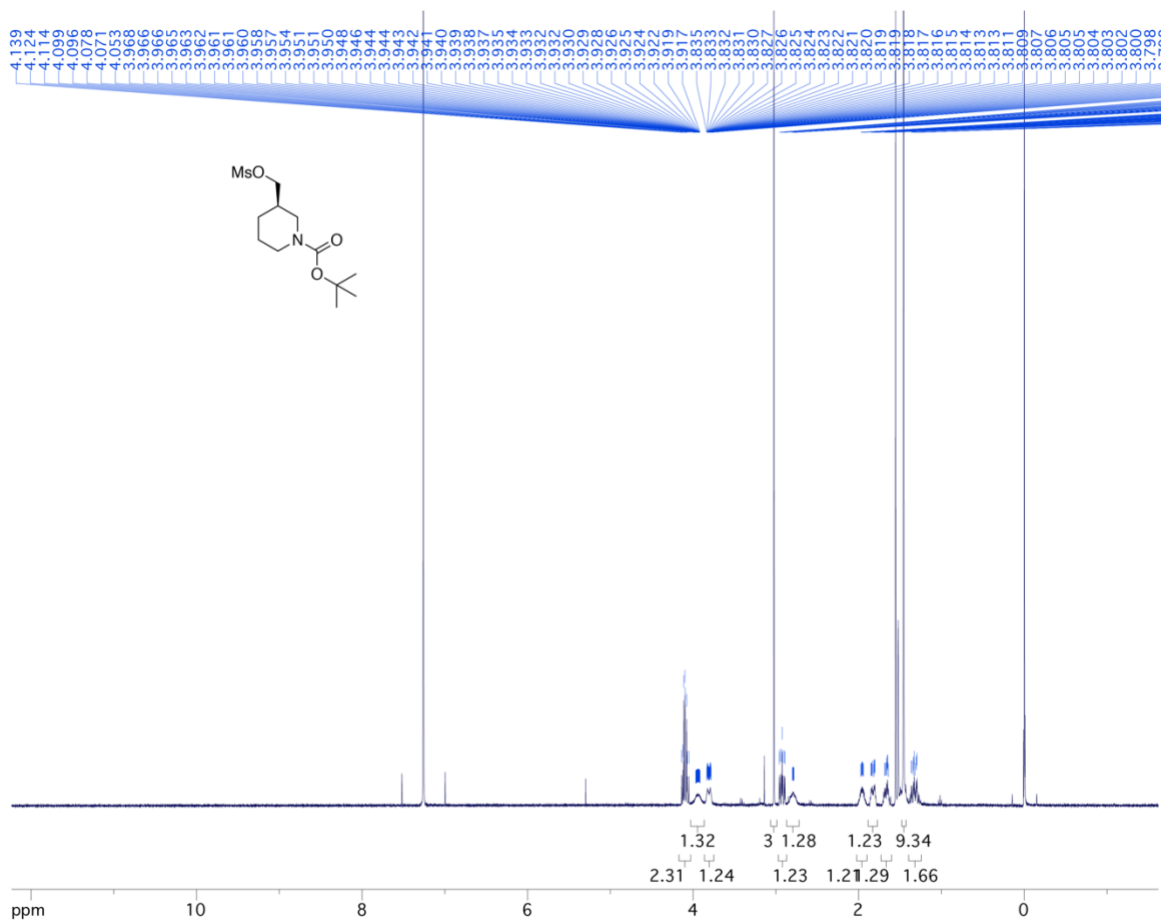
tert-butyl 3-(((methylsulfonyl)oxy)methyl)pyrrolidine-1-carboxylate



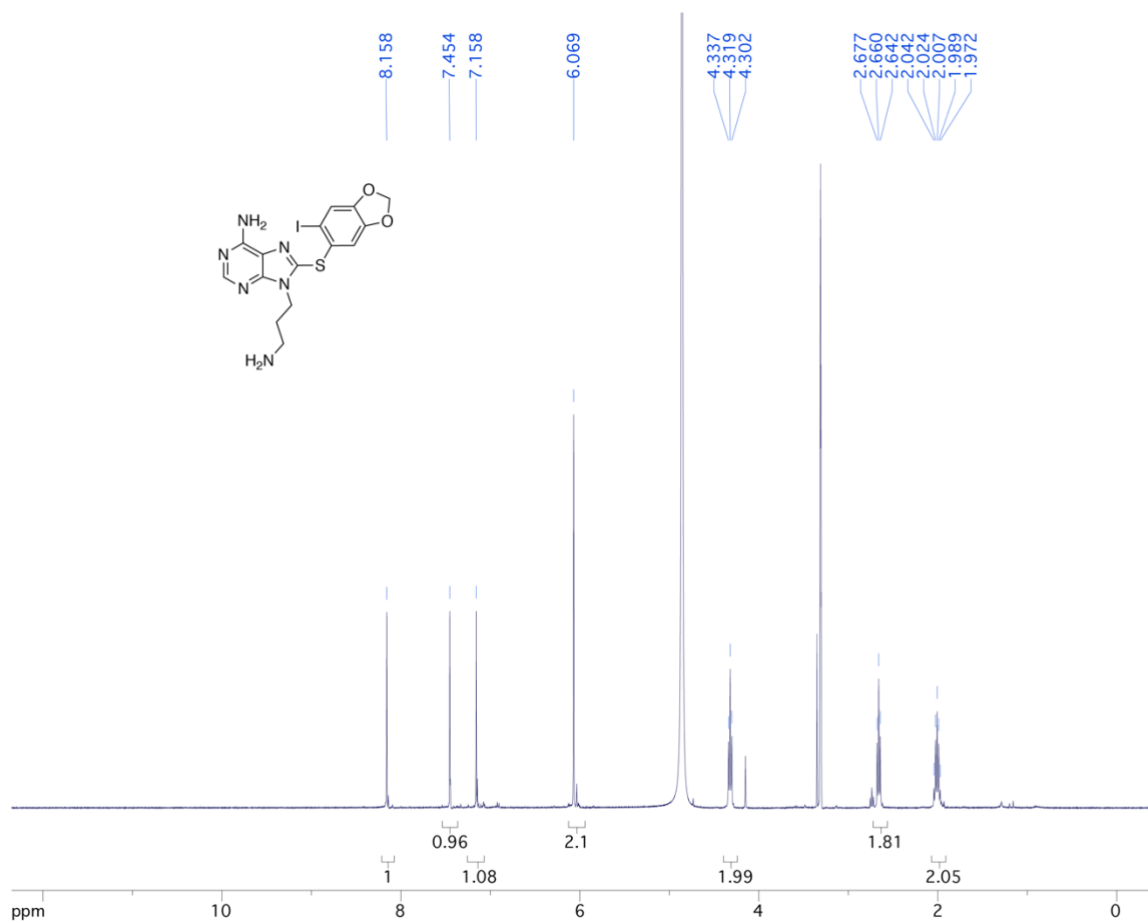
(S)-tert-butyl 3-(((methylsulfonyl)oxy)methyl)piperidine-1-carboxylate



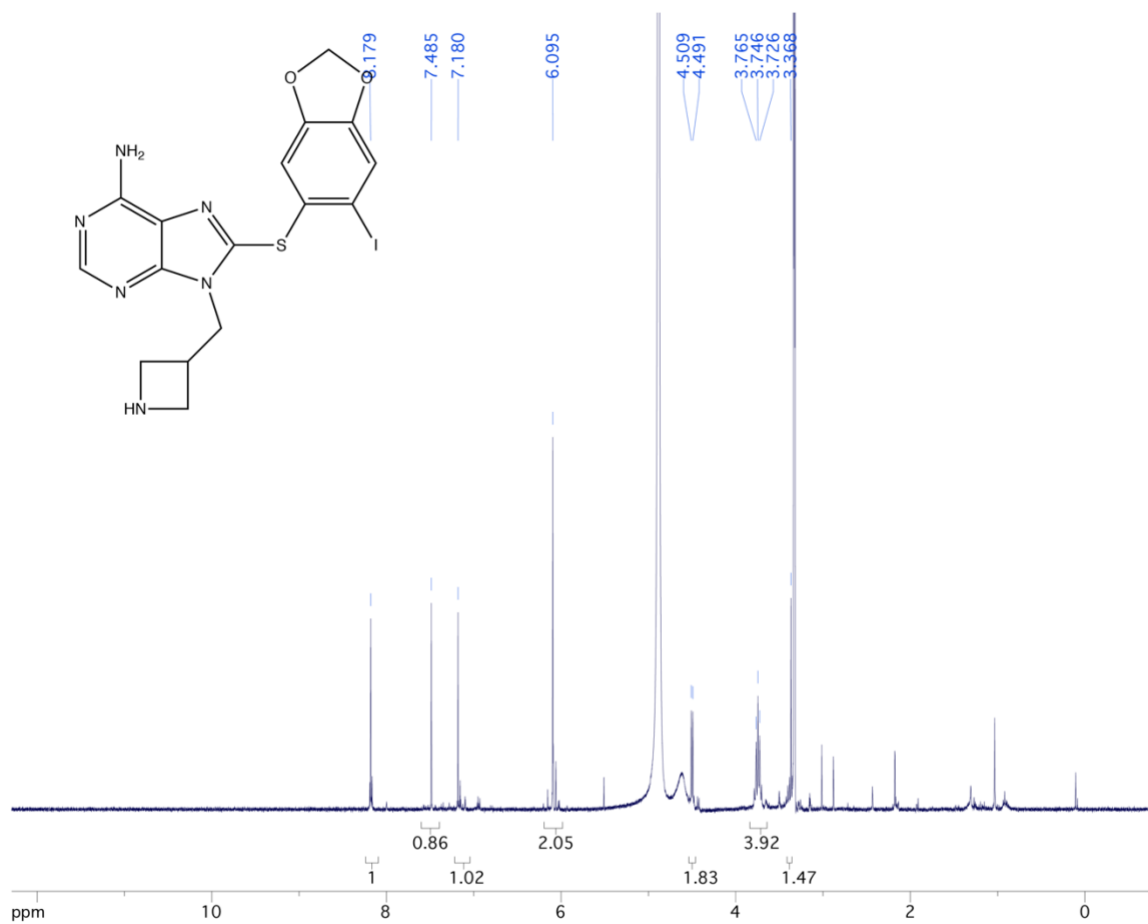
(R)-tert-butyl 3-(((methylsulfonyl)oxy)methyl)piperidine-1-carboxylate:



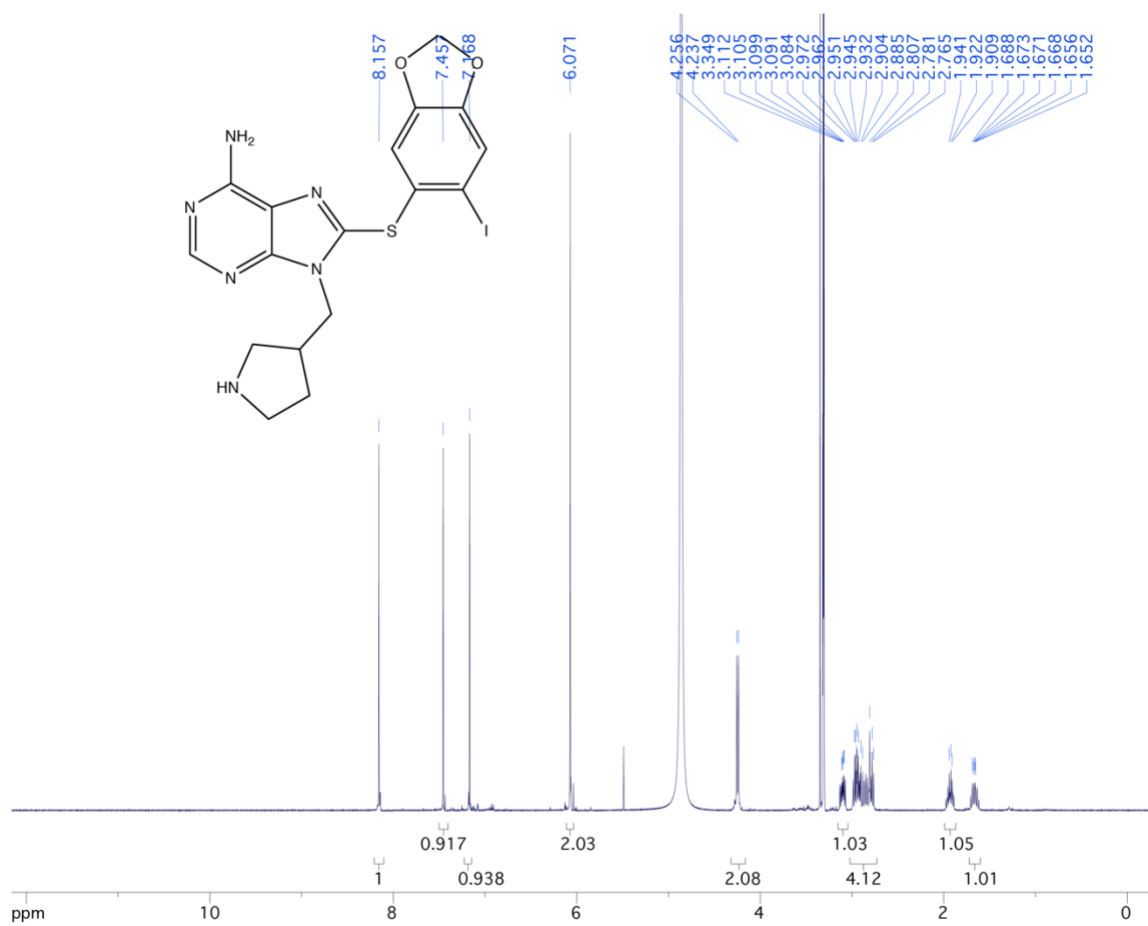
9-(3-aminopropyl)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine



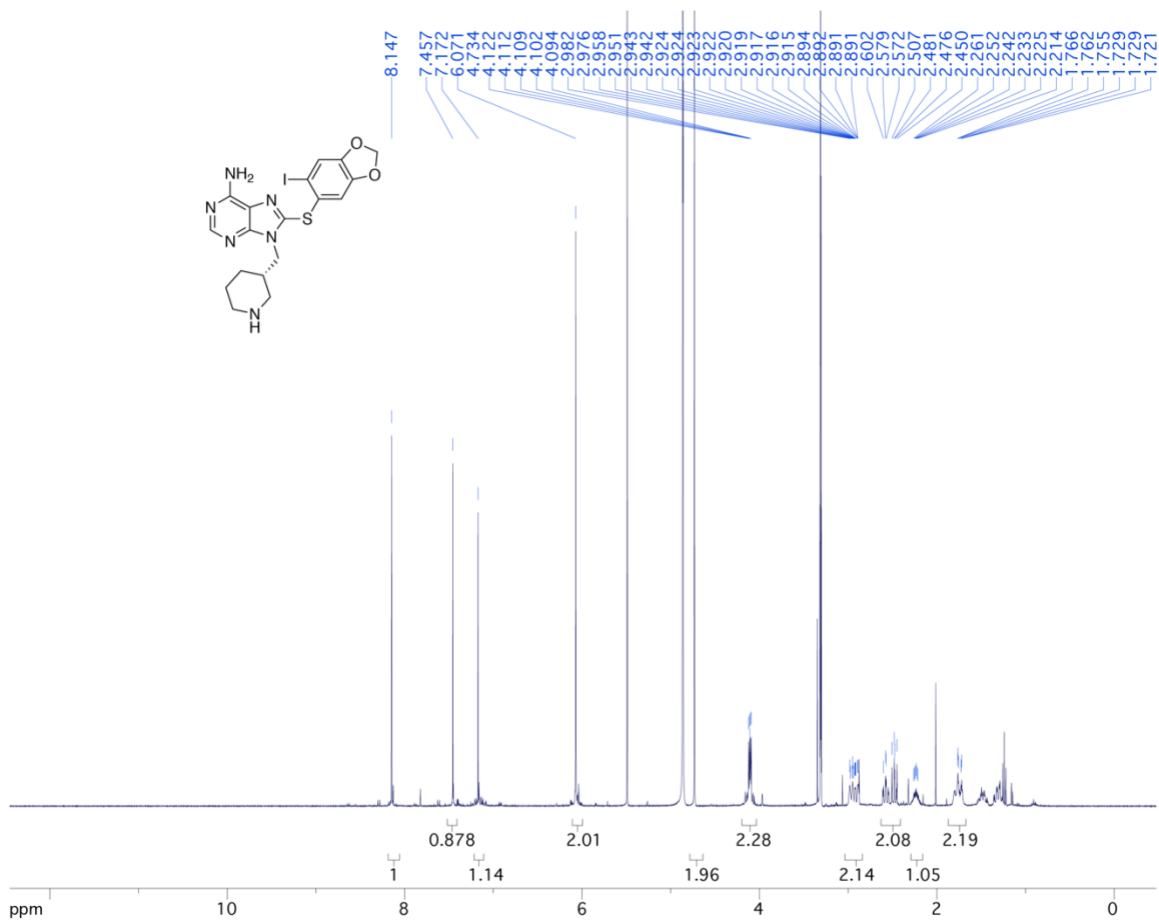
9-(azetidin-3-ylmethyl)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-6-amine:



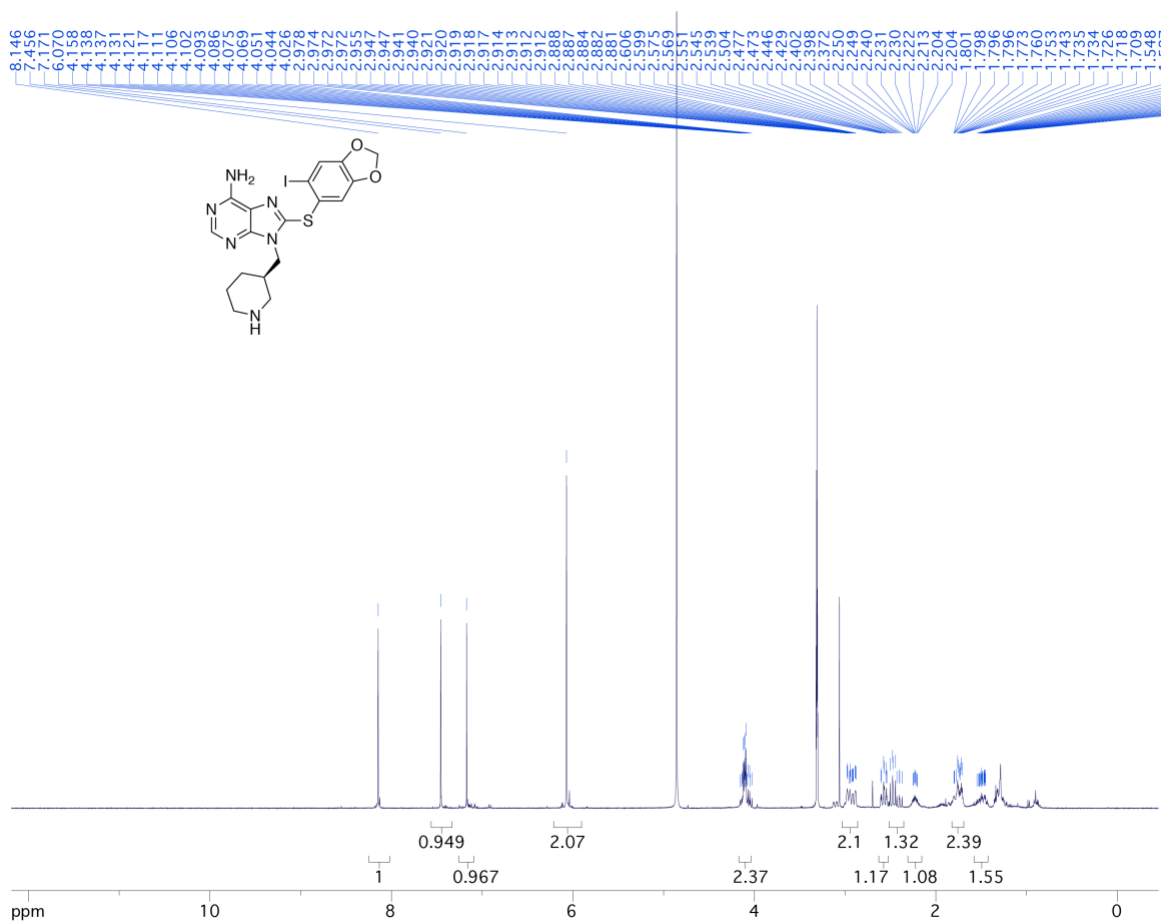
8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9-(pyrrolidin-3-ylmethyl)-9H-purin-6-amine:



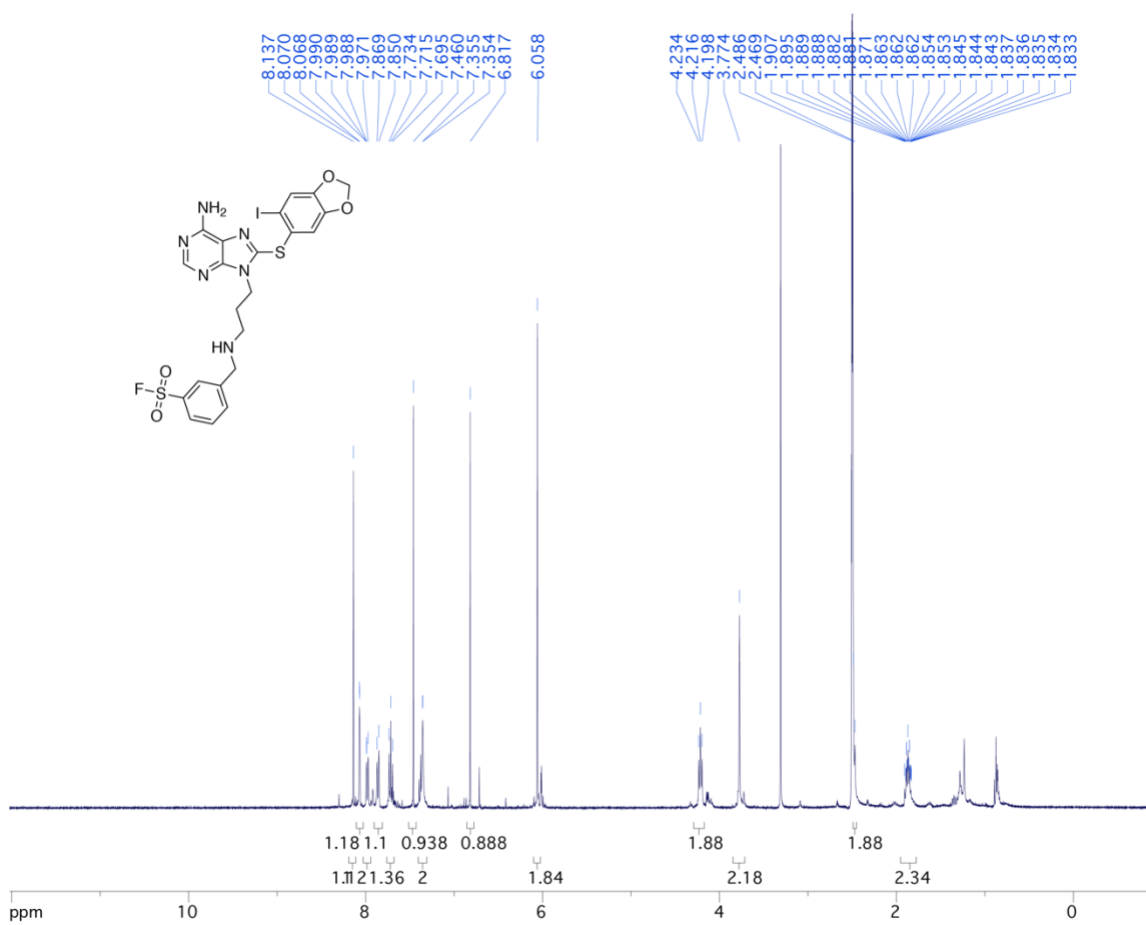
(S)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9-(piperidin-3-ylmethyl)-9H-purin-6-amine



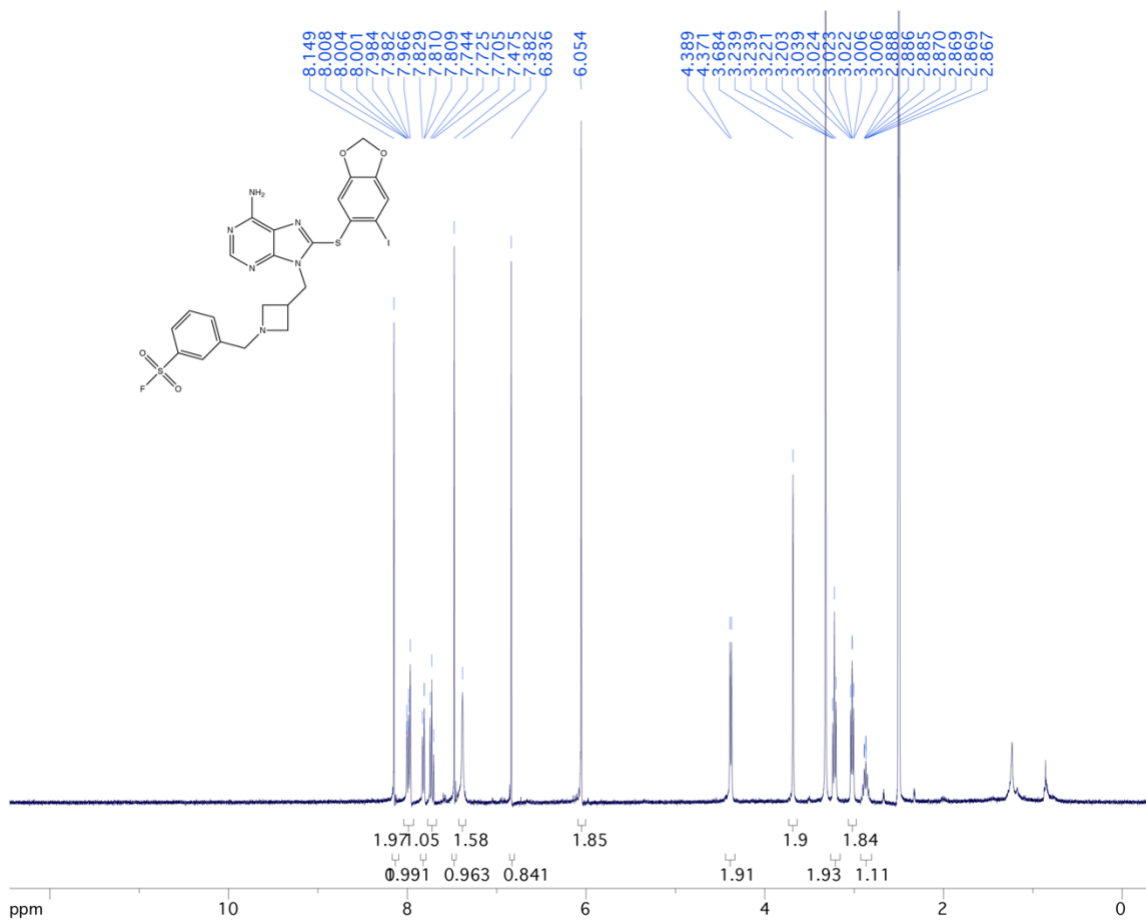
(R)-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9-(piperidin-3-ylmethyl)-9H-purin-6-amine:



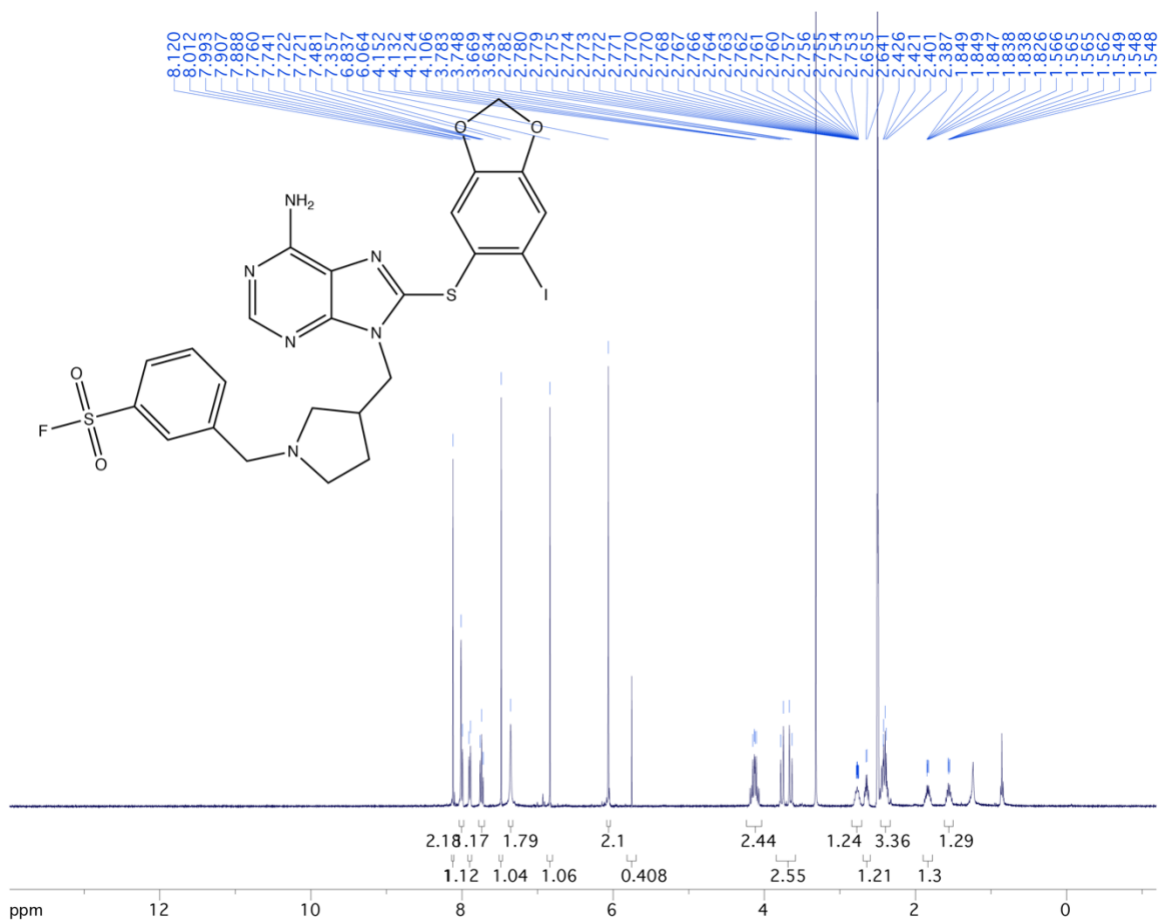
3-(((3-(6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-yl)propyl)amino)methyl)benzene-1-sulfonyl fluoride (1)



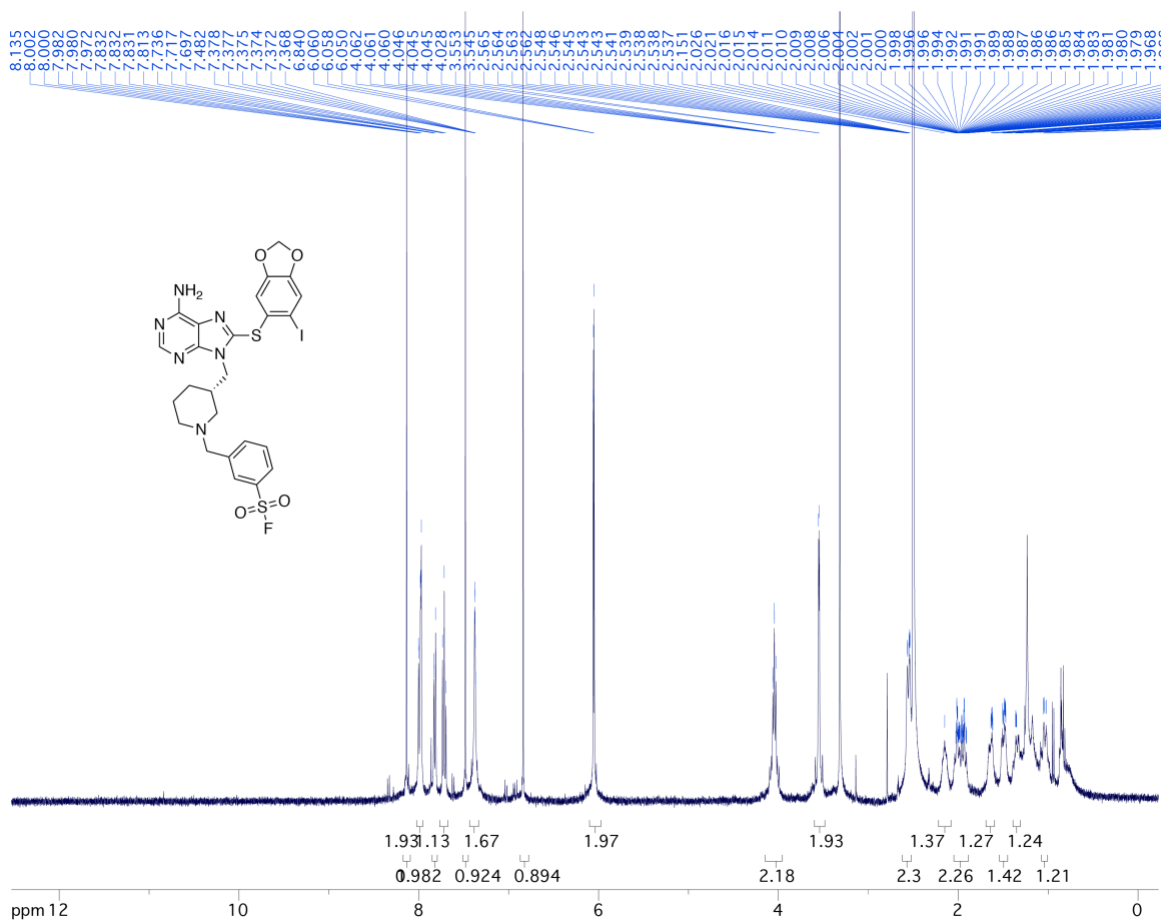
3-((3-((6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-yl)methyl)azetidin-1-yl)methyl)benzene-1-sulfonyl fluoride (3):



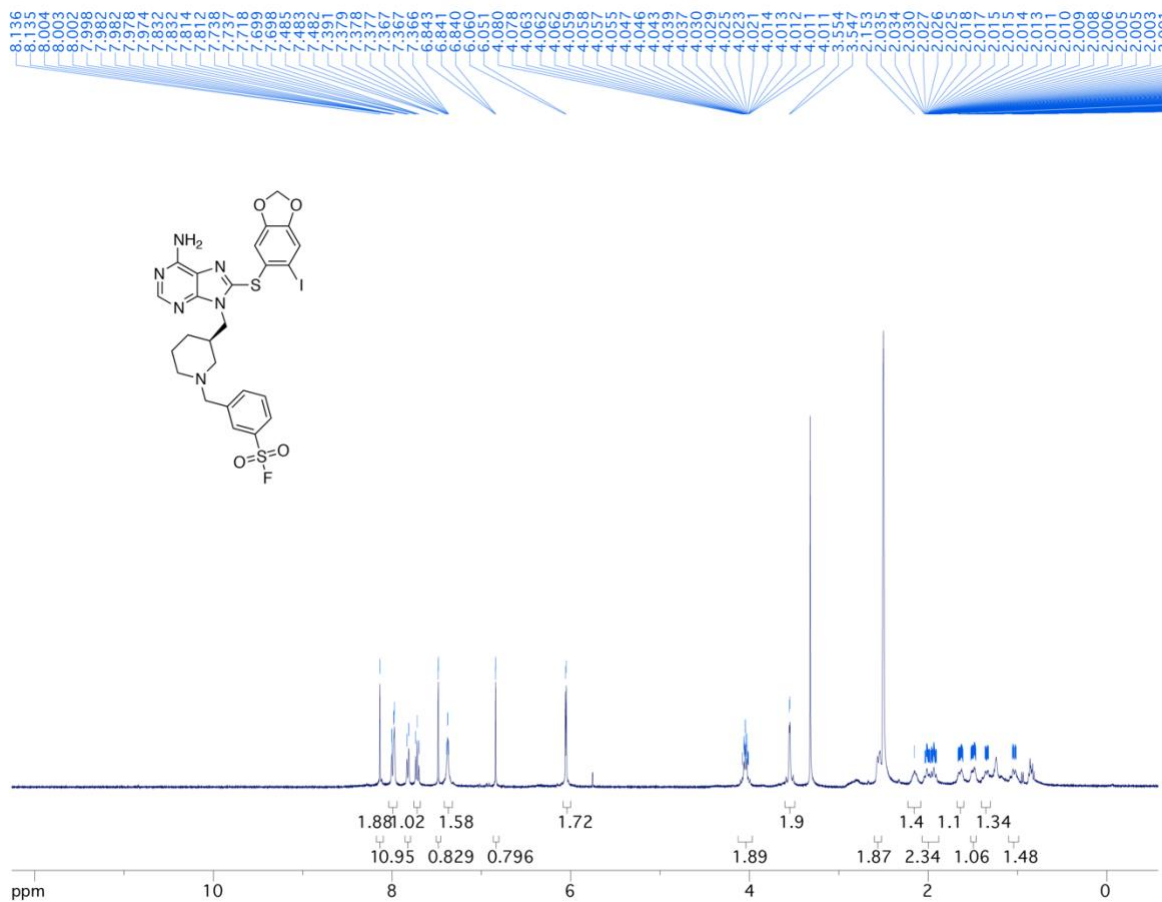
3-((3-((6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-yl)methyl)pyrrolidin-1-yl)methyl)benzene-1-sulfonyl fluoride (4):



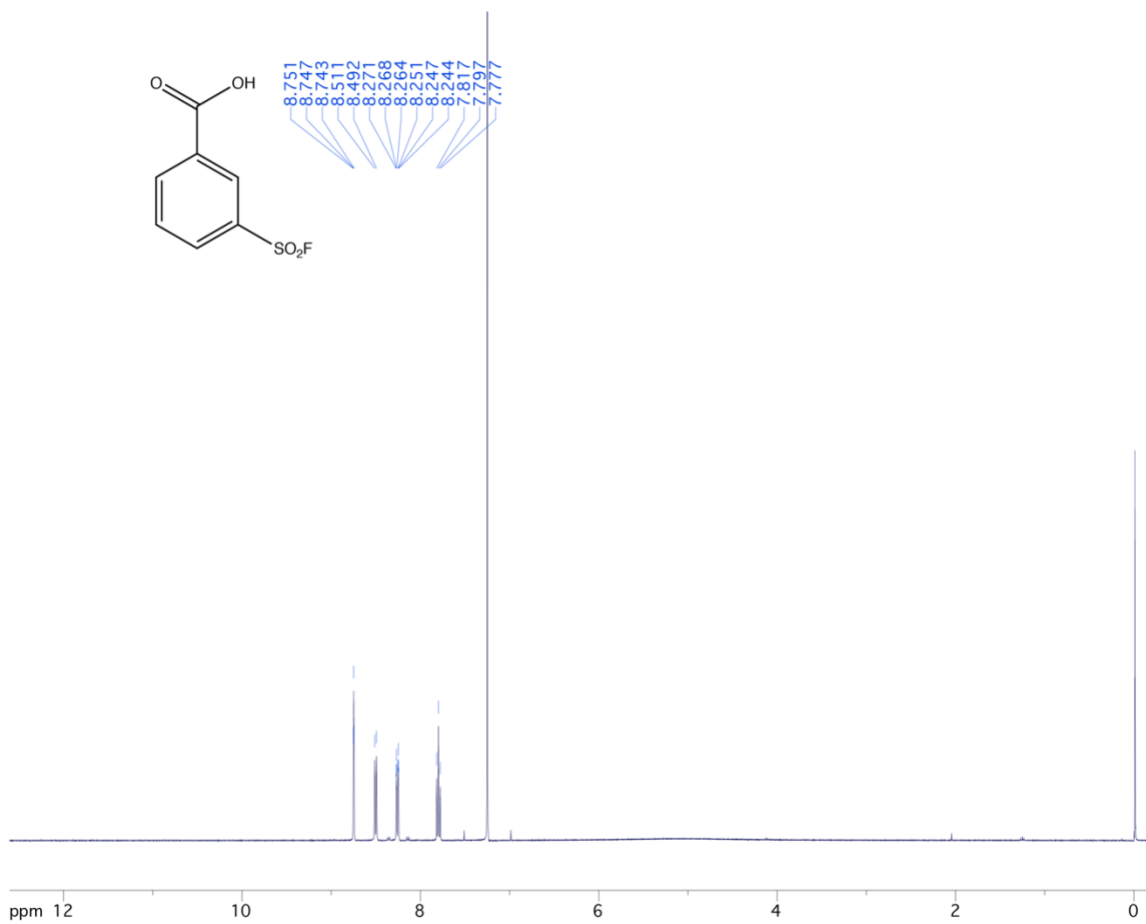
(S)-3-((3-((6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-yl)methyl)piperidin-1-yl)methyl)benzene-1-sulfonyl fluoride (5):



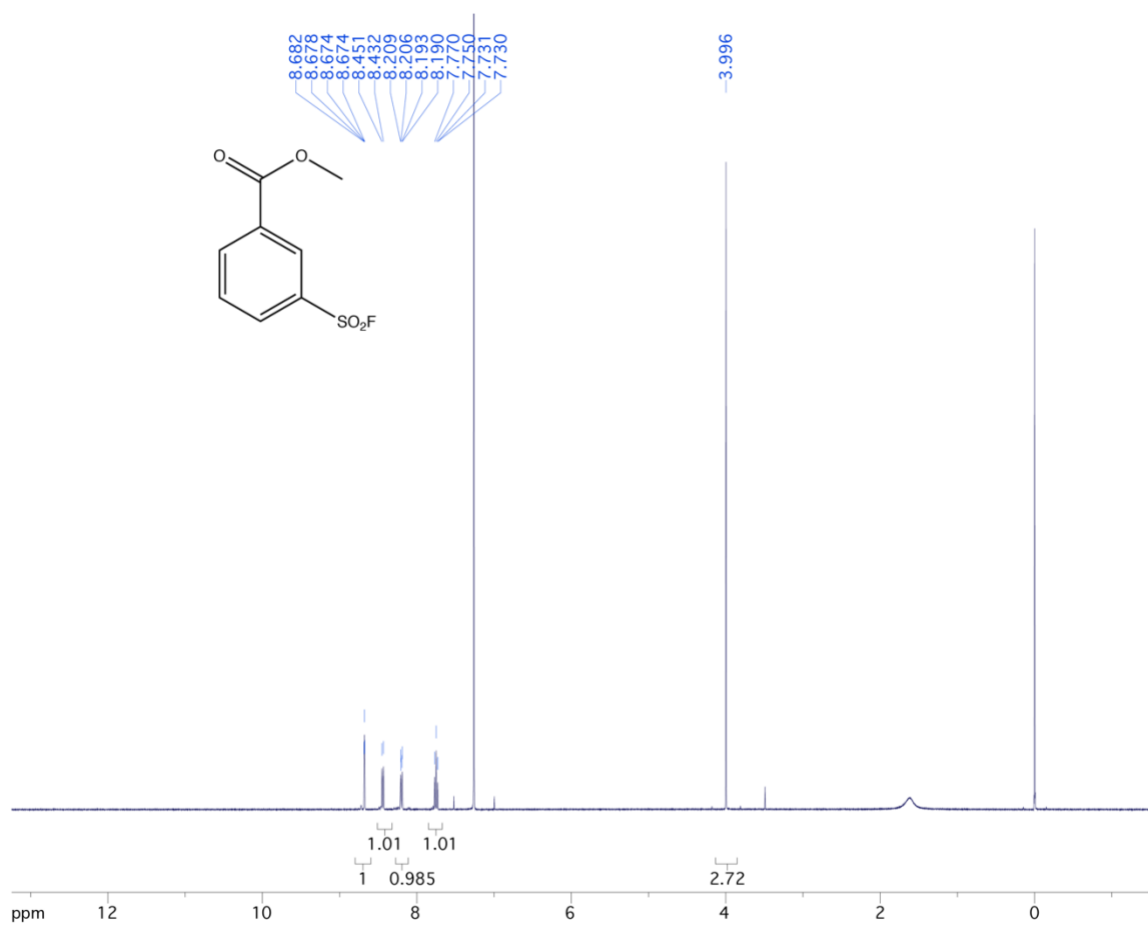
(R)-3-((3-((6-amino-8-((6-iodobenzo[d][1,3]dioxol-5-yl)thio)-9H-purin-9-yl)methyl)piperidin-1-yl)methyl)benzene-1-sulfonyl fluoride (6):



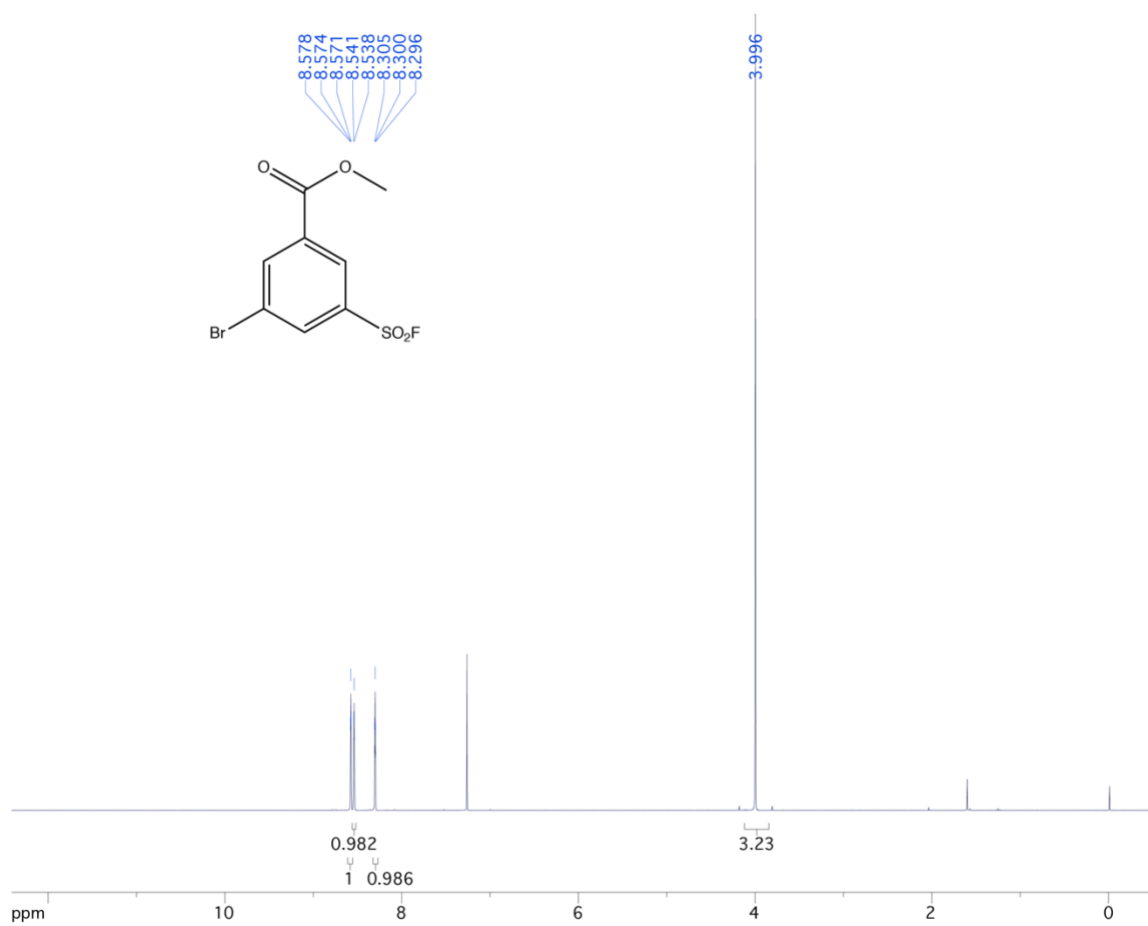
3-fluorosulfonyl benzoic acid:



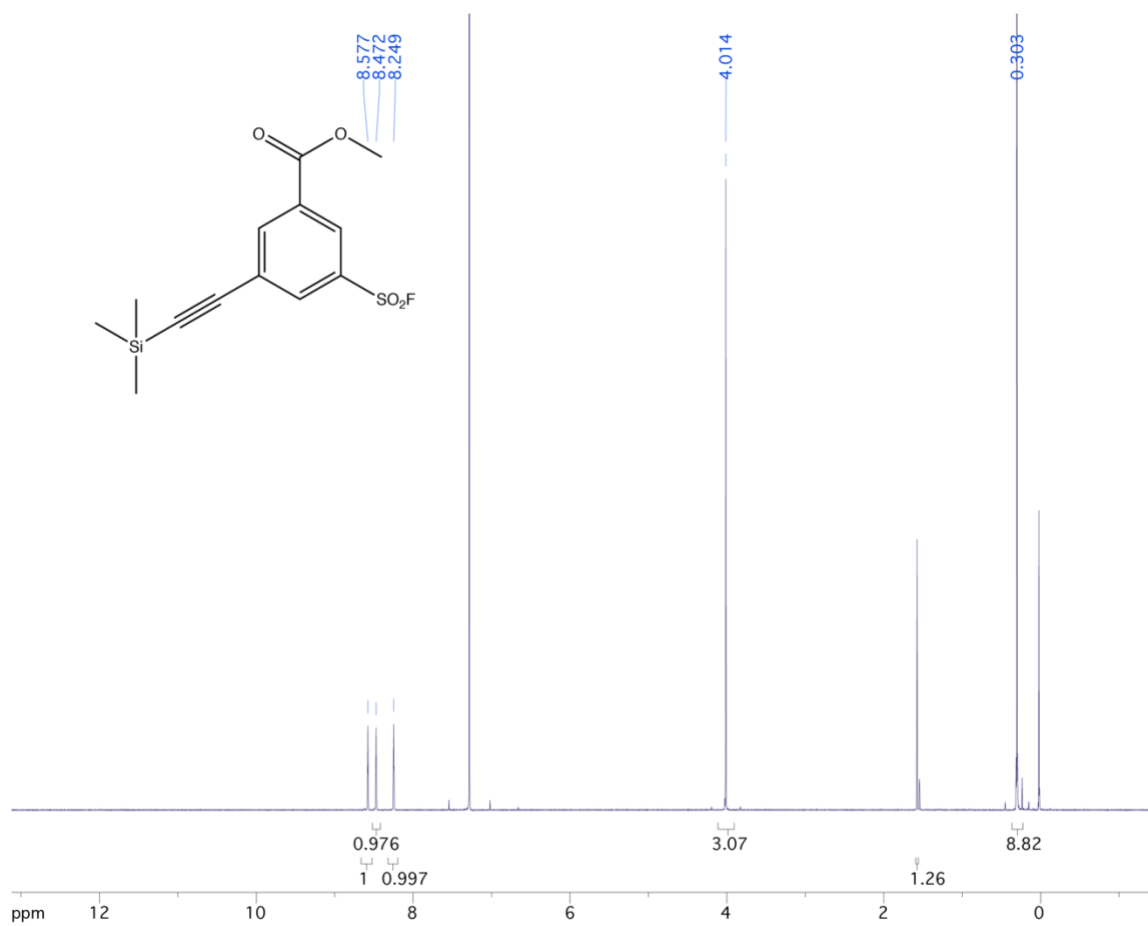
3-(fluorosulfonyl) methyl benzoate



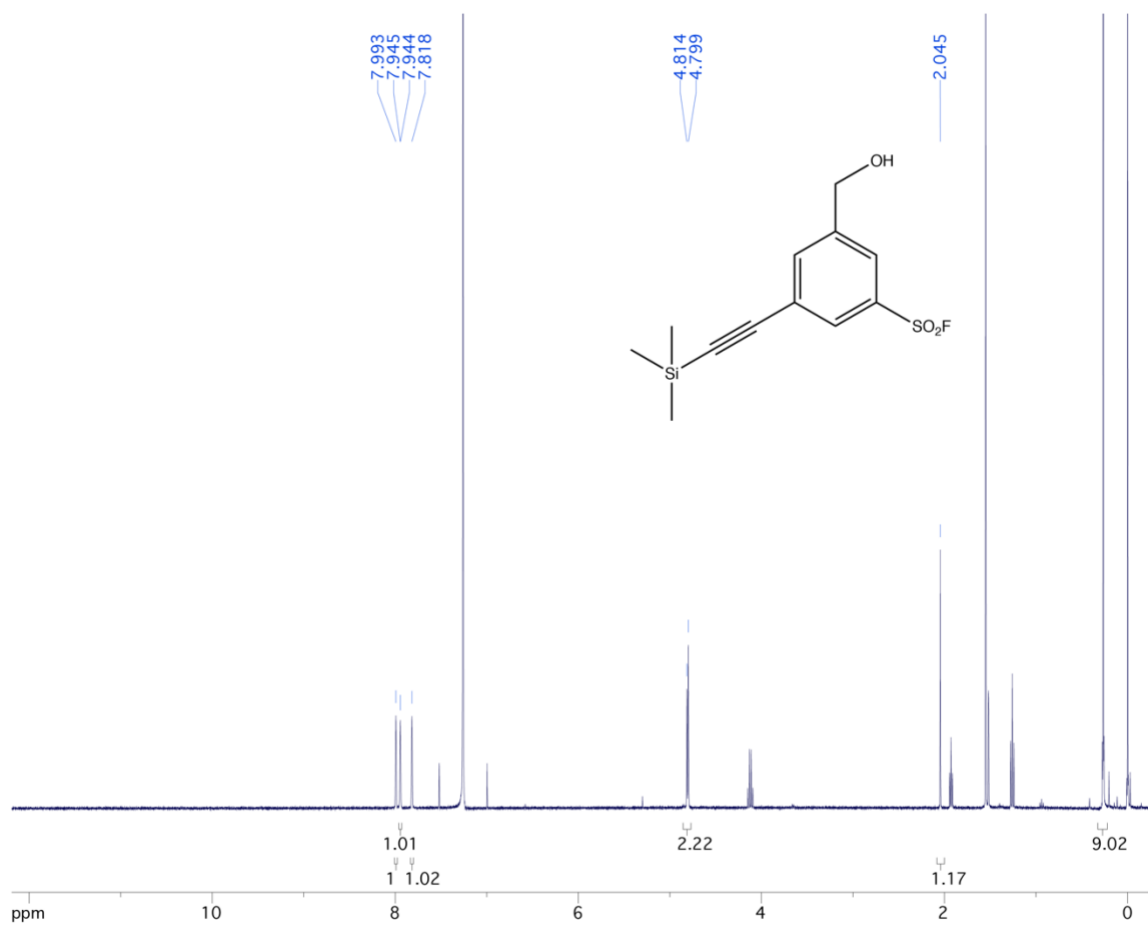
3-bromo-5-(fluorosulfonyl) methylbenzoate



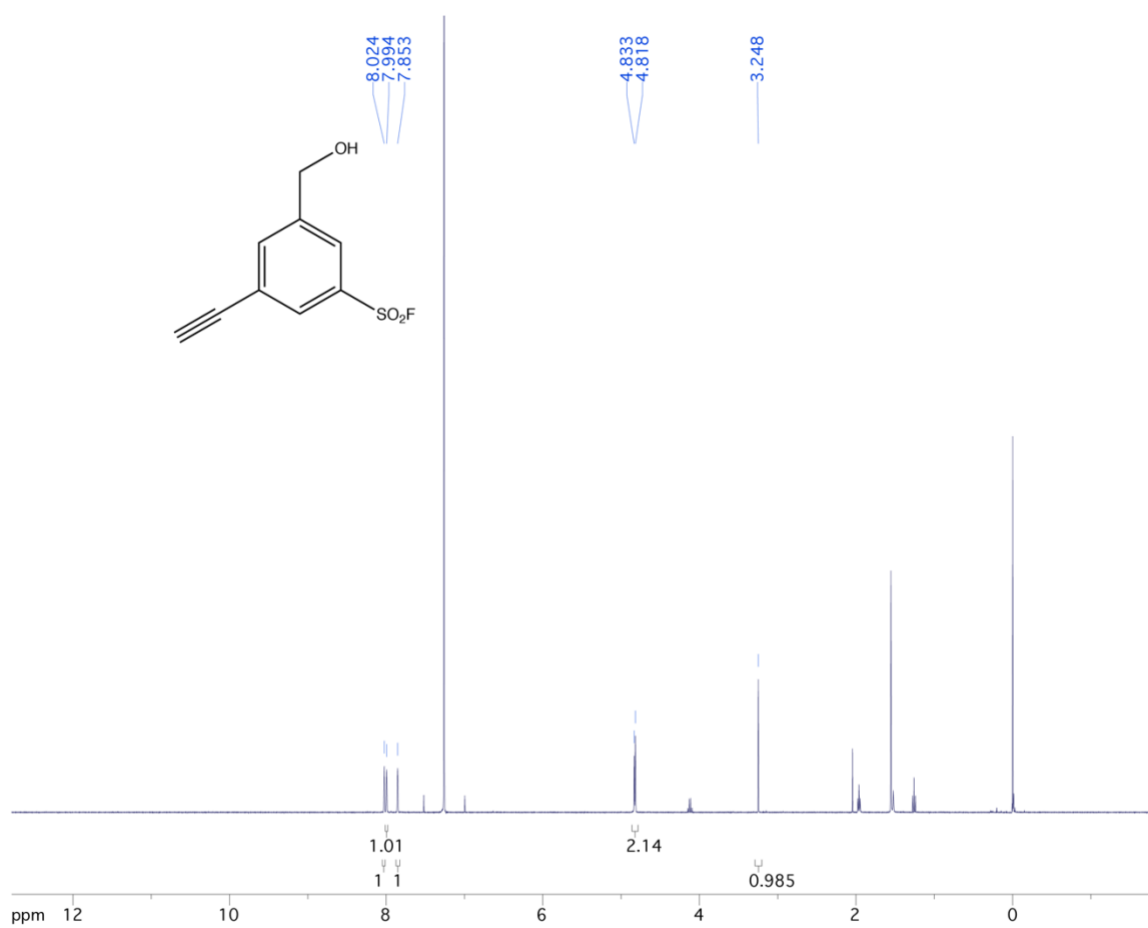
3-(fluorosulfonyl)-5-((trimethylsilyl)ethynyl)methylbenzoate



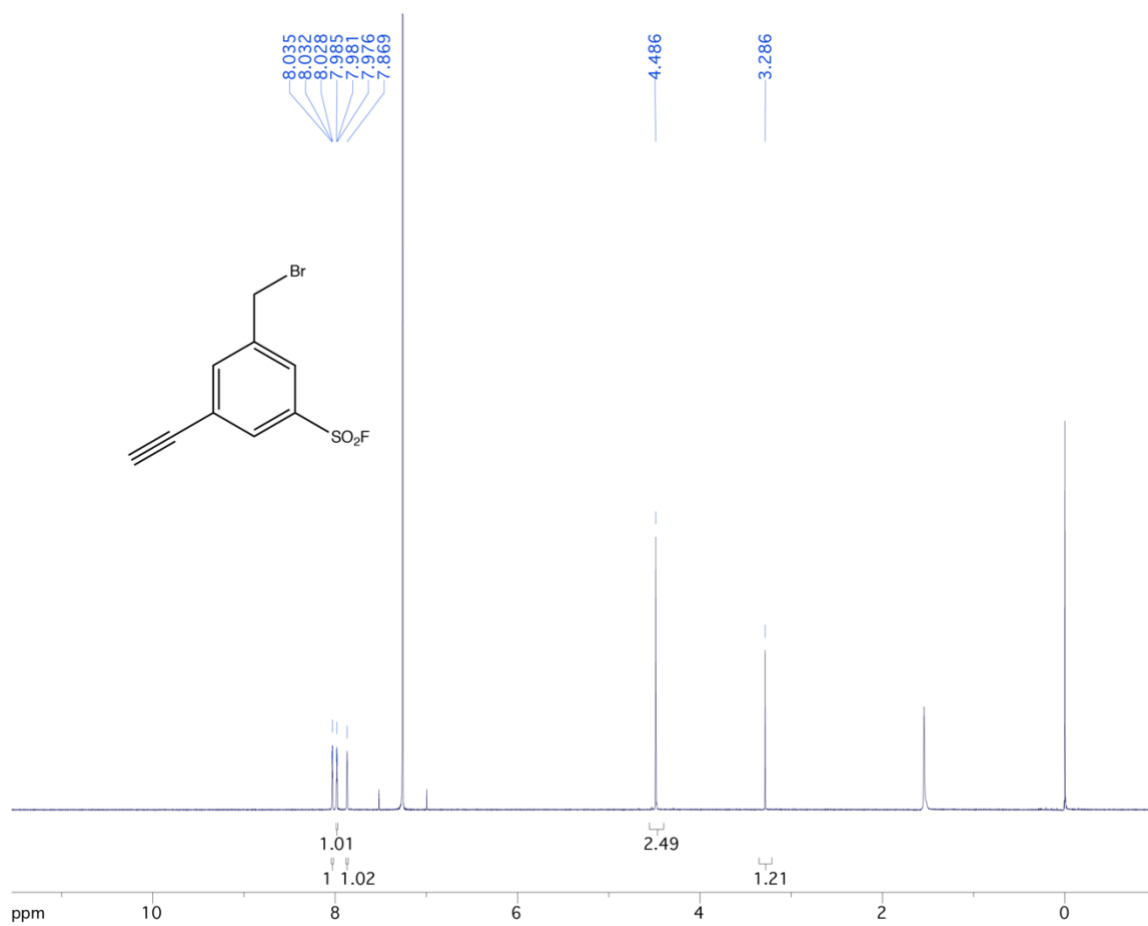
3-(hydroxymethyl)-5-((trimethylsilyl)ethynyl)benzene-1-sulfonyl fluoride



3-ethynyl-5-(hydroxymethyl)benzene-1-sulfonyl fluoride



3-(bromomethyl)-5-ethynylbenzene-1-sulfonyl fluoride



Appendix B: Python scripts for data processing

1. AA_per_pocket.py

```
import sys, csv

sipath = sys.argv[1]

datadict = {}

aa3L = ["ALA", "CYS", "ASP", "GLU", "PHE", "GLY", "HIS", "ILE", "LYS", "LEU", "MET", "ASN", "PRO", "GLN", "ARG",
        "SER", "THR", "VAL", "TRP", "TYR"]
outputdict = {el:[] for el in aa3L}
print outputdict

with open(sipath, 'rbU') as sicsv:
    csvreader = csv.reader(sicsv)
    next(csvreader)#skip header
    for line in csvreader:
        pdbid = line[0]
        datadict[pdbid] = {}
        residues = line[2].strip().split(",")
        for line in residues:
            aa = line.split("-")[0]
            if aa not in datadict[pdbid].keys():
                datadict[pdbid][aa] = 1
            else:
                datadict[pdbid][aa] += 1

for aa in outputdict.keys():
    for id in datadict.keys():
        try:
            count = datadict[id][aa]
        except:
            count = 0

        while len(outputdict[aa]) <= count :
            outputdict[aa].append(0.)

        outputdict[aa][count] += 1

with open("Amino acids per pocket.csv", 'w') as csvfile:
    csvwriter = csv.writer(csvfile)
    csvwriter.writerow(["Amino Acid"] + range(0, 30))
    for aa in outputdict.keys():
        csvwriter.writerow([aa] + outputdict[aa])
```

2. has_good_K_v3.3_CSV_Output.py

```
#!/usr/bin/python
import os, string, sys, math, subprocess, csv
import cPickle as pic
from html import HTML
from time import gmtime, strftime

print "Starting Script\n"
print strftime("%Y-%m-%d %H:%M:%S")

#program will take three things as arguments: The location of the PDB files, a list of names of the pdbfiles,
#and the path/filename for the output
#This version (v3_3) aims to use a less restrictive set of ligands and will not use Xiaobo's set of bad ligands
#It will only use those ligands which are explicitly indicated in the script. It does not remove nucleotides explicitly
#Version 3_3 does not remove phosphate containing ligands (unlike V3_2).

pdbpath = sys.argv[1]
pdblast= sys.argv[2]
outputname = sys.argv[3]
#badligpath = sys.argv[4]

""" help """
""" usage: ./filter.py rec.crg topdock.pdb """

"""global variables used throughout the script"""
def dist (a1,a2):
    return math.sqrt((a1[0]-a2[0])*(a1[0]-a2[0])+(a1[1]-a2[1])*(a1[1]-a2[1])+(a1[2]-a2[2])*(a1[2]-a2[2]))

def unique(seq):
    checked = []
    for e in seq:
        if e not in checked:
            checked.append(e)
    return checked
"""
Identify all lysine sidechain amines, glutamate or aspartate and store their coordinates

Return 4 arrays, one each for Lysine, Glutamate, Aspartate or the Ligand.

Amino Acid arrays have tuples as elements, starting with the Name, Chain and ResidueNUmber,
followed by a list containing the cartesian coordinates

Ligand atom arrays only contain cartesian coordinates
"""
def getResidues(pdbfilename, ligand):
    """Open the file so that it can be accessed"""
    f = open(pdbfilename)
    pdbfile = f.readlines()
    f.close

    """lists where I will store info from the pdb file"""
    lysines = []
    ligandatoms = []
    ligandatomserialdict = {}
    conectDict = {}
    headerinfo = []
    crossref = []

    for line in pdbfile:
        if ((line[0:6] == "COMPND") and (line[11:19] == "MOLECULE")):
            headerinfo.append(line[11:len(line)].strip())
        elif ((line[0:6] == "COMPND") and (line[11:16] == "CHAIN")):
            headerinfo.append(line[11:len(line)].strip())
        elif line[0:4] == "ATOM":
            if ((line[12:16]==" NZ ") and (line[17:20]=="LYS")):
                lysines.append((line[17:26], [float(line[30:38]),float(line[38:46]),float(line[46:54])]))
            #
            elif line[0:4] == "ATOM":
                if ((line[12:16]==" OH ") and (line[17:20]=="TYR")):
                    lysines.append((line[17:26], [float(line[30:38]),float(line[38:46]),float(line[46:54])]))
                #
                elif line[0:4] == "ATOM":
                    if ((line[12:16]==" SG ") and (line[17:20]=="CYS")):
                        lysines.append((line[17:26], [float(line[30:38]),float(line[38:46]),float(line[46:54])]))
                    #
                elif line[0:6] == "HETATM":
```

```

        if (line[17:20]==ligand):
            ligandatoms.append([float(line[30:38]),float(line[38:46]),float(line[46:54])])
            serial = line[6:11].strip()
            ligandatoms[serial] = [float(line[30:38]),float(line[38:46]),float(line[46:54])]
    elif line[0:6] == "DBREF ":
        #Contains related DB information
        #PDB Chain ID, DB Abbr, DB ID, Entry Description
        crossref.append([line[12:13], line[26:32], line[33:41], line[42:54]])
    elif line[0:6] == "CONNECT":
        atomSerial = line[6:11].strip()
        bondedSerials = line[11:].split()
        connectDict[atomSerial] = bondedSerials

    return headerinfo, lysines, ligandatoms, crossref, ligandatomsdict, connectDict

"""
Takes an array of Lysine name and coordinate tuples, and an array of Ligand coordinates
Calculates whether the lysine amine is within 10A of any ligand atom
Returns a tuple of unique lysine residues, coordinates and the distance from the lysine
amine to the closest ligand atom
"""
def kNearLigand(lysinesArray, ligandatomsArray):
    proxlysines = []
    uniquelylysines = []
    for l in lysinesArray:
        for k in ligandatomsArray:
            if dist(l,k[1]) < 5:
                proxlysines.append(k)
    proxlysines = unique(proxlysines)

    if (len(proxlysines) > 0):
        for pk in proxlysines:
            distance = 10
            for l in ligandatomsArray:
                if (dist(l,pk[1]) <= distance):
                    distance = dist(l,pk[1])
            uniquelylysines.append((pk, distance))

    print uniquelylysines
    return uniquelylysines

def formatK(kArray):
    outputArray = []
    if (len(kArray) > 0):
        for k in kArray:
            outputArray.append((k[0][0], None, k[1], None))

    output = unique(outputArray)
    return output

"""
Takes 3 Arrays: An Array of Lys Coordinates, Glu Coordinates, and Asp Coordinates
Determine whether any Glu or Asp are within 3.0A of Lys Amine
Returns an Array containing the Name and Number of the Lys - X pair and the distance btwn them
"""
def edNearK(kArray, eArray, dArray):
    pairs = []
    if (len(kArray) > 0):
        for k in kArray:
            if len(eArray) > 0:
                for e in eArray:
                    if dist(k[0][1], e[1]) <= 3.3:
                        pairs.append((k[0][0], e[0], k[1], dist(k[0][1], e[1]))) #Name of lysine, name of glu,
#Name of lysine, name of glu,
lys distance from ligand, lys distance from glu
            if len(dArray) > 0:
                for d in dArray:
                    if dist(k[0][1], d[1]) <= 3.3:
                        pairs.append((k[0][0], d[0], k[1], dist(k[0][1], d[1]))) #Name of lysine, name of asp,
#Name of lysine, name of asp,
lys distance from ligand, lys distance from asp
    pairs = unique(pairs)
    return pairs

"""
For each atom in the ligand dictionary, determine the hybridization state based off of distances.
Calculate the distance between the each ligand atom and its connected atoms.
Based on these distances, decide whether the bond is a single, double or triple bond.
"""
#def hybridization(ligDict, connectDict):

```



```

"""
Takes the molecular info from the pdb file and the list of interesting lysines and outputs
an array with the results
"""
def makeOutput (headerinfo, goodK, outputTable, ligandAbbr, referencedb, ligandNameTuple):
    oneStringTxt = ""
    oneStringHTML = ""
    for line in headerinfo:
        oneStringTxt = oneStringTxt + line + "\n"
        oneStringHTML = oneStringHTML + line + "<br>"
    pdbcode = pdbfilepath
    pdbcode = pdbcode.strip('./')
    if "pdb" in pdbcode:
        start = pdbcode.find(".pdb")
        pdbcode = pdbcode[start-4:start]
    if ligandNameTuple[1] == "HETSYN":
        ligandName = ligandNameTuple[0]
    else:
        ligandName = ligandNameTuple[1]
    outputTable.append((pdbcode, oneStringHTML, goodK, ligandAbbr, referencedb, ligandName))
    return outputTable

"""
Takes a results array and outputs an HTML file with a sortable table
Edited in this latest version to also output a CSV file with the same table without the links
"""
def writeHTML (outputTable, name):
    h = HTML()
    ta = h.script(' ', src='./scripts/sortable.js')
    ta.text("</script>", escape=False)
    print "This line works fine!"
    ta.table(border='1', klass='sortable')

    row = ta.tr
    row.th("PDB Code")
    row.th("Protein Description")
    row.th("Ligand")
    row.th("Lysine")
    #row.th("Activator")
    row.th("Distance from Ligand (A)")
    #row.th("Distance from Activator (A)")
    row.th("Other DB Reference", newlines=True)

    csv_output = []

    total = len(outputTable)
    totalEntries = 0
    for output in outputTable:
        for o in output[2]:
            #csv output
            outline = []
            outline.append(output[0])#PDB ID
            outline.append(output[1])#Protein Description Placeholder
            outline.append(output[5])#Ligand Name
            outline.append(o[0])#Lysine
            outline.append(o[2])#Distance from Ligand
            chainid = o[0][4]
            prot_desc = output[1].split("<br>")

            #Pull out the protein description for the lys of interest
            for index, val in enumerate(prot_desc):
                if "CHAIN:" in val:
                    chains = val[6:]
                    if chainid in chains:
                        outline[1] = prot_desc[index-1][10:-1]
                        break

            #html output
            r = ta.tr
            r.td(h.a(output[0], href="http://www.pdb.org/pdb/explore/explore.do?structureId="+output[0]), escape=False)
            r.td(output[1], escape=False)
            r.td(h.a(output[5] + " (" + output[3] + ")", href="http://www.rcsb.org/pdb/ligand/ligandsummary.do?hetId="+output[3]),
            escape=False)

            r.td(o[0], escape=False)
            #r.td(o[1], escape=False)
            r.td(o[2], escape=False)
            #r.td(o[3], escape=False)
            id = unique(output[4])

```

```

for item in id:
    if o[0][4] == item[0]:
        #print item[1]
        if (str(item[1]).strip() == 'UNP'):
            r.td(h.a(item[3], href="http://uniprot.org/uniprot/"+item[2]), escape=False,
newlines=True)

            outline.append(item[2])#UniprotID
        elif (str(item[1]).strip() == 'PDB'):
            print "skip"
        elif (str(item[1]).strip() == 'GB'):
            r.td(h.a(item[3], href="http://www.ncbi.nlm.nih.gov/protein/"+item[2]),
escape=False, newlines=True)

            outline.append(item[2])#GenebankID
        else:
            r.td(item, escape=False, newlines=True)

totalEntries = totalEntries + 1
csv_output.append(outline)

#print ta
row.td("Total Number of Proteins: " + str(total) + "<br>", escape=False)
row.td("Total Number of Entries: " + str(totalEntries))

f = open(name, 'w')
f.write(str(ta))
f.close()

with open("%s.csv" % name, 'wb') as csvfile:
    csvwriter = csv.writer(csvfile)
    csvwriter.writerow(['PDB Code', 'Protein Description', 'Ligand', 'Lysine', 'Distance from Ligands (A)', 'Uniprot/Genebank ID'])
    for line in csv_output:
        csvwriter.writerow(line)

```

Accept a PDB file and return a list of unique useful ligands.
Useful ligands are those that are not salts, crystallographic buffers or non-competable cofactors like heme, and entries that are part of the chain that are listed as heteroatoms such as non-standard amino acids or amino acids with post-translational modifications.

Items filtered include:

MSE: Selenomethionine, PTR: Phosphotyrosine, ORN: Ornithine, 4BF: 4-Bromo-Phenylalanine, TY5: Sulfotyrosine, TPO: Phosphothreonine, SEP: Phosphoserine, CCS: CarboxymethylCysteine, CSW: Cysteine Dioxide, HEM: Heme, HOH: Water, CSX: S Oxy Cysteine, HDD: CIS-HEME D HYDROXYCHLORIN GAMMA-SPIROLACTONE SRM: Siroheme

Ions such as:

SO4: Sulfate, CA: Calcium, CL:Chloride, NA: Sodium, PO4: Phosphate, MN: Manganese, K: Potassium, FE: Iron, F: Fluoride, LI: Lithium, MN3: Manganese 3+, NO2: Nitrate, PD: Palladium ion, VO4: Vanadate YT3: Yttrium(III), W: Tungsten, WO4, Tungstate, 3PO: Triphosphate, MGF: Trifluoromagnesate SO3: Sulfite, RB: Rubidium, POP: Pyrophosphate, PR: Praseodymium, PO3: Phosphite, OS: Osmium MF4: Magnesium Tetrafluoride, LU: Lutetium, PB: Lead(II), GD: Gadolinium, EMC: Ethylmercury HG2: Dibromomercury, CS: Cesium, ALF: Tetrafluoroaluminat, AF3: Aluminum Floride, PPK: (DIPHOSPHONO)AMINOPHOSPHONIC ACID TMO: Trimethylamine Oxide, SAT: Sulfoacetic Acid, SIN: Succinic Acid, RH3: Rhodium (III), PI: Hydrogenphosphate HO: Holmium, 6BP: Hexabromoplatinate, DVT: Decavanadate, ARS: Arsenic, SM: Samarium

PTM:

BMA: Beta-D-Mannose, CRQ: Fluorescent Protein Chromophore, MLZ: N-Methyl Lysine, MME: N Methyl Methionine NAG: N-Acetyl-Glucosamine, SMC: S-MethylCysteine, OFM: Fluorescent Protein Chromophore, OIM: Fluorescent Protein Chromophore, NRQ: Fluorescent Protein Chromophore, CRO: Fluorescent Protein Chromophore, CH6: Chromophore, EYG: Chromophore, CGU: Gamma Carboxy glutamate

Crystallographic Artifacts and Detergents:

BOG: Beta-Octyl Glucoside, GOL: Glycerol, CIT: Citrate, CPS: CHAPS Detergent, DMF: Dimethylformamide, DOD: Deuterated Water, FLC: Citrate, GAI: Guanidinium, NME: Methylamine, PCA: Pyroglutamic Acid (N terminal cyclized glutamine) PEO: Hydrogen Peroxide, LMU: DODECYL-ALPHA-D-MAL TOSIDE, MES: 2-(N-MORPHOLINO)-ETHANESULFONIC ACID, OXL: Oxalate URE: Urea, TAM: Tris, BET: Trimethyl Glycine, TFA: Trifluoroacetic Acid, TGL: Triacylglycerol, STE: Stearic Acid, TTN: Tartronate, PGO: Propanediol, 12P: Polyethylene glycol P4C: Polyethylene, P3A: Phosphatidyl Glycerol, 7PH: Phosphatidic Acid, PLM: Palmitic Acid, OLA: Oleic Acid DR6: Big big detergent, 37X: Octaglucose neopentyl glycol, HSJ: OCTYL BETA-L-TALOPYRANOSIDE, IDS: O2-SULFO-GLUCURONIC ACID 2PE: Nonaethylene glycol, OCT: N-Octane, MYR: Myristic Acid, SGM: Monothioglycerol, MOH: Methanol DAO: Lauric Acid, GOA: Glycolic Acid, HEZ: Hexane-1,6-Diol, PEE: DOPE, LMT: Dodecyl-Beta-D-maltoside, SDS: Sodium Dodecyl Sulfate PLC: Diundecyl Phosphatidyl Choline, CDL: Cardioplin , DMA: Dimethylallyl diphosphate, DKA: Decanoic Acid ACM: Acetamide, PC1: Big lipid, PGA: Phosphoglycolic acid, FGL: 2-AMINOPROPANEDIOIC ACID, 1PG: PEG COM: 1-THIOETHANESULFONIC ACID, OLC: (2R)-2,3-dihydroxypropyl (9Z)-octadec-9-enoate, DIO: Dioxin BU2: 1,3-Butanediol, 52N: 1,2-dioctanoyl phosphatidyl epi-inositol (3,4)-bisphosphate, PIE: 1,2-DIACYL-SN-GLYCERO-3-PHOSPHOINOSITOL, L2O: (2S,3R)-3-amino-2-hydroxy-5-methylhexanoic acid, LYK: (2S)-2,6-diaminohexane-1,1-diol, DTU: (2R,3S)-1,4-DIMERCAPTOBUTANE-2,3-DIOL, TMA: Tetramethylammonium, PPI: Propanoic Acid, 7PE: PEG Fragment, PE3: PEG, PPF: PHOSPHONOFORMIC ACID, PGH: PHOSPHOGLYCOLOHYDROXAMIC ACID IPH: Phenol, C15: N-DODECYL-N,N-DIMETHYL-3-AMMONIO-1-PROPANESULFONATE, HEX: Hexane, 6NA: Hexanoic Acid P33: PEG330, B7G: HEPTYL-BETA-D-GLUCOPYRANOSIDE (B7G), XPE: Decaethylene Glycol, ETE: Some Polyester B3P: DiTris, 2PG: 2-PHOSPHOGLYCERIC ACID, MXE: 2-MethoxyEthanol, ETX: 2-EthoxyEthanol, 3PP: 3-PHOSPHONOPROPANOIC ACID PG5: Some polyester, P4G: Some Polyester, BU1: 1,4-Butanediol, DXE: Dimethoxyethane, BU3: 2,3-Butanediol,

0V5: (2R)-2-(phosphonoxy)propanoic acid, TBU: TERTIARY-BUTYL ALCOHOL, PLD: DIUNDECYL PHOSPHATIDYL CHOLINE
PEF: Big big lipid, EPH: Big lipid, OPE: Cholamine Phosphoric Acid, 15P: PEG

Buffers:

MPO: 3 N-Morpholino Propane Sulfonic Acid, CPQ: Deoxy-Bigchap, NHE: CHES

Sugars:

SUC: Sucrose, SPV: Sulfopyruvate, PYR: Pyruvic Acid, PEP: Phosphoenolpyruvate, SIA: Sialic Acid
MAL: Maltose, MLI: Malonate, MLT: Malate, PEQ: L-phospholactate, LMR: L-malate, IDR: Iduronic Acid
ICT: Isocitric Acid, ISD: Isoascorbic acid, 4IP: INOSITOL-(1,3,4,5)-TETRAKISPHOSPHATE,
F6P: Fructose-6-Phosphate, F6R: Fructose-6-Phosphate, MLA: Malonic Acid, TAR: Tartaric Acid
LAT: Beta-lactose, M6D: Mannose-6-Phosphate, BG6: Glucose-6-Phosphate, BGC: Beta-d-glucose
GLS: BETA-D-GLUCOPYRANOSE SPIROHYDANTOIN, GAL: Beta-d-galactose, MLR: Maltrose,
RAM: Rhamnose, FUC: Fucose, MAK: Ketomalonic Acid, MAN: Mannose, G6P: Alpha Glucose 6 Phosphate,
VG1: Glucose-6-P-1-Vanadate, G1P: Glucose-1-P, G16: Glucose-1,6-Bisphosphate, GLC: Glucose
TRE: Trehalose, GLG: ALPHA-D-GLUCOPYRANOSYL-2-CARBOXYLIC ACID AMIDE, GLA: Alpha galactose
FUL: 6-Deoxy-Beta-L-Galactose, G6D: 6-Deoxy-Alpha-D-Glucose, MAH: 3-HYDROXY-3-METHYL-GLUTARIC ACID
AKG: 2-oxoglutaric acid, 5MM: DiphosphonoMannitol, ALS: 2-AMINO-3-OXO-4-SULFO-BUTYRIC ACID,
FGP: 2-AMINO-3-HYDROXY-3-PHOSPHONOXY-PROPIONIC ACID, NDG: 2-(ACETYLAMINO)-2-DEOXY-A-D-GLUCOPYRANOSE
NBG: 1-N-ACETYL-BETA-D-GLUCOSAMINE, 2FP: Linear FructoseDiphosphate, MA1: Dithioglucopyranose
13P: Dihydroxyacetonephosphate, SRT: S,R MESO-TARTARIC ACID, AGL: 4,6-DIDEOXY-4-AMINO-ALPHA-D-GLUCOSE
OAA: Oxaloacetate, IHP: INOSITOL HEXAKISPHOSPHATE, INS: Myo-Inositol, FUM: FUMARIC ACID,
IP5: Myoinositol pentaphosphate, I3P: Inositol triphosphate, I7P: Inositol Phosphate, I8P: Inositol Phosphate
GLF: Fluoro-Alpha-Deoxy-Glucose, M2P: Manitol-Diphosphate, ASO: Anhydrosorbitol

Biological Lipids:

FPP: Farnesyl Diphosphate, GRG: GeranylGeranyl Diphosphate, FPS: Farnesyl thiopyrophosphate
FAR: Farnesyl, EIC: Linoleic Acid, MGM: 3-AZAGeranylgeranyl diphosphate, FII: FPP ANALOG, PCW: Phospholipid
P3A: Phospholipid, P6L: Phospholipid, PEF: Phospholipid

Other:

UNX: Unknown, UNL: Unknown Ligand, CPT: Cisplatin, PUT: 1,4-DIAMINOBTANE, ICF: Isofluorane, UNK: Unknown

def getGoodLigands(pdbfilename):

```
f = open(pdbfilename)
pdbfile = f.readlines()
f.close

#f = open(badligpath, 'r')
#badligands = pic.load(f)
ligandsDict = {}
formulaDict = {}
ligandsList = []
badligands = ['NBN', 'ENC', 'MNC', 'NPN', 'GSH', '0HH', 'GTX', 'BEN', 'OAR', 'PYX', 'CFC', 'ETF', 'TDG', 'ADC', 'RIP', 'CFO', 'OFO', '9DI',
'HPA', 'NOS', 'MET', 'LEU', 'DME', 'TRA', '1DA', 'O', 'MBG', 'SCR', 'AYA', 'ETA', 'HAE', 'MIC', 'DAS', 'BE2', 'DCE', 'ASP', 'HEA', 'LDA', 'SNN', 'IAS', 'HC',
'BHD', 'EDR', 'SGN', 'NTP', 'PYJ', 'SAC', 'TSU', 'CYH', 'AOP', 'CRA', 'MBR', 'CTO', 'UMG', 'GUM', 'BLV', 'FMS', 'A2G', 'DAR', 'PMB', 'MMC', '2FU', 'PBP',
'INN', 'BRC', 'SSB', 'SNC', 'HGB', 'AJ3', 'ADN', 'BAM', 'NVA', 'H2S', 'ADE', 'ACA', 'MTA', 'DEC', 'NON', 'COH', 'HMC', 'TRI', 'PHG', 'CAD', 'PG6', 'SAH',
'NH3', 'DDQ', 'ABH', 'SO2', 'SET', 'NLE', 'DSS', 'BUQ', 'CYS', 'AR7', '0QE', 'PYZ', 'ITU', 'LA', 'V7O', 'PHQ', 'HRG', 'CMT', 'GER', 'CSS', 'LPB', 'BOM',
'ASG', 'GLU', '4MZ', '1MZ', 'HAR', 'DHM', 'UNA', 'GUP', 'DDZ', 'BO3', 'AAE', 'DTD', 'PBR', 'HLT', 'ALA', 'PRO', 'GCS', 'IPU', 'IUM', 'CDU', 'HEV', 'CRS',
'RCO', 'IOL', 'HDS', 'ASJ', 'CNN', 'MGS', 'MGU', 'DCY', 'ESA', 'CE', 'GCU', 'HXA', 'LAX', 'AZE', 'AIB', 'SE4', 'ATH', 'MOS', 'FRU', 'FGA', 'DAM', 'PUR',
'HNI', 'ARG', 'TL', 'GL2', 'GL5', 'GL7', 'GL9', 'CR1', 'CR6', 'SGA', 'MAG', 'ACN', '3OL', 'AEM', 'EMM', '3OM', 'IUR', 'IDK', 'URA', 'GLY', 'NTA', 'CDG',
'GM2', 'D12', 'TWT', 'AOA', 'OGA', 'BEZ', 'G2F', 'DO3', 'LO', 'CXL', 'AOE', 'ANL', 'NA1', 'NAA', 'TBR', 'PFB', 'BRB', 'FEO', 'ELA', 'NNH', 'PP9', 'UNK',
'PEF', 'P6L', 'P3A', 'PCW', 'NHE', '15P', 'CPQ', 'SM', 'MPO', 'OPE', 'EPH', 'PEF', 'PLD', 'CGU', 'TMA', 'TBU', '0V5', 'BU3', 'DXE', 'ASO', 'M2P', 'BU1',
'GLF', 'P4G', 'PG5', 'I7P', 'I8P', '3PP', 'ETX', 'MXE', '2PG', 'B3P', 'ETE', 'ARS', 'EYG', 'I3P', 'IP5', 'XPE', 'DVT', 'FII', 'FUM', 'B7G', 'P33', 'HO', '6BP',
'HEX', '6NA', 'PI', 'INS', 'IHP', 'C15', 'OAA', 'IPH', 'PGH', 'PPF', '7PE', 'PE3', 'PPI', 'PUT', 'AGL', 'RH3', 'SRT', 'SIN', 'SAT', 'TMO', 'DTU', 'LYK', 'L2O',
'PPK', 'PIE', '52N', 'BU2', '13P', 'DIO', 'MA1', '2FP', 'NBG', 'OLC', 'COM', '1PG', 'NDG', 'ALS', 'FGP', 'FGL', '5MM', 'AKG', 'PGA', 'MGM', 'MAH', 'PC1',
'FUL', 'G6D', 'EIC', 'ACM', 'GLA', 'GLG', 'TRE', 'GLC', 'VG1', 'G1P', 'G16', 'G6P', 'MAK', 'MAN', 'RAM', 'FUC', 'AF3', 'MLR', 'GLS', 'GAL', 'BGC', 'M6D',
'BG6', 'LAT', 'CS', 'CH6', 'TAR', 'DKA', 'HG2', 'CPT', 'MLA', 'DMA', 'CDL', 'LMT', 'SDS', 'PLC', 'PEE', 'EMC', 'FAR', 'FPP', 'GRG', 'F6P', 'F6R', 'GD',
'HEZ', 'GOA', '4IP', 'ISD', 'ICT', 'IDR', 'LMR', 'DAO', 'PEQ', 'PB', 'LU', 'MF4', 'MLT', 'MLI', 'MAL', 'MOH', 'SGM', 'MYR', 'OCT', '2PE', 'IDU', 'SIA', 'IDS',
'37X', 'HSJ', 'DR6', 'OLA', 'OS', 'PLM', '7PH', 'PO3', 'PEP', '12P', 'P4C', 'PR', 'POP', 'PYR', 'RB', 'PGO', 'TTN', 'STE', 'SPV', 'SO3', 'SUC', 'ALF', 'TGL',
'TFA', 'MGF', 'BET', 'TAM', '3PO', 'WO4', 'W', 'UNL', 'URE', 'OXL', 'YT3', 'VO4', 'MES', 'LMU', 'CRO', 'NRQ', 'UNX', 'OIM', 'OFM', 'PD', 'SMC', 'PEO',
'PCA', 'NO2', 'NME', 'NAG', 'NHC', 'MME', 'MLZ', 'LI', 'HDD', 'GAI', 'FLC', 'DOD', 'DMF', 'BMA', 'CRQ', 'BOG', 'CIT', 'CPS', 'CSX', 'F', '202', 'SO4', 'GOL',
'ZN', 'MG', 'CA', 'CL', 'NA', 'EDO', 'PO4', 'HEM', 'ACT', 'MN', 'K', 'PEG', 'FE', 'ACE', 'CU', 'MPD', 'NI', 'NH2', 'DMS', 'BME', 'TRS', 'CD', 'PG4',
'FMT', 'ACY', 'SEP', 'EPE', 'SF4', 'PGE', 'CO', 'TPO', 'IOD', 'FE2', 'FES', 'IMD', 'NO3', 'HG', 'CSO', 'PTR', 'IPA', '1PE', 'KCX', 'MRD', 'HEC', 'CMO', 'BR',
'CO3', 'CME', 'TLA', 'SCN', 'OXY', 'P6G', 'NCO', 'CAC', 'TYS', 'CSD', 'MLY', 'DAL', 'F3S', 'EOH', 'CU1', 'AZI', 'HED', 'M3L', 'DTT', 'NH4', 'HYP', 'OCS',
'CYN', 'FME', 'CAS', 'ALY', 'BTB', 'ABA', 'XE', 'CSW', 'BCT', 'SR', 'BA', 'NO', 'C8E', 'BCL', 'BEF', 'PE4', 'AU', 'CP', 'PT', 'YB', 'MSE', 'ORN', '4BF',
'CCS', 'HOH', 'OH']
for line in pdbfile:
    if line[0:6] == "HETATM":
        ligandsList.append(line[17:20])
    if line[0:6] == "HETNAM":
        id = line[11:14]
        name = line[15:].strip()
        if id not in ligandsDict.keys():
            ligandsDict[id] = (name, "HETSYN")
        else:
            ligandsDict[id] = (ligandsDict[id][0] + name, ligandsDict[id][1])
    if line[0:6] == "HETSYN":
        id = line[11:14]
```

```

        name = line[15:].strip()
        if id not in ligandsDict.keys():
            ligandsDict[id] = ("HETNAM", name)
        elif ligandsDict[id][1] == "HETSYN":
            ligandsDict[id] = (ligandsDict[id][0], name)
        else:
            ligandsDict[id] = (ligandsDict[id][0], ligandsDict[id][1] + name)
    if line[0:6] == "FORMUL":
        id = line[12:15]
        formula = line[20:].strip()
        if id not in formulaDict.keys():
            formulaDict[id] = formula
        else:
            formulaDict[id] += formula

ligandsList = unique(ligandsList)
goodligands = []
for lig in ligandsList:
    for bad in badligands:
        if bad == lig:
            print "Not including: %s" % lig
            break
    else:
        print "This is the formula dictionary:"
        print formulaDict
        if lig not in formulaDict.keys():
            print "Lig not in Formul Entry: %s" % lig
        #elif 'P' not in formulaDict[lig]:
        #print "Added %s!" % lig
        #goodligands.append(lig)
    else:
        #print "Skipped %s!" % lig
        print "Added %s!" % lig
        goodligands.append(lig)

return goodligands, ligandsDict

"""
Take a list of molecules, create header lines with molecule name, and organize the molecules in order by their name
Return list containing header entries followed by molecules all having the same name
"""
def organizeMoleculeName(outputarray):
    newOutput = []
    #add in header entries to output
    for line in outputarray:
        headerlines = line[1].split("<br>")
        lysine = line[2][0]
        moleculeName = ("header", getName(headerlines, lysine))
        if moleculeName in newOutput:
            newOutput.insert(newOutput.index(moleculeName)+1, line)
        else:
            newOutput.append(moleculeName)
            newOutput.append(line)
    return newOutput

"""
Take a list of molecules organized by name (and containing header lines). Organize these molecules by Uniprot ID.
Update header lines to include all molecule names and all uniprot id's that match
Return a list containing header files followed by the appropriate name or ID matched entry
"""
def organizeMoleculeID(outputarray):
    newOutput = []
    newHeaderList = []
    headerIndex = 0

    #First, Update each header with the list of uniprotIDs that correspond to it
    for i, line in enumerate(outputarray):
        #If header, add new element that is a list that we can add uniprotIDs to.
        #Hold on to its index and add it to the list.
        if line[0] == 'header':
            print line
            newHeader = (line[0], line[1], [])
            newOutput.append(newHeader)
            headerIndex = newOutput.index(newHeader)
        #If you aren't at a header, gather uniprotIDs and add them to the right header
        else:
            id = getUniprot(line)
            print headerIndex
            print newOutput[headerIndex]
            if (id not in newOutput[headerIndex][2]) and (id != None):

```

```

        newOutput[headerIndex][2].append(id)
    """
    for item in line[4]:
        if item[1].strip() == 'UNP':
            id = item[2].strip()
            if id not in newOutput[headerIndex][2]:
                newOutput[headerIndex][2].append(id)
    """

    newOutput.append(line)

#Next, start combining headers that that share uniprot IDs. Put them into a list containing only headers
for i, line in enumerate(newOutput):
    Added = False
    if line[0] == 'header':
        if newHeaderList == []:
            newHeaderList.append(line)
        else:
            for i, item in enumerate(newHeaderList):
                if Added == False:
                    for id in line[2]:
                        if (id in item[2]) and (id != None):

                            #combine the IDs of the two matching molecules
                            for uniprotID in line[2]:
                                if (uniprotID not in item[2]) and (uniprotID !=
None):
                                    item[2].append(uniprotID)

                            #combine the names of the two matching molecules
                            if type(item[1]) is str:
                                nameList = [item[1], line[1]]
                                newHeader = ("header", nameList, item[2])
                                newHeaderList[i] = newHeader
                            else:
                                item[1].append(line[1])
                            Added = True
                            break

            if Added == False:
                newHeaderList.append(line)

#Lastly, go through the original list.
#For each entry, check uniprotID against the headers. Insert entry after the appropriate header
#Return List
for line in outputarray:
    if line[0] != "header":
        id = getUniprot(line)
        for index, entry in enumerate(newHeaderList):
            if id in entry[2]:
                try:
                    newHeaderList.insert(index+1, line)
                    break
                except IndexError:
                    newHeaderList.append(line)
                    break
            elif id == None:
                headerlines = line[1].split("<br>")
                lysine = line[2][0]
                moleculeName = getName(headerlines, lysine)
                if moleculeName in entry[1]:
                    try:
                        newHeaderList.insert(index+1, line)
                        break
                    except IndexError:
                        newHeaderList.append(line)
                        break

    return newHeaderList

"""
Takes the output array and returns an array grouped by molecule name
"""
def collapseDuplicateMolecules(newOutput, name, count):
    g = HTML()

    ta = g.script(' ', src='sortable.js')
    ta.text("</script>", escape=False)
    ta.text("<div class='layer_1'>", escape=False)

```

```

total = count
totalEntries = 0
moleculeCount = 0
for index, output in enumerate(newOutput):
    if output[0] == "header":
        moleculeCount = moleculeCount + 1
        print output
        if index > 0:
            ta.text("</table>", escape=False)
            ta.text("</div>", escape=False)
            label = ta.h3(str(output[1]), klass="heading", newlines=True)
            content = ta.div(klass="content")
            tableflag = True
        elif tableflag == True:
            ta.table(border='1', klass='sortable')

            row = ta.tr
            row.th("PDB Code", newlines=True)
            row.th("Protein Description", newlines=True)
            row.th("Ligand", newlines=True)
            row.th("Lysine", newlines=True)
            #row.th("Activator", newlines=True)
            row.th("Distance from Ligand (A)", newlines=True)
            #row.th("Distance from Activator (A)", newlines=True)
            row.th("Other DB Reference", newlines=True)
            tableflag = False
        if tableflag == False:
            for o in output[2]:
                r = ta.tr(newlines=True)
                r.td(g.a(output[0], href="http://www.pdb.org/pdb/explore/explore.do?structureId="+output[0]),
escape=False, newlines=True)
                r.td(output[1], escape=False, newlines=True)
                r.td(g.a(output[5] + " (" + output[3] + ")"),
href="http://www.rcsb.org/pdb/ligand/ligandsummary.do?hetId="+output[3]), escape=False, newlines=True)
                r.td(o[0], escape=False, newlines=True)
                #r.td(o[1], escape=False, newlines=True)
                r.td(o[2], escape=False, newlines=True)
                #r.td(o[3], escape=False, newlines=True)
                for item in output[4]:
                    if o[0][4] == item[0]:
                        #print item[1]
                        if (str(item[1]).strip() == 'UNP'):
                            r.td(g.a(item[3], href="http://uniprot.org/uniprot/"+item[2]),
escape=False, newlines=True)
                        elif (str(item[1]).strip() == 'PDB'):
                            print "skip"
                        elif (str(item[1]).strip() == 'GB'):
                            r.td(g.a(item[3], href="http://www.ncbi.nlm.nih.gov/protein/"+item[2]),
escape=False, newlines=True)
                        else:
                            r.td(item, escape=False, newlines=True)

            totalEntries = totalEntries + 1

#print ta
ta.text("</table>", escape=False)
ta.h3(" ", klass="heading", newlines=True)
ta.tf("Total Number of Proteins: " + str(total) + "<br>", escape=False)
ta.tf("Total Number of Entries: " + str(totalEntries))

f = open(name, 'wa')
head = htmlHead()
f.write(str(head))
f.write("<p>Total Molecule Count: " + str(moleculeCount) + "</p>")
f.write("<p>Total Number of PDB Files: " + str(total) + "</p>")
f.write("<p>Total Number of Entries: " + str(totalEntries) + "</p>")
f.write(str(ta))
f.close()

"""
Return the name of the molecule that contains the given residue.
Header is an array where each element is a line in the header
Residue is a string containing the residue information: XXX A####
"""
def getName(header, residue):
    for index, value in enumerate(header):
        if value[0:7] == "CHAIN: ":
            if residue[0][4] in value[7:len(value)]:
                return header[index-1][10:len(header[index-1])]

```

```

        return "Error: "+ str(header)

def getUniprot(line):
    print line
    lysine_chain = line[2][0][0][4]
    print lysine_chain
    for uniprotEntry in line[4]:
        if uniprotEntry[1].strip() == 'UNP':
            if uniprotEntry[0] == lysine_chain:
                return uniprotEntry[2].strip()
        elif uniprotEntry[1].strip() == 'GB':
            if uniprotEntry[0] == lysine_chain:
                return uniprotEntry[2].strip()

def htmlHead():
    h = HTML()
    top = h.head(newlines=True)
    code1 = top.script(" ", src="./scripts/jquery-1.11.1.js", escape=False, newlines=True)
    code2 = top.script(" ", src="./scripts/jquery-ui.js", escape=False, newlines=True)
    code3 = top.script(" ", src="./scripts/collapse.js", escape=False, newlines=True)
    code4 = top.script(" ", src="./scripts/sorttable.js", escape=False, newlines=True)
    style = top.link(" ", rel="stylesheet", type="text/css", href="./scripts/mystyle.css", escape=False, newlines=True)

    return h

def removeLigandEntries(outputList):
    AdenineLigs = ['112', 'ATP', 'ADP', 'ANP', 'IMP', 'CMP', 'CXR', 'AGS', 'AVU', 'APR', 'ACP', 'A3P', 'AMP', 'PPS', '3AM']
    GuanineLigs = ['GSP', 'GNP', 'GDP', 'GCP', 'GP2', 'GTG', 'GTP', '5GP', 'G3D', 'G2R', 'G2Q', 'G1R', 'PCG', 'CGR', 'DGP']
    NADLigs = ['NAP', 'NDP', 'DN4', 'NAD', 'NMN', '2NF', 'NAI']
    FADLigs = ['FAD', 'FMN']
    UridineLigs = ['UPG', 'U5P', 'U2F', 'UDP', 'U2P', 'GDU', 'OMP', 'UMP', 'S5P', 'BMP', '6CN', 'NUP', '6AU', 'XMP', '5FU', 'CNU', '5BU', 'DUR',
    'UD1']
    ThymidineLigs = ['TMP', 'TTP', 'TDP', 'ATM', '2DT']
    CytosineLigs = ['TKW', 'DCP']

    adenineOutput = []
    guanineOutput = []
    nadOutput = []
    fadOutput = []
    uridineOutput = []
    thymidineOutput = []
    cytosineOutput = []
    remainingOutput = []
    for line in outputList:
        if line[3] in AdenineLigs:
            adenineOutput.append(line)
        elif line[3] in GuanineLigs:
            guanineOutput.append(line)
        elif line[3] in NADLigs:
            nadOutput.append(line)
        elif line[3] in FADLigs:
            fadOutput.append(line)
        elif line[3] in UridineLigs:
            uridineOutput.append(line)
        elif line[3] in ThymidineLigs:
            thymidineOutput.append(line)
        elif line[3] in CytosineLigs:
            cytosineOutput.append(line)
        else:
            remainingOutput.append(line)

    return adenineOutput, guanineOutput, nadOutput, fadOutput, uridineOutput, thymidineOutput, cytosineOutput, remainingOutput

def writeOutput(log, name, extra):

    total = len(log)
    writeHTML(log, extra + name)
    organizedLog = organizeMoleculeName(log)
    collapseDuplicateMolecules(organizedLog, extra + "Collapsible_By_Name_" + name, total)
    collapseDuplicateMolecules(organizedMoleculeID(organizedLog), extra + "Collapsible_By_Name_and_ID_" + name, total)

    return organizedLog

"""
MAIN
"""
outputLog = []
f = open(pdblist)

```

```

pdbs = f.readlines()
f.close

ligandList = []
for file in pdbs:
    print strftime("%Y-%m-%d %H:%M:%S")
    pdbfilepath = pdbpath + "/" + file[0:4] + ".pdb"

    ligands, ligandNamesDict = getGoodLigands(pdbfilepath)
    print file
    print ligands

    for ligandName in ligands:
        header, lys, lig, refdb, ligSerialDict, connectDict = getResidues(pdbfilepath, ligandName)
        lys10 = kNearLigand(lys, lig)
        activatedLys = formatK(lys10)
        print activatedLys
        if len(activatedLys) > 0:
            outputLog = makeOutput(header, activatedLys, outputLog, ligandName, refdb, ligandNamesDict[ligandName])
            ligandList.append(ligandName)

ligandList = unique(ligandList)

total = len(outputLog)
organizedOutput = organizeMoleculeName(outputLog)

adenineTable, guanineTable, nadTable, fadTable, uridineTable, thymidineTable, cytosineTable, remainingTable = removeLigandEntries(outputLog)

orgLog = writeOutput(outputLog, outputname, "")
#print orgLog
writeOutput(remainingTable, outputname, "Nucleotide_Free_")

f = open(outputname + "_Ligandlist.pkl", 'w')
pic.dump(ligandList, f)
f.close()

#with open("%s.csv" % outputname, 'wb') as csvfile:
#    csvwriter = csv.writer(csvfile)
#    for line in orgLog:
#        csvwriter.writerow(line)

```


3. PP_LFQ_v5.py

```
import sys, csv, datetime

inputpath = sys.argv[1]
decoderpath = sys.argv[2]
outputpath = sys.argv[3]

print datetime.datetime.now()

#v3 organizes the output by the sample name, based on the raw data file name

#import the decoder into a dictionary. column 1 is the fraction, column 2 is the sample name
decoder = {}
with open(decoderpath, 'rU') as csvfile:
    csv_reader = csv.reader(csvfile)
    for line in csv_reader:
        decoder[line[0].strip()] = line[1].strip()

#import the text file into a list
prospector_output = []
with open(inputpath, 'rU') as csvfile:
    dialect = csv.Sniffer().sniff(csvfile.readline(), delimiters=['\t', ','])
    csvfile.seek(0)
    #csv_reader = csv.reader(csvfile, delimiter='\t')
    csv_reader = csv.reader(csvfile, dialect)
    for line in csv_reader:
        if len(line) > 1:
            prospector_output.append(line)

#extract the header information
header = []
for line in prospector_output:
    #print line
    if "# in DB" in line:
        header = line
        break

#get the index number for "# in DB" so you know where to look for unique peptides
#print header
#put all of the peptides that have a unique entry in the DB in one list
#Add an additional line that converts Cdk12 and 13 into a combined entry. Do the same thing for Cdk16 and 17
#Set the number in DB to 1 if it is 3 or less for the combined entry
#Check to see whether this new line is in the output before adding it, since PP will double count shared peptides

uniquecol = header.index("# in DB")
uniprot_index = header.index("Acc #")
prot_name = header.index("Protein Name")
percent_cov = header.index("% Cov")
mw_index = header.index("Protein MW")

unique_output = []
Cdk12or13 = ["Q9NYV4", "Q14004"]
Cdk16or17 = ["Q00536", "Q00537"]
combinelist = Cdk12or13 + Cdk16or17

header_flag = False
for line in prospector_output:
    if header_flag == False:
        if line == header:
            header_flag = True
    else:
        #print line
        uid = line[uniprot_index]
        if uid not in combinelist:
            if float(line[uniquecol]) == 1:
                unique_output.append(line)
        else:
            if uid in Cdk12or13:
```

```

        if float(line[uniquecol]) == 1:
            unique_output.append(line)
        if float(line[uniquecol]) < 4:
            newline = line[:]
            newline[uniprot_index] = "Cdk12_or_Cdk13"
            newline[prot_name] = "Cdk12_or_Cdk13"
            newline[percent_cov] = "0"
            newline[mw_index] = "0"
            finalline = ((uniprot_index)*[""]) + newline[uniprot_index:]
            if finalline in unique_output:
                print "This line was already added"
            else:
                print "This line hasn't been added yet"
                print finalline
            if finalline not in unique_output:
                print finalline
                unique_output.append(finalline)
elif uid in Cdk16or17:
    if float(line[uniquecol]) == 1:
        unique_output.append(line)
    if float(line[uniquecol]) < 4:
        newline = line[:]
        newline[uniprot_index] = "Cdk16_or_Cdk17"
        newline[prot_name] = "Cdk16_or_Cdk17"
        newline[percent_cov] = "0"
        newline[mw_index] = "0"
        finalline = ((uniprot_index) *[""]) + newline[uniprot_index:]
        if finalline not in unique_output:
            print finalline
            unique_output.append(finalline)

#####
#testing output
with open("Modified Input.csv", 'wb') as csvfile:
    f = csv.writer(csvfile)
    for line in unique_output:
        f.writerow(line)
#####

#split by Fraction (essentially by the name of the raw data file) into separate lists stored in a dictionary
sample_dict = {}
if "Fraction" in header:
    id = header.index("Fraction")
    for line in unique_output:
        sample_id = line[id]
        if "_ITMSms2cid" in sample_id:
            sample_id = sample_id[:-11]
            #print "This is the sample ID %s" % sample_id
        elif "_FTMSms2hcd" in sample_id:
            sample_id = sample_id[:-11]
        elif "_FTMSms2etd" in sample_id:
            sample_id = sample_id[:-11]
        if sample_id in sample_dict.keys():
            sample_dict[sample_id].append(line)
        else:
            sample_dict[sample_id] = [line]

#make a dictionary with all the accession numbers within each id
uniprot_dict = {}

print sample_dict.keys()
for id in sample_dict.keys():
    for line in sample_dict[id]:
        uniprotID = line[uniprot_index]
        if id in uniprot_dict.keys():
            if uniprotID not in uniprot_dict[id]:
                uniprot_dict[id].append(uniprotID)
        else:
            if uniprotID not in uniprot_dict.keys():

```

```

        uniprot_dict[id] = [uniprotID]

#add the intensity of all the peptides corresponding to an individual protein
#each line that is read is set aside in a new list
#before processing each line, it makes sure that the line has not been previously processed
#by checking it against the already processed list. (Guards against duplicates)

begin_new = header.index("Acc #")
intensity_index = header.index("Intensity")
already_checked = {}

intensities = {}
all_unique_peptides = {}

for sample in sample_dict.keys():
    #print sample_dict.keys()
    intensities[sample] = {}
    all_unique_peptides[sample] = {}
    already_checked[sample] = []
    #print uniprot_dict[sample]
    for id in uniprot_dict[sample]:
        #print uniprot_dict[sample]
        #intensities[sample][id] = 0.
        #print "Set to 0"
        #print all_unique_peptides[sample]
        all_unique_peptides[sample][id] = []
        #print all_unique_peptides[sample]
        #print intensities
        for line in sample_dict[sample]:
            #print already_checked
            #print line[begin_new:]
            #print line[uniprot_index]
            #print id
            #if line[begin_new:] not in already_checked[sample]:
            #print "Made it in!"
            if line[uniprot_index] == id:
                print line
                already_checked[sample].append(line[begin_new:])

                intensity = line[intensity_index]
                if intensity == "":
                    intensity = 0.
                else:
                    intensity = float(intensity)
                try:
                    intensities[sample][id] += intensity
                except KeyError:
                    intensities[sample][id] = intensity
                    #print id
                    #print intensity
                    #print intensities[sample][id]
                all_unique_peptides[sample][id].append(line)
        #print uniprot_dict[sample]
        #print intensities[sample]

#print intensities
#go through the uniprotIDs
#use these to go through the sample dictionary
#update the first instance of each uniprotID with the sum of the intensities calculated in the previous loop
#break out of it

##Realistically want to also include the number of unique peptides identified for each protein!!
##Need to keep track of this as well!
"""
protein_sample_dict = {}
for sample in uniprot_dict.keys():
    protein_sample_dict[sample] = []
    for id in uniprot_dict[sample]:

```

```

        for line in sample_dict[sample]:
            if id == line[uniprot_index]:
                protein_line = line
                for each in intensity_indices:
                    protein_line[each] = intensities[sample][id]
            protein_sample_dict[sample].append(protein_line)
        """

#output each sample as its own file.
#Output the Sample#, UniprotID, Intensity, and number unique for each file
#all_unique_peptides contains all instances identified of a peptides matching to a single
#protein. The sum of the elements in this structure is the spectral counts of the peptides
#The number of kinds of peptides uniquely identified can be determine by checking for the
#uniqueness of each entry. Peptides that have different variable modifications are counted
#as being different.

#get the number of the kinds of peptides that have been uniquely identified.
pep_seq_num = header.index("DB Peptide")
try:
    var_mod_num = header.index("Variable Mods")
except ValueError:
    var_mod_num = header.index("Protein Mods")
unique_kinds_of_peptides = {}
for each in all_unique_peptides.keys():
    unique_kinds_of_peptides[each] = {}
    for id in all_unique_peptides[each].keys():
        entry = all_unique_peptides[each][id]
        print entry
        for line in entry:
            print line
            pep_seq = line[pep_seq_num]
            var_mod = line[var_mod_num]
            pepvar = (pep_seq, var_mod)
            try:
                if pepvar not in unique_kinds_of_peptides[each][id]:
                    unique_kinds_of_peptides[each][id].append(pepvar)
            except KeyError:
                unique_kinds_of_peptides[each][id] = [pepvar]

print decoder.keys()
for sample in intensities.keys():
    with open("%s %s" % (decoder[sample].strip(), outputpath), 'w') as csvfile:
        f = csv.writer(csvfile)
        output_header = ["Sample ID", "Uniprot ID", "Unique Peptides", "Spectral Counts", "Protein Intensity"]
        f.writerow(output_header)
        for uniprot in intensities[sample].keys():
            protein_intensity = intensities[sample][uniprot]
            number_unique = len(unique_kinds_of_peptides[sample][uniprot])
            spectral_counts = len(all_unique_peptides[sample][uniprot])
            output_line = [sample, uniprot, number_unique, spectral_counts, protein_intensity]
            f.writerow(output_line)

date = datetime.datetime.today().strftime("%Y%m%d")
with open("%s_Sample_List.txt" % date, 'w') as txtfile:
    for sample in intensities.keys():
        outputline = "%s %s\n" % (decoder[sample], outputpath)
        txtfile.write(outputline)
    print outputline

```

4. format_and_combine_protein_value-v2.py

#Process the output from PP_LFQ_v5.py with this script to get a well organized table of protein intensities.

```
import sys, csv, datetime

inputpath = sys.argv[1]
decoderpath = sys.argv[2]
outputpath = sys.argv[3]

print datetime.datetime.now()

#v3 organizes the output by the sample name, based on the raw data file name

#import the decoder into a dictionary. column 1 is the fraction, column 2 is the sample name
decoder = {}
with open(decoderpath, 'rU') as csvfile:
    csv_reader = csv.reader(csvfile)
    for line in csv_reader:
        decoder[line[0].strip()] = line[1].strip()

#import the text file into a list
prospector_output = []
with open(inputpath, 'rU') as csvfile:
    dialect = csv.Sniffer().sniff(csvfile.readline(), delimiters=['\t', ','])
    csvfile.seek(0)
    #csv_reader = csv.reader(csvfile, delimiter='\t')
    csv_reader = csv.reader(csvfile, dialect)
    for line in csv_reader:
        if len(line) > 1:
            prospector_output.append(line)

#extract the header information
header = []
for line in prospector_output:
    #print line
    if "# in DB" in line:
        header = line
        break

#get the index number for "# in DB" so you know where to look for unique peptides
#print header
#put all of the peptides that have a unique entry in the DB in one list
#Add an additional line that converts Cdk12 and 13 into a combined entry. Do the same thing for Cdk16 and 17
#Set the number in DB to 1 if it is 3 or less for the combined entry
#Check to see whether this new line is in the output before adding it, since PP will double count shared peptides

uniquecol = header.index("# in DB")
uniprot_index = header.index("Acc #")
prot_name = header.index("Protein Name")
percent_cov = header.index("% Cov")
mw_index = header.index("Protein MW")

unique_output = []
Cdk12or13 = ["Q9NYV4", "Q14004"]
Cdk16or17 = ["Q00536", "Q00537"]
combinelist = Cdk12or13 + Cdk16or17

header_flag = False
for line in prospector_output:
    if header_flag == False:
        if line == header:
            header_flag = True
    else:
        #print line
        uid = line[uniprot_index]
        if uid not in combinelist:
            if float(line[uniquecol]) == 1:
                unique_output.append(line)
        else:
            if uid in Cdk12or13:
```

```

        if float(line[uniquecol]) == 1:
            unique_output.append(line)
        if float(line[uniquecol]) < 4:
            newline = line[:]
            newline[uniprot_index] = "Cdk12_or_Cdk13"
            newline[prot_name] = "Cdk12_or_Cdk13"
            newline[percent_cov] = "0"
            newline[mw_index] = "0"
            finalline = ((uniprot_index)*[""]) + newline[uniprot_index:]
            if finalline in unique_output:
                print "This line was already added"
                print finalline
            else:
                print "This line hasn't been added yet"
                print finalline
            if finalline not in unique_output:
                print finalline
                unique_output.append(finalline)
    elif uid in Cdk16or17:
        if float(line[uniquecol]) == 1:
            unique_output.append(line)
        if float(line[uniquecol]) < 4:
            newline = line[:]
            newline[uniprot_index] = "Cdk16_or_Cdk17"
            newline[prot_name] = "Cdk16_or_Cdk17"
            newline[percent_cov] = "0"
            newline[mw_index] = "0"
            finalline = ((uniprot_index) *[""]) + newline[uniprot_index:]
            if finalline not in unique_output:
                print finalline
                unique_output.append(finalline)

#####
#testing output
with open("Modified Input.csv", 'wb') as csvfile:
    f = csv.writer(csvfile)
    for line in unique_output:
        f.writerow(line)
#####

#split by Fraction (essentially by the name of the raw data file) into separate lists stored in a dictionary
sample_dict = {}
if "Fraction" in header:
    id = header.index("Fraction")
    for line in unique_output:
        sample_id = line[id]
        if "_ITMSms2cid" in sample_id:
            sample_id = sample_id[:-11]
            #print "This is the sample ID %s" % sample_id
        elif "_FTMSms2hcd" in sample_id:
            sample_id = sample_id[:-11]
        elif "_FTMSms2etd" in sample_id:
            sample_id = sample_id[:-11]
        if sample_id in sample_dict.keys():
            sample_dict[sample_id].append(line)
        else:
            sample_dict[sample_id] = [line]

#make a dictionary with all the accession numbers within each id
uniprot_dict = {}

print sample_dict.keys()
for id in sample_dict.keys():
    for line in sample_dict[id]:
        uniprotID = line[uniprot_index]
        if id in uniprot_dict.keys():
            if uniprotID not in uniprot_dict[id]:
                uniprot_dict[id].append(uniprotID)
        else:
            if uniprotID not in uniprot_dict.keys():

```

```

        uniprot_dict[id] = [uniprotID]

#add the intensity of all the peptides corresponding to an individual protein
#each line that is read is set aside in a new list
#before processing each line, it makes sure that the line has not been previously processed
#by checking it against the already processed list. (Guards against duplicates)

begin_new = header.index("Acc #")
intensity_index = header.index("Intensity")
already_checked = {}

intensities = {}
all_unique_peptides = {}

for sample in sample_dict.keys():
    #print sample_dict.keys()
    intensities[sample] = {}
    all_unique_peptides[sample] = {}
    already_checked[sample] = []
    #print uniprot_dict[sample]
    for id in uniprot_dict[sample]:
        #print uniprot_dict[sample]
        #intensities[sample][id] = 0.
        #print "Set to 0"
        #print all_unique_peptides[sample]
        all_unique_peptides[sample][id] = []
        #print all_unique_peptides[sample]
        #print intensities
        for line in sample_dict[sample]:
            #print already_checked
            #print line[begin_new:]
            #print line[uniprot_index]
            #print id
            #if line[begin_new:] not in already_checked[sample]:
            #print "Made it in!"
            if line[uniprot_index] == id:
                print line
                already_checked[sample].append(line[begin_new:])

                intensity = line[intensity_index]
                if intensity == "":
                    intensity = 0.
                else:
                    intensity = float(intensity)
                try:
                    intensities[sample][id] += intensity
                except KeyError:
                    intensities[sample][id] = intensity
                    #print id
                    #print intensity
                    #print intensities[sample][id]
                all_unique_peptides[sample][id].append(line)
        #print uniprot_dict[sample]
        #print intensities[sample]

#print intensities
#go through the uniprotIDs
#use these to go through the sample dictionary
#update the first instance of each uniprotID with the sum of the intensities calculated in the previous loop
#break out of it

##Realistically want to also include the number of unique peptides identified for each protein!!
##Need to keep track of this as well!
"""
protein_sample_dict = {}
for sample in uniprot_dict.keys():
    protein_sample_dict[sample] = []
    for id in uniprot_dict[sample]:

```

```

        for line in sample_dict[sample]:
            if id == line[uniprot_index]:
                protein_line = line
                for each in intensity_indices:
                    protein_line[each] = intensities[sample][id]
                protein_sample_dict[sample].append(protein_line)
        """

#output each sample as its own file.
#Output the Sample#, UniprotID, Intensity, and number unique for each file
#all_unique_peptides contains all instances identified of a peptides matching to a single
#protein. The sum of the elements in this structure is the spectral counts of the peptides
#The number of kinds of peptides uniquely identified can be determine by checking for the
#uniqueness of each entry. Peptides that have different variable modifications are counted
#as being different.

#get the number of the kinds of peptides that have been uniquely identified.
pep_seq_num = header.index("DB Peptide")
try:
    var_mod_num = header.index("Variable Mods")
except ValueError:
    var_mod_num = header.index("Protein Mods")
unique_kinds_of_peptides = {}
for each in all_unique_peptides.keys():
    unique_kinds_of_peptides[each] = {}
    for id in all_unique_peptides[each].keys():
        entry = all_unique_peptides[each][id]
        print entry
        for line in entry:
            print line
            pep_seq = line[pep_seq_num]
            var_mod = line[var_mod_num]
            pepvar = (pep_seq, var_mod)
            try:
                if pepvar not in unique_kinds_of_peptides[each][id]:
                    unique_kinds_of_peptides[each][id].append(pepvar)
            except KeyError:
                unique_kinds_of_peptides[each][id] = [pepvar]

print decoder.keys()
for sample in intensities.keys():
    with open("%s %s" % (decoder[sample].strip(), outputpath), 'w') as csvfile:
        f = csv.writer(csvfile)
        output_header = ["Sample ID", "Uniprot ID", "Unique Peptides", "Spectral Counts", "Protein Intensity"]
        f.writerow(output_header)
        for uniprot in intensities[sample].keys():
            protein_intensity = intensities[sample][uniprot]
            number_unique = len(unique_kinds_of_peptides[sample][uniprot])
            spectral_counts = len(all_unique_peptides[sample][uniprot])
            output_line = [sample, uniprot, number_unique, spectral_counts, protein_intensity]
            f.writerow(output_line)

date = datetime.datetime.today().strftime("%Y%m%d")
with open("%s_Sample_List.txt" % date, 'w') as txtfile:
    for sample in intensities.keys():
        outputline = "%s %s\n" % (decoder[sample], outputpath)
        txtfile.write(outputline)
    print outputline

```


5. batch_exp_fit_triplicates.py

#Use this script to fit triplicate intensity data to a single exponential to calculate half-life

```
#Assumes that the data is already organized as:
#[uniprot, 0h, 0h, 0h, 1h, 1h, 1h, 3h, 3h, 3h, 6h, 6h, 6h]

from lmfit.models import ExponentialModel
import matplotlib.pyplot as plt
import sys, csv, math
from matplotlib.backends.backend_pdf import PdfPages
import numpy as np

inputpath = sys.argv[1]
outputpath = sys.argv[2]
figname = sys.argv[3]

protein_int = []
#import the file
with open(inputpath, 'rU') as csvfile:
    csvreader = csv.reader(csvfile)
    for line in csvreader:
        if "Uniprot ID" not in line:
            protein_int.append(line)

#Store figures until ready
pp = PdfPages(figname)

def r_squared(results, data):
    res = results.residual
    ss_res = 0.

    for each in res:
        ss_res += each*each

    tot_mean = np.mean(data)

    ss_tot = 0.

    for each in data:
        ss_tot += (each - tot_mean)*(each - tot_mean)

    r_sq = 1 - ss_res/ss_tot

    return r_sq

def get_stats(x_axis, y_axis, fig_storage, plotname):
    #print x_axis
    #print y_axis
    mod = ExponentialModel()
    pars = mod.guess(y_axis, x=x_axis, decay=1)
    out = mod.fit(y_axis, pars, x=x_axis)

    results = out.best_values
    #print results.keys()
    decay = out.params['decay'].value
    hl = 0.69314 * decay # 1/decay to get k and ln(2)/k to get half-life
    hl_err = 0.69314 * out.params['decay'].stderr
    amp = out.params['amplitude'].value

    r_sq = r_squared(out, y_axis)

    #to make x coordinates for smooth fit.
    large_x = np.linspace(min(x), max(x), len(x)*100)

    #to make y coordinates for smooth fit
    fit_y = []
    for each in large_x:
        fit_y.append(amp * math.exp(-1*(1/decay)*each))
```

```

f = plt.figure()
plt.plot(x_axis,y_axis, 'bo')
ax = f.get_axes()[0]
f.text(0.6, 0.8, 'half-life = %s +/- %s\nr-squared = %s' % (round(hl,2), round(hl_err,2), round(r_sq,2)),
horizontalalignment='left',
verticalalignment='top',
transform=ax.transAxes)
plt.plot(large_x, fit_y, 'r')
plt.title(plotname)
plt.xlabel('Time (h)')
plt.ylabel('Intensity')
plt.ylim(ymin=0)
plt.xlim(xmin=-0.25, xmax=6.25)
pp.savefig(f)
return hl, hl_err, r_sq

#take each line
#convert the intensities from log-space
#feed in to the half life function
#store the results in a new list
x = [0,0,0,1,1,1,3,3,3,6,6,6]
hl_list = []
halflife_error = []
percent_err = []
r_sq_list = []
for line in protein_int:
    print line
    log_int = line[2:]
    norm_int = []
    for each in log_int:
        #print each
        if each == "#N/A" or each == "NA" or each == "":
            norm_int.append(2**-1)
        else:
            norm_int.append(2**float(each))
    halflife, hl_error, r_sq = get_stats(x, norm_int, pp, line[1])
    hl_list.append([line[0], line[1], halflife])
    percent_err.append([line[0], line[1], (hl_error/halflife)*100])
    r_sq_list.append([line[0], line[1], r_sq])
    halflife_error.append([line[0], line[1], halflife, hl_error])

pp.close()

with open("Halflife %s" % outputpath, 'wb') as csvfile:
    csvwriter = csv.writer(csvfile)
    for line in hl_list:
        csvwriter.writerow(line)

with open("Percent Error %s" % outputpath, 'wb') as csvfile:
    csvwriter = csv.writer(csvfile)
    for line in percent_err:
        csvwriter.writerow(line)

with open("R-squared %s" % outputpath, 'wb') as csvfile:
    csvwriter = csv.writer(csvfile)
    for line in r_sq_list:
        csvwriter.writerow(line)

with open("Halflife and Error %s" % outputpath, 'wb') as csvfile:
    csvwriter = csv.writer(csvfile)
    for line in halflife_error:
        csvwriter.writerow(line)

```

Appendix C: Human proteins with cofactors $\leq 5\text{\AA}$ from a Lys e-amine

Supplementary Table 1: Human proteins with cofactors $\leq 5\text{\AA}$ from a Lys e-amine

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
A2RUC4	tRNA wybutosine-synthesizing protein 5	TYW5	3AL6	AKG	1	['LYS B 175']	['K175']
O14607	Histone demethylase UTY	UTY	3ZLI	AKG	1	['LYS A1084']	['K1084']
O14832	Phytanoyl-CoA dioxygenase, peroxisomal	PHYH	2A1X	AKG	1	['LYS A 120']	['K120']
O15054	Lysine-specific demethylase 6B	KDM6B	2XUE	AKG	1	['LYS A1381']	['K1381']
O15550	Lysine-specific demethylase 6A	KDM6A	3AVR	OGA	1	['LYS A1137']	['K1137']
O75151	Lysine-specific demethylase PHF2	PHF2	3PU3	OGA	1	['LYS B 266']	['K266']
O75164	Lysine-specific demethylase 4A	KDM4A	2GP5	AKG	1	['LYS A 206']	['K206']
O75874	Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1	3INM	AKG	1	['LYS B 212']	['K212']
O75936	Gamma-butyrobetaine dioxygenase	BBOX1	3MS5	OGA	0	['None']	[]
P0C870	JmjC domain-containing protein 7	JMJD7	5NFN	AKG	1	['LYS A 193']	['K193']
P29375	Lysine-specific demethylase 5A	KDM5A	5E6H	AKG	1	['LYS A 501']	['K501']
Q12797	Aspartyl/asparaginyl beta-hydroxylase	ASPH	5JZA	OGA	0	['None']	[]
Q5SRE7	Phytanoyl-CoA dioxygenase domain-containing protein 1	PHYHD1	3OBZ	AKG	1	['LYS A 145']	['K145']
Q6B0I6	Lysine-specific demethylase 4D	KDM4D	3DXU	OGA	1	['LYS A 210']	['K210']
Q6N021	Methylcytosine dioxygenase TET2	TET2	4NM6	OGA	0	['None']	[]
Q6NS38	DNA oxidative demethylase ALKBH2	ALKBH2	3BU0	AKG	0	['None']	[]
Q6NYC1	Bifunctional arginine demethylase and lysyl-hydroxylase JMJD6	JMJD6	3LDB	AKG	1	['LYS A 204']	['K204']
Q6P6C2	RNA demethylase ALKBH5	ALKBH5	4NRP	OGA	1	['LYS A 132']	['K132']
Q6ZMT4	Lysine-specific demethylase 7A	KDM7A	3KV5	OGA	1	['LYS A 299']	['K299']
Q7LBC6	Lysine-specific demethylase 3B	KDM3B	4C8D	OGA	1	['LYS A1699']	['K1699']
Q8IU8	Ribosomal oxygenase 2	RIOX2	2XDV	OGA	1	['LYS A 194']	['K194']
Q8N371	JmjC domain-containing protein 5	KDM8	3UYJ	AKG	1	['LYS A 336']	['K336']
Q8N543	Prolyl 3-hydroxylase OGFOD1	OGFOD1	4NHX	OGA	0	['None']	[]
Q8N5Z0	Kynurenine/alpha-aminoadipate aminotransferase, mitochondrial	AADAT	3DC1	AKG	0	['None']	[]
Q96Q83	Alpha-ketoglutarate-dependent dioxygenase alkB homolog 3	ALKBH3	2IUW	AKG	0	['None']	[]
Q9BT30	Alpha-ketoglutarate-dependent dioxygenase alkB homolog 7, mitochondrial	ALKBH7	4QKB	AKG	0	['None']	[]
Q9C0B1	Alpha-ketoglutarate-dependent dioxygenase FTO	FTO	3LFM	OGA	0	['None']	[]
Q9GZT9	Egl nine homolog 1	EGLN1	3HQR	OGA	0	['None']	[]
Q9H3R0	Lysine-specific demethylase 4C	KDM4C	2XML	OGA	1	['LYS A 208']	['K208']
Q9H6W3	Ribosomal oxygenase 1	RIOX1	4CCK	OGA	1	['LYS A 355']	['K355']
Q9NWT6	Hypoxia-inducible factor 1-alpha inhibitor	HIF1AN	1H2N	AKG	1	['LYS A 214']	['K214']
Q9UGL1	Lysine-specific demethylase 5B	KDM5B	5A1F	OGA	1	['LYS A 517']	['K517']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
Q9UPP1	Histone lysine demethylase PHF8	PHF8	3KV4	OGA	1	['LYS A 264']	['K300']
Q9Y2K7	Lysine-specific demethylase 2A	KDM2A	2YU1	AKG	1	['LYS A 229']	['K229']
O00139	Kinesin-like protein KIF2A	KIF2A	2GRY	ADP	1	['LYS A 292']	['K319']
O00571	ATP-dependent RNA helicase DDX3X	DDX3X	5E7M	ANP	2	['LYS A 230', 'LYS A 288']	['K230', 'K288']
O00764	Pyridoxal kinase	PD XK	2AJP	ANP	1	['LYS A 225']	['K225']
O14656	Torsin-1A	TOR1A	5J1S	ATP	3	['LYS A 108', 'LYS A 320', 'LYS A 113']	['K108', 'K320', 'K113']
O14727	Apoptotic protease-activating factor 1	APAF1	1Z6T	ADP	1	['LYS A 160']	['K160']
O14782	Kinesin-like protein KIF3C	KIF3C	3B6V	ADP	1	['LYS A 103']	['K103']
O14994	Synapsin-3	SYN3	2P0A	ANP	4	['LYS A 204', 'LYS A 352', 'LYS A 258', 'LYS A 248']	['K204', 'K352', 'K258', 'K248']
O15066	Kinesin-like protein KIF3B	KIF3B	3B6U	ADP	1	['LYS B 102']	['K102']
O43143	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	DHX15	5XDR	ADP	1	['LYS A 166']	['K166']
O43252	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 1	PAPSS1	2OFX	ADP	2	['LYS A 65', 'LYS A 171']	['K65', 'K171']
O43314	Inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase 2	PPIP5K2	3T54	ATP	2	['LYS A 248', 'LYS A 187']	['K248', 'K187']
O60306	RNA helicase aquarius	AQR	4PJ3	ANP	2	['LYS A 829', 'LYS A 865']	['K829', 'K865']
O60825	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 2	PFKFB2	5HTK	ATP	1	['LYS A 51']	['K51']
O75191	Xylulose kinase	XYLB	4BC2	ADP	0	['None']	[]
O75417	DNA polymerase theta	POLQ	5AGA	ANP	1	['LYS A 121']	['K121']
O75643	U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	4F93	ADP	2	['LYS B 509', 'LYS B1356']	['K509', 'K1356']
O94762	ATP-dependent DNA helicase Q5	RECQL5	5LB3	ADP	2	['LYS B 58', 'LYS B 30']	['K58', 'K30']
O95340	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 2	PAPSS2	2AX4	ADP	2	['LYS A 65', 'LYS A 171']	['K55', 'K161']
O95786	Probable ATP-dependent RNA helicase DDX58	DDX58	4AY2	ADP	1	['LYS A 270']	['K270']
P00558	Phosphoglycerate kinase 1	PGK1	2WZB	ADP	2	['LYS A 219', 'LYS A 215']	['K220', 'K216']
P06746	DNA polymerase beta	POLB	8ICN	ATP	0	['None']	[]
P07814	Bifunctional glutamate/proline--tRNA ligase	EPRS	4HVC	ANP	2	['LYS A1156', 'LYS B1477']	['K1156', 'K1477']
P07900	Heat shock protein HSP 90-alpha	HSP90AA1	3T0Z	ATP	2	['LYS A 112', 'LYS A 58']	['K112', 'K58']
P10153	Non-secretory ribonuclease	RNASE2	1HI5	ADP	1	['LYS A 38']	['K65']
P11021	Endoplasmic reticulum chaperone BiP	HSPA5	3LDL	ATP	2	['LYS A 96', 'LYS A 296']	['K96', 'K296']
P11142	Heat shock cognate 71 kDa protein	HSPA8	3FZF	ATP	1	['LYS A 271']	['K271']
P11388	DNA topoisomerase 2-alpha	TOP2A	1ZX M	ANP	2	['LYS A 378', 'LYS A 168']	['K378', 'K168']
P12277	Creatine kinase B-type	CKB	3B6R	ADP	0	['None']	[]
P12883	Myosin-7	MYH7	4DB1	ANP	1	['LYS A 184']	['K184']
P13569	Cystic fibrosis transmembrane conductance regulator	CFTR	1XMI	ATP	1	['LYS A 464']	['K464']
P14618	Pyruvate kinase PKM	PKM	4FXF	ATP	2	['LYS D 270', 'LYS D 207']	['K270', 'K207']
P15104	Glutamine synthetase	GLUL	2OJW	ADP	1	['LYS E 43']	['K43']
P15531	Nucleoside diphosphate kinase A	NME1	1UCN	ADP	1	['LYS A 12']	['K12']
P16118	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 1	PFKFB1	1K6M	AGS	2	['LYS A 54', 'LYS A 174']	['K55', 'K175']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
P17066	Heat shock 70 kDa protein 6	HSPA6	3FE1	ADP	1	['LYS A 273']	['K273']
P17540	Creatine kinase S-type, mitochondrial	CKMT2	4Z9M	ADP	0	['None']	[]
P17844	Probable ATP-dependent RNA helicase DDX5	DDX5	3FE2	ADP	1	['LYS A 144']	['K144']
P19367	Hexokinase-1	HK1	1DGK	ADP	2	['LYS N 866', 'LYS N 429']	['K866', 'K429']
P22102	Trifunctional purine biosynthetic protein adenosine-3 [Includes: Phosphoribosylamine--glycine ligase]	GART	2QK4	ATP	2	['LYS A 148', 'LYS B 162']	['K148', 'K162']
P23381	Tryptophan--tRNA ligase, cytoplasmic	WARS	2QUI	ATP	2	['LYS A 349', 'LYS A 200']	['K349', 'K200']
P23677	Inositol-trisphosphate 3-kinase A	ITPKA	1W2C	ANP	3	['LYS A 264', 'LYS A 209', 'LYS A 336']	['K264', 'K209', 'K336']
P23919	Thymidylate kinase	DTYMK	1E2D	ADP	1	['LYS A 19']	['K19']
P27707	Deoxycytidine kinase	DCK	1P5Z	ADP	1	['LYS B 34']	['K34']
P31327	Carbamoyl-phosphate synthase [ammonia], mitochondrial	CPS1	5DOU	ADP	0	['None']	[]
P33176	Kinesin-1 heavy chain	KIF5B	1BG2	ADP	1	['LYS A 91']	['K91']
P33527	Multidrug resistance-associated protein 1	ABCC1	2CBZ	ATP	1	['LYS A 684']	['K684']
P34931	Heat shock 70 kDa protein 1-like	HSPA1L	3GDQ	ADP	1	['LYS A 273']	['K273']
P35557	Glucokinase	GCK	3FGU	ANP	1	['LYS A 169']	['K169']
P35790	Choline kinase alpha	CHKA	3G15	ADP	0	['None']	[]
P38919	Eukaryotic initiation factor 4A-III	EIF4A3	2HYI	ANP	1	['LYS C 88']	['K88']
P40692	DNA mismatch repair protein Mlh1	MLH1	4P7A	ADP	1	['LYS A 84']	['K84']
P41250	Glycine--tRNA ligase	GARS	2ZT7	ATP	0	['None']	[]
P42285	Exosome RNA helicase MTR4	MTREX	6C90	ADP	1	['LYS A 167']	['K167']
P43246	DNA mismatch repair protein Msh2	MSH2	2O8D	ADP	1	['LYS A 675']	['K675']
P46063	ATP-dependent DNA helicase Q1	RECQL	2V1X	ADP	1	['LYS A 119']	['K119']
P48637	Glutathione synthetase	GSS	2HGS	ADP	3	['LYS A 305', 'LYS A 364', 'LYS A 452']	['K305', 'K364', 'K452']
P49773	Histidine triad nucleotide-binding protein 1	HINT1	5O8I	ADP	0	['None']	[]
P49902	Cytosolic purine 5'-nucleotidase	NT5C2	2XCW	ATP	2	['LYS A 362', 'LYS A 359']	['K362', 'K359']
P49903	Selenide, water dikinase 1	SEPHS1	3FD6	ADP	1	['LYS A 32']	['K32']
P49914	5-formyltetrahydrofolate cyclo-ligase	MTHFS	3HY6	ADP	1	['LYS A 10']	['K10']
P49917	DNA ligase 4	LIG4	3W5O	ATP	5	['LYS A 273', 'LYS A 451', 'LYS A 449', 'LYS A 432', 'LYS B 345']	['K273', 'K451', 'K449', 'K432', 'K345']
P50053	Ketohexokinase	KHK	3NBV	ANP	1	['LYS B 193']	['K193']
P51553	Isocitrate dehydrogenase [NAD] subunit gamma, mitochondrial	IDH3G	5GRE	ADP	1	['LYS B 276']	['K315']
P51570	Galactokinase	GALK1	1WUU	ANP	0	['None']	[]
P51649	Succinate-semialdehyde dehydrogenase, mitochondrial	ALDH5A1	2W8R	ADP	1	['LYS A 228']	['K228']
P51795	H(+)/Cl(-) exchange transporter 5	CLCN5	2J9L	ATP	3	['LYS A 726', 'LYS B 587', 'LYS C 730']	['K726', 'K587', 'K730']
P52701	DNA mismatch repair protein Msh6	MSH6	2O8B	ADP	1	['LYS B1140']	['K1140']
P52732	Kinesin-like protein KIF11	KIF11	1I16	ADP	2	['LYS A 111', 'LYS A 34']	['K111', 'K34']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
P53396	ATP-citrate synthase	ACLY	3PFF	ADP	1	['LYS A 58']	['K58']
P54132	Bloom syndrome protein	BLM	4CDG	ADP	2	['LYS A 695', 'LYS A 726']	['K695', 'K726']
P54278	Mismatch repair endonuclease PMS2	PMS2	1H7U	AGS	1	['LYS A 86']	['K86']
P54652	Heat shock-related 70 kDa protein 2	HSPA2	3I33	ADP	1	['LYS A 274']	['K274']
P54922	[Protein ADP-ribosylarginine] hydrolase	ADPRH	3HFW	ADP	0	['None']	[]
P55072	Transitional endoplasmic reticulum ATPase	VCP	4KO8	AGS	2	['LYS A 251', 'LYS B 236']	['K251', 'K236']
P56373	P2X purinoceptor 3	P2RX3	5SVK	ATP	1	['LYS A 299']	['K299']
P67775	Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	PPP2CA	4LAC	AGS	0	['None']	[]
P78356	Phosphatidylinositol 5-phosphate 4-kinase type-2 beta	PIP4K2B	3X03	ANP	6	['LYS A 214', 'LYS A 150', 'LYS A 96', 'LYS A 94', 'LYS A 76', 'LYS B 239']	['K214', 'K150', 'K96', 'K94', 'K76', 'K239']
Q01415	N-acetylgalactosamine kinase	GALK2	2A2C	ADP	1	['LYS A 234']	['K234']
Q01813	ATP-dependent 6-phosphofructokinase, platelet type	PFKP	4U1R	ATP	0	['None']	[]
Q02224	Centromere-associated protein E	CENPE	1T5C	ADP	1	['LYS A 92']	['K92']
Q03518	Antigen peptide transporter 1	TAP1	1JJ7	ADP	1	['LYS A 544']	['K604']
Q06830	Peroxisome oxidoreductin-1	PRDX1	3HY2	ATP	1	['LYS A 93']	['K93']
Q08211	ATP-dependent RNA helicase A	DHX9	3LLM	ADP	1	['LYS A 417']	['K417']
Q08AH3	Acyl-coenzyme A synthetase ACSM2A, mitochondrial	ACSM2A	3C5E	ATP	2	['LYS A 229', 'LYS A 557']	['K229', 'K557']
Q10588	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 2	BST1	1ISG	AGS	0	['None']	[]
Q12931	Heat shock protein 75 kDa, mitochondrial	TRAP1	5HPH	ANP	1	['LYS A 126']	['K126']
Q13206	Probable ATP-dependent RNA helicase DDX10	DDX10	2PL3	ADP	1	['LYS A 119']	['K119']
Q13232	Nucleoside diphosphate kinase 3	NME3	1ZS6	ADP	1	['LYS A 29']	['K29']
Q13572	Inositol-tetrakisphosphate 1-kinase	ITPK1	2Q7D	ANP	3	['LYS A 199', 'LYS A 237', 'LYS A 157']	['K199', 'K237', 'K157']
Q13838	Spliceosome RNA helicase DDX39B	DDX39B	1XTJ	ADP	1	['LYS A 95']	['K95']
Q14807	Kinesin-like protein KIF22	KIF22	3BFN	ADP	1	['LYS A 133']	['K133']
Q15046	Lysine-tRNA ligase	KARS	3BJU	ATP	0	['None']	[]
Q15257	Serine/threonine-protein phosphatase 2A activator	PTPA	2HV7	ADP	0	['None']	[]
Q15645	Pachytene checkpoint protein 2 homolog	TRIP13	5VQA	ATP	1	['LYS A 185']	['K185']
Q16875	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3	PFKFB3	2AXN	ADP	2	['LYS A 47', 'LYS A 168']	['K48', 'K169']
Q2M1P5	Kinesin-like protein KIF7	KIF7	2XT3	ADP	1	['LYS A 97']	['K100']
Q460N5	Poly [ADP-ribose] polymerase 14	PARP14	4D86	ADP	1	['LYS A1141']	['K1141']
Q9UQ90	Paraplegin	SPG7	2QZ4	ADP	1	['LYS A 355']	['K355']
P0DMV8	Heat shock 70 kDa protein 1A	HSPA1A	2E8A	ANP	2	['LYS A 71', 'LYS A 271']	['K71', 'K271']
Q6PIW4	Fidgetin-like protein 1	FIGNL1	3D8B	ADP	1	['LYS A 447']	['K447']
Q6ZT98	Tubulin polyglutamylase TTL7	TTL7	4YLR	ADP	3	['LYS A 107', 'LYS A 160', 'LYS A 201']	['K107', 'K160', 'K201']
Q8IYB8	ATP-dependent RNA helicase SUPV3L1, mitochondrial	SUPV3L1	3RC3	ANP	2	['LYS A 213', 'LYS A 243']	['K213', 'K243']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
Q8N884	Cyclic GMP-AMP synthase	CGAS	6CTA	ATP	3	['LYS A 414', 'LYS A 219', 'LYS A 439']	['K414', 'K219', 'K439']
Q8NFU5	Inositol polyphosphate multikinase	IPMK	5W2H	ADP	1	['LYS A 75']	['K75']
Q8NI77	Kinesin-like protein KIF18A	KIF18A	3LRE	ADP	1	['LYS A 119']	['K119']
Q8TBC4	NEDD8-activating enzyme E1 catalytic subunit	UBA3	2NVU	ATP	2	['LYS B2103', 'LYS B2126']	['K281']
Q8WYK0	Acyl-coenzyme A thioesterase 12	ACOT12	4MOB	ADP	0	['None']	[]
Q92900	Regulator of nonsense transcripts 1	UPF1	2GJK	ANP	2	['LYS A 498', 'LYS A 533']	['K509', 'K544']
Q969G6	Riboflavin kinase	RFK	1NB0	ADP	1	['LYS A 28']	['K21']
Q96EP1	E3 ubiquitin-protein ligase CHFR	CHFR	2XOC	ADP	0	['None']	[]
Q96EY8	Cob(II)yrinic acid a,c-diamide adenosyltransferase, mitochondrial	MMAB	2IDX	ATP	1	['LYS C 78']	['K78']
Q96MK3	Pseudokinase FAM20A	FAM20A	5WRS	ATP	2	['LYS A 129', 'LYS A 233']	['K129', 'K233']
Q96PN6	Adenylate cyclase type 10	ADCY10	4USW	ATP	1	['LYS A 144']	['K144']
Q99661	Kinesin-like protein KIF2C	KIF2C	4UBF	ADP	2	['LYS A 354', 'LYS B 371']	['K354', 'K371']
Q9BTU6	Phosphatidylinositol 4-kinase type 2-alpha	PI4K2A	4HNE	ADP	1	['LYS A 152']	['K152']
Q9BVA6	Adenosine monophosphate-protein transferase FICD	FICD	4U07	ATP	0	['None']	[]
Q9BVG8	Kinesin-like protein KIFC3	KIFC3	5WDE	ADP	1	['LYS A 534']	['K534']
Q9BW19	Kinesin-like protein KIFC1	KIFC1	2REP	ADP	1	['LYS A 416']	['K416']
Q9BYN0	Sulfiredoxin-1	SRXN1	1XW4	ADP	1	['LYS X 61']	['K61']
Q9BZ23	Pantothenate kinase 2, mitochondrial	PANK2	5E+26	ADP	1	['LYS A 224']	['K224']
Q9BZX2	Uridine-cytidine kinase 2	UCK2	1UJ2	ADP	1	['LYS A 33']	['K33']
Q9GZZ9	Ubiquitin-like modifier-activating enzyme 5	UBA5	3GUC	ANP	1	['LYS A 127']	['K127']
Q9H477	Ribokinase	RBKS	2FV7	ADP	0	['None']	[]
Q9H6E5	Speckle targeted PIP5K1A-regulated poly	TUT1	5WU4	ATP	0	['None']	[]
Q9H8M5	Metal transporter CNNM2	CNNM2	4IY0	ADP	0	['None']	[]
Q9H981	Actin-related protein 8	ACTR8	4FO0	ATP	2	['LYS A 288', 'LYS A 347']	['K288', 'K347']
Q9H999	Pantothenate kinase 3	PANK3	3SMS	ADP	1	['LYS A 24']	['K24']
Q9HA47	Uridine-cytidine kinase 1	UCK1	2UVQ	ADP	1	['LYS A 36']	['K36']
Q9HAQ2	Kinesin-like protein KIF9	KIF9	3NWN	ADP	2	['LYS A 99', 'LYS A 14']	['K99', 'K14']
Q9HBA0	Transient receptor potential cation channel subfamily V member 4	TRPV4	4DX2	ATP	2	['LYS A 197', 'LYS A 192']	['K197', 'K192']
Q9HD67	Unconventional myosin-X	MYO10	5I0H	ADP	1	['LYS A 163']	['K163']
Q9NP58	ATP-binding cassette sub-family B member 6, mitochondrial	ABCB6	3NH9	ATP	1	['LYS A 629']	['K629']
Q9NQG6	Mitochondrial dynamics protein MID51	MIEF1	4NXU	ADP	2	['LYS A 368', 'LYS A 326']	['K368', 'K326']
Q9NQX4	Unconventional myosin-Vc	MYO5C	4ZG4	ADP	1	['LYS B 167']	['K167']
Q9NRF8	CTP synthase 2	CTPS2	3IHL	ADP	1	['LYS A 16']	['K16']
Q9NS87	Kinesin-like protein KIF15	KIF15	4BN2	ADP	1	['LYS A 101']	['K115']
Q9NWW6	Nicotinamide riboside kinase 1	NMRK1	2QL6	ADP	2	['LYS A 16', 'LYS F 25']	['K16', 'K25']
Q9NYA1	Sphingosine kinase 1	SPHK1	3VZD	ADP	0	['None']	[]
Q9UBT2	SUMO-activating enzyme subunit 2	UBA2	1Y8Q	ATP	2	['LYS B 346', 'LYS B 72']	['K346', 'K72']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
Q9UGM6	Tryptophan--tRNA ligase, mitochondrial	WARS2	5EKD	ATP	2	['LYS A 229', 'LYS A 226']	['K229', 'K226']
Q9UHI6	Probable ATP-dependent RNA helicase DDX20	DDX20	2OXC	ADP	1	['LYS A 112']	['K112']
Q9UJ70	N-acetyl-D-glucosamine kinase	NAGK	2CH6	ADP	1	['LYS B 274']	['K274']
Q9UKV8	Protein argonaute-2	AGO2	3QX9	ATP	4	['LYS A 533', 'LYS A 570', 'LYS A 566', 'LYS B 525']	['K533', 'K570', 'K566', 'K525']
Q9UMR2	ATP-dependent RNA helicase DDX19B	DDX19B	3EWS	ADP	2	['LYS A 64', 'LYS A 144']	['K64', 'K144']
Q9UNA4	DNA polymerase iota	POLI	3G6V	ATP	2	['LYS A 77', 'LYS A 214']	['K102', 'K239']
Q9Y223	Bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase	GENE	2YHY	ADP	0	['None']	[]
Q9Y230	RuvB-like 2	RUVBL2	3UK6	ADP	1	['LYS A 83']	['K83']
Q9Y253	DNA polymerase eta	POLH	5EWG	ATP	1	['LYS A 231']	['K231']
Q9Y259	Choline/ethanolamine kinase	CHKB	3FEG	ADP	0	['None']	[]
Q9Y265	RuvB-like 1	RUVBL1	2C9O	ADP	1	['LYS A 76']	['K76']
Q9Y2R4	Probable ATP-dependent RNA helicase DDX52	DDX52	3DKP	ADP	2	['LYS A 215', 'LYS A 270']	['K215', 'K270']
Q9Y3D8	Adenylate kinase isoenzyme 6	AK6	5JZV	ADP	1	['LYS A 16']	['K16']
Q9Y6X9	MORC family CW-type zinc finger protein 2	MORC2	5OFA;5 OFB	ANP	4	['LYS B 427', 'LYS B 89', 'LYS B 105']; ['LYS A 86']	['K427', 'K89', 'K105']; ['K86']
Q9NTK5	Obg-like ATPase 1	OLA1	2OHF	ACP	1	['LYS A 35']	['LYS A 35']
Q9NRK6	ATP-binding cassette sub-family B member 10, mitochondrial	ABCB10	4AYT	ACP	1	['LYS A 533']	['LYS A 533']
O00194	Ras-related protein Rab-27B	RAB27B	2F7S	GDP	2	['LYS A 22', 'LYS A 134']	['K22', 'K134']
O00212	Rho-related GTP-binding protein RhoD	RHOD	2J1L	GDP	2	['LYS A 30', 'LYS A 130']	['K30', 'K130']
O00429	Dynamin-1-like protein	DNM1L	3W6O	GCP	2	['LYS A 38', 'LYS A 216']	['K38', 'K216']
O75628	GTP-binding protein REM 1	REM1	2NZJ	GDP	2	['LYS A 93', 'LYS A 196']	['K93', 'K196']
O95057	GTP-binding protein Di-Ras1	DIRAS1	2GF0	GDP	2	['LYS A 20', 'LYS A 122']	['K20', 'K122']
O95716	Ras-related protein Rab-3D	RAB3D	2GF9	GDP	3	['LYS A 35', 'LYS A 136', 'LYS A 167']	['K35', 'K136', 'K167']
P01111	GTPase NRas	NRAS	3CON	GDP	3	['LYS A 16', 'LYS A 117', 'LYS A 147']	['K16', 'K117', 'K147']
P01112	GTPase HRas	HRAS	1PLK	GTP	3	['LYS A 16', 'LYS A 117', 'LYS A 147']	['K16', 'K117', 'K147']
P01116	GTPase KRas	KRAS	4DSO	GSP	3	['LYS A 16', 'LYS A 117', 'LYS A 147']	['K16', 'K117', 'K147']
P08134	Rho-related GTP-binding protein RhoC	RHOC	2GCP	GSP	3	['LYS A 18', 'LYS A 118', 'LYS A 162']	['K18', 'K118', 'K162']
P08754	Guanine nucleotide-binding protein G	GNAI3	2IHB	GDP	2	['LYS A 46', 'LYS A 270']	['K46', 'K270']
P10114	Ras-related protein Rap-2a	RAP2A	2RAP	GTP	3	['LYS A 16', 'LYS A 117', 'LYS A 147']	['K16', 'K117', 'K148']
P10301	Ras-related protein R-Ras	RRAS	2FN4	GDP	3	['LYS A 42', 'LYS A 143', 'LYS A 174']	['K42', 'K143', 'K174']
P11233	Ras-related protein Ral-A	RALA	2A78	GDP	3	['LYS A 27', 'LYS A 128', 'LYS A 159']	['K27', 'K128', 'K159']
P15153	Ras-related C3 botulinum toxin substrate 2	RAC2	2W2V	GTP	2	['LYS A 16', 'LYS A 116']	['K16', 'K116']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
P18085	ADP-ribosylation factor 4	ARF4	1Z6X	GDP	2	[LYS A 30', 'LYS A 127']	['K30', 'K127']
P20337	Ras-related protein Rab-3B	RAB3B	3DZ8	GDP	3	[LYS A 35', 'LYS A 62', 'LYS A 136']	['K35', 'K62', 'K136']
P20338	Ras-related protein Rab-4A	RAB4A	1Z0K	GTP	2	[LYS A 21', 'LYS A 122']	['K26', 'K127']
P20339	Ras-related protein Rab-5A	RAB5A	1N6L	GTP	3	[LYS A 33', 'LYS A 134', 'LYS A 165']	['K33', 'K134', 'K165']
P20340	Ras-related protein Rab-6A	RAB6A	2GIL	GTP	3	[LYS A 26', 'LYS A 127', 'LYS A 158']	['K26', 'K127', 'K158']
P20591	Interferon-induced GTP-binding protein Mx1	MX1	4P4T	GDP	2	[LYS A 248', 'LYS A 279']	['K248', 'K279']
P21980	Protein-glutamine gamma-glutamyltransferase 2	TGM2	4PYG	GTP	0	['None']	[]
P22392	Nucleoside diphosphate kinase B	NME2	1NUE	GDP	1	[LYS A 12']	['K12']
P23258	Tubulin gamma-1 chain	TUBG1	1Z5V	GSP	0	['None']	[]
P30520	Adenylosuccinate synthetase isozyme 2	ADSS	2V40	GDP	4	[LYS A 43', 'LYS A 362', 'LYS A 447', 'LYS A 45']	['K43', 'K362', 'K447', 'K45']
P35558	Phosphoenolpyruvate carboxykinase, cytosolic [GTP]	PCK1	1KHB	GCP	2	[LYS A 290', 'LYS A 244']	['K290', 'K244']
P36404	ADP-ribosylation factor-like protein 2	ARL2	3DOE	GTP	2	[LYS A 29', 'LYS A 126']	['K29', 'K126']
P40616	ADP-ribosylation factor-like protein 1	ARL1	1UPT	GTP	2	[LYS A 30', 'LYS A 127']	['K30', 'K127']
P51149	Ras-related protein Rab-7a	RAB7A	1T91	GTP	2	[LYS A 21', 'LYS A 126']	['K21', 'K126']
P51151	Ras-related protein Rab-9A	RAB9A	1WMS	GDP	3	[LYS A 20', 'LYS A 125', 'LYS A 156']	['K20', 'K125', 'K156']
P55040	GTP-binding protein GEM	GEM	2CJW	GDP	2	[LYS A 88', 'LYS A 192']	['K88', 'K192']
P55042	GTP-binding protein RAD	RRAD	2DPX	GDP	2	[LYS A 104', 'LYS A 204']	['K104', 'K204']
P57735	Ras-related protein Rab-25	RAB25	2OIL	GDP	2	[LYS A 25', 'LYS A 126']	['K25', 'K126']
P57772	Selenocysteine-specific elongation factor	EEFSEC	5IZL	GCP	2	[LYS A 20', 'LYS A 147']	['K20', 'K147']
P60763	Ras-related C3 botulinum toxin substrate 3	RAC3	2C2H	GDP	2	[LYS A 16', 'LYS A 116']	['K16', 'K116']
P60953	Cell division control protein 42 homolog	CDC42	2WMO	GTP	1	[LYS B 16']	['K16']
P61006	Ras-related protein Rab-8A	RAB8A	4LHV	GDP	3	[LYS A 21', 'LYS A 122', 'LYS A 153']	['K21', 'K122', 'K153']
P61018	Ras-related protein Rab-4B	RAB4B	2O52	GDP	2	[LYS A 21', 'LYS A 122']	['K21', 'K122']
P61019	Ras-related protein Rab-2A	RAB2A	1Z0A	GDP	3	[LYS A 19', 'LYS A 120', 'LYS B 151']	['K19', 'K120', 'K151']
P61020	Ras-related protein Rab-5B	RAB5B	2HEI	GDP	3	[LYS A 33', 'LYS A 134', 'LYS A 165']	['K33', 'K134', 'K165']
P61106	Ras-related protein Rab-14	RAB14	4D0G	GTP	3	[LYS A 24', 'LYS A 125', 'LYS A 156']	['K24', 'K125', 'K156']
P61224	Ras-related protein Rap-1b	RAP1B	4DXA	GSP	2	[LYS A 16', 'LYS A 117']	['K16', 'K117']
P61586	Transforming protein RhoA	RHOA	3KZ1	GSP	3	[LYS E 18', 'LYS E 118', 'LYS E 162']	['K18', 'K118', 'K162']
P61587	Rho-related GTP-binding protein RhoE	RND3	1M7B	GTP	2	[LYS A 36', 'LYS A 136']	['K36', 'K136']
P62070	Ras-related protein R-Ras2	RRAS2	2ERY	GDP	3	[LYS A 27', 'LYS A 128', 'LYS A 159']	['K27', 'K128', 'K159']
P62330	ADP-ribosylation factor 6	ARF6	2A5D	GTP	2	[LYS A 26', 'LYS A 123']	['K26', 'K123']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
P62491	Ras-related protein Rab-11A	RAB11A	2D7C	GTP	2	[LYS A 24', 'LYS A 125']	['K24', 'K125']
P62745	Rho-related GTP-binding protein RhoB	RHOB	2FV8	GDP	2	[LYS A 18', 'LYS A 118']	['K18', 'K118']
P62820	Ras-related protein Rab-1A	RAB1A	3TKL	GTP	3	[LYS A 24', 'LYS A 125', 'LYS A 156']	['K24', 'K125', 'K156']
P62826	GTP-binding nuclear protein Ran	RAN	4OL0	GTP	3	[LYS A 23', 'LYS A 123', 'LYS A 152']	['K23', 'K123', 'K152']
P62834	Ras-related protein Rap-1A	RAP1A	4KVG	GTP	3	[LYS A 16', 'LYS A 117', 'LYS A 149']	['K16', 'K117', 'K149']
P63000	Ras-related C3 botulinum toxin substrate 1	RAC1	2FJU	GSP	2	[LYS A 16', 'LYS A 116']	['K16', 'K116']
P63092	Guanine nucleotide-binding protein G	GNAS	6AU6	GDP	2	[LYS A 53', 'LYS A 293']	['K53', 'K293']
P63096	Guanine nucleotide-binding protein G	GNAI1	1Y3A	GDP	3	[LYS A 46', 'LYS A 51', 'LYS A 270']	['K46', 'K51', 'K270']
P84077	ADP-ribosylation factor 1	ARF1	1HUR	GDP	2	[LYS A 30', 'LYS A 127']	['K30', 'K127']
P84085	ADP-ribosylation factor 5	ARF5	2B6H	GDP	2	[LYS A 30', 'LYS A 127']	['K30', 'K127']
Q05193	Dynamin-1	DNM1	2X2E	GDP	2	[LYS A 44', 'LYS A 206']	['K44', 'K206']
Q13637	Ras-related protein Rab-32	RAB32	4CYM	GCP	3	[LYS A 38', 'LYS A 144', 'LYS A 176']	['K38', 'K144', 'K176']
Q15019	Septin-2	SEPT2	2QNR	GDP	2	[LYS A 50', 'LYS A 183']	['K50', 'K183']
Q15382	GTP-binding protein Rheb	RHEB	1XTS	GTP	3	[LYS A 19', 'LYS A 120', 'LYS A 151']	['K19', 'K120', 'K151']
Q15907	Ras-related protein Rab-11B	RAB11B	4OJK	GDP	3	[LYS A 24', 'LYS A 41', 'LYS A 125']	['K24', 'K41', 'K125']
Q6DD88	Atlastin-3	ATL3	5VGR	GDP	1	[LYS A 73']	['K73']
Q86YS6	Ras-related protein Rab-43	RAB43	2HUP	GDP	3	[LYS A 31', 'LYS A 132', 'LYS A 164']	['K31', 'K132', 'K164']
Q8IVH4	Methylmalonic aciduria type A protein, mitochondrial	MMAA	2WWW	GDP	2	[LYS A 156', 'LYS A 290']	['K156', 'K290']
Q8IWA4	Mitofusin-1	MFN1	5GNS	GTP	2	[LYS A 88', 'LYS A 286']	['K88', 'K286']
Q8IX1	Mitochondrial Rho GTPase 2	RHOT2	5KUT	GDP	2	[LYS A 429', 'LYS A 525']	['K429', 'K525']
Q8IX12	Mitochondrial Rho GTPase 1	RHOT1	5KTY	GDP	5	[LYS A 454', 'LYS A 431', 'LYS A 528', 'LYS A 447', 'LYS A 427']	['K454', 'K431', 'K528', 'K447', 'K427']
Q8IYK8	GTP-binding protein REM 2	REM2	3CBQ	GDP	2	[LYS A 127', 'LYS A 230']	['K127', 'K230']
Q8IZ41	Ras and EF-hand domain-containing protein	RASEF	2P5S	GDP	3	[LYS A 554', 'LYS A 655', 'LYS A 692']	['K554', 'K655', 'K692']
Q8WUD1	Ras-related protein Rab-2B	RAB2B	2A5J	GDP	3	[LYS A 24', 'LYS A 125', 'LYS A 156']	['K19', 'K120', 'K151']
Q8WWP7	GTPase IMAP family member 1	GIMAP1	3V70	GDP	1	[LYS A 40']	['K40']
Q8WXF7	Atlastin-1	ATL1	3Q5D	GDP	1	[LYS A 80']	['K80']
Q92730	Rho-related GTP-binding protein Rho6	RND1	2CLS	GTP	2	[LYS A 26', 'LYS A 126']	['K26', 'K126']
Q92963	GTP-binding protein Rit1	RIT1	4KLZ	GDP	2	[LYS A 34', 'LYS A 135']	['K34', 'K135']
Q96A58	Ras-related and estrogen-regulated growth inhibitor	RERG	2ATV	GDP	2	[LYS A 19', 'LYS A 119']	['K19', 'K119']
Q96BM9	ADP-ribosylation factor-like protein 8A	ARL8A	1ZD9	GDP	3	[LYS A 32', 'LYS A 130', 'LYS A 164']	['K33', 'K131', 'K165']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
Q96HU8	GTP-binding protein Di-Ras2	DIRAS2	2ERX	GDP	3	['LYS A 20', 'LYS A 122', 'LYS A 153']	['K20', 'K122', 'K153']
Q96KC2	ADP-ribosylation factor-like protein 5B	ARL5B	1YZG	GDP	2	['LYS A 29', 'LYS A 126']	['K29', 'K126']
Q9H0F7	ADP-ribosylation factor-like protein 6	ARL6	2H57	GTP	2	['LYS A 30', 'LYS A 131']	['K30', 'K131']
Q9H0U4	Ras-related protein Rab-1B	RAB1B	411O	GDP	3	['LYS A 21', 'LYS A 122', 'LYS A 153']	['K21', 'K122', 'K153']
Q9NR31	GTP-binding protein SAR1a	SAR1A	2GAO	GDP	2	['LYS A 38', 'LYS A 135']	['K38', 'K135']
Q9NRW1	Ras-related protein Rab-6B	RAB6B	2FFQ	GSP	3	['LYS A 26', 'LYS A 127', 'LYS A 158']	['K26', 'K127', 'K158']
Q9NUV9	GTPase IMAP family member 4	GIMAP4	3LXX	GDP	2	['LYS A 43', 'LYS A 192']	['K43', 'K192']
Q9NVJ2	ADP-ribosylation factor-like protein 8B	ARL8B	2AL7	GDP	3	['LYS A 32', 'LYS A 130', 'LYS A 164']	['K33', 'K131', 'K165']
Q9NYN1	Ras-like protein family member 12	RASL12	3C5C	GDP	2	['LYS A 33', 'LYS A 135']	['K33', 'K135']
Q9UG22	GTPase IMAP family member 2	GIMAP2	2XTN	GTP	1	['LYS A 35']	['K35']
Q9UH03	Neuronal-specific septin-3	SEPT3	4Z54	GDP	3	['LYS A 36', 'LYS A 60', 'LYS A 170']	['K74', 'K98', 'K208']
Q9UHD8	Septin-9	SEPT9	5CYP	GSP	2	['LYS A 36', 'LYS A 170']	['K311', 'K445']
Q9UL25	Ras-related protein Rab-21	RAB21	1Z0I	GDP	3	['LYS A 32', 'LYS A 133', 'LYS A 164']	['K32', 'K133', 'K164']
Q9UQ16	Dynamin-3	DNM3	3L43	GDP	2	['LYS A 44', 'LYS A 206']	['K44', 'K206']
Q9Y3Z3	Deoxynucleoside triphosphate triphosphohydrolase SAMHD1	SAMHD1	4QG2	GTP	4	['LYS A 455', 'LYS C 523', 'LYS D 116', 'LYS A 377']	['K455', 'K523', 'K116', 'K377']
Q9Y689	ADP-ribosylation factor-like protein 5A	ARL5A	1Z6Y	GDP	2	['LYS A 29', 'LYS A 126']	['K29', 'K126']
O00214	Galectin-8	LGALS8	4HAN	NAD	0	['None']	[]
O14556	Glyceraldehyde-3-phosphate dehydrogenase, testis-specific	GAPDHS	3H9E	NAD	0	['None']	[]
O43175	D-3-phosphoglycerate dehydrogenase	PHGDH	2G76	NAD	0	['None']	[]
O43488	Aflatoxin B1 aldehyde reductase member 2	AKR7A2	2BP1	NDP	2	['LYS A 237', 'LYS A 106']	['K236', 'K105']
O43865	S-adenosylhomocysteine hydrolase-like protein 1	AHCYL1	3MTG	NAD	1	['LYS B 524']	['K524']
O60218	Aldo-keto reductase family 1 member B10	AKR1B10	1ZUA	NAP	3	['LYS X 22', 'LYS X 263', 'LYS X 78']	['K22', 'K263', 'K78']
Q9BUP3	Oxidoreductase HTATIP2	HTATIP2	2BKA	NDP	2	['LYS A 98', 'LYS A 147']	['K98', 'K147']
O60547	GDP-mannose 4,6 dehydratase	GMDS	1T2A	NDP	1	['LYS A 183']	['K183']
O60701	UDP-glucose 6-dehydrogenase	UGDH	2Q3E	NAI	2	['LYS A 279', 'LYS A 220']	['K279', 'K220']
O75828	Carbonyl reductase [NADPH] 3	CBR3	2HRB	NAP	1	['LYS A 198']	['K198']
O75874	Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1	1T09	NAP	2	['LYS A 72', 'LYS B 260']	['K72', 'K260']
O94788	Retinal dehydrogenase 2	ALDH1A2	4X2Q	NAD	1	['LYS A 192']	['K210']
O95154	Aflatoxin B1 aldehyde reductase member 3	AKR7A3	2CLP	NDP	2	['LYS A 237', 'LYS A 106']	['K208', 'K77']
O95831	Apoptosis-inducing factor 1, mitochondrial	AIFM1	4BUR	NAD	2	['LYS A 342', 'LYS B 177']	['K342', 'K177']
P00325	Alcohol dehydrogenase 1B	ADH1B	1DEH	NAD	1	['LYS A 228']	['K229']
P00326	Alcohol dehydrogenase 1C	ADH1C	1HT0	NAD	1	['LYS A 228']	['K229']
P00338	L-lactate dehydrogenase A chain	LDHA	1110	NAI	1	['LYS A 56']	['K57']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
P00352	Retinal dehydrogenase 1	ALDH1A1	4WB9	NAI	2	['LYS A 193', 'LYS A 353']	['K193', 'K353']
P00374	Dihydrofolate reductase	DHFR	1BOZ	NDP	2	['LYS A 54', 'LYS A 55']	['K55', 'K56']
P00390	Glutathione reductase, mitochondrial	GSR	1GRB	NAD	1	['LYS A 66']	['K110']
P04035	3-hydroxy-3-methylglutaryl-coenzyme A reductase	HMGCR	1DQA	NAP	1	['LYS A 691']	['K691']
P04040	Catalase	CAT	1DGF	NDP	2	['LYS A 237', 'LYS C 306']	['K237', 'K306']
P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	1U8F	NAD	0	['None']	[]
P05091	Aldehyde dehydrogenase, mitochondrial	ALDH2	1CW3	NAD	2	['LYS A 192', 'LYS A 352']	['K209', 'K369']
P07195	L-lactate dehydrogenase B chain	LDHB	1I0Z	NAI	0	['None']	[]
P07327	Alcohol dehydrogenase 1A	ADH1A	1HSO	NAD	1	['LYS A 228']	['K229']
P08319	Alcohol dehydrogenase 4	ADH4	3COS	NAD	1	['LYS A 234']	['K234']
P09417	Dihydropteridine reductase	QDPR	1HDR	NAD	1	['LYS A 153']	['K154']
P09622	Dihydropolyl dehydrogenase, mitochondrial	DLD	1ZMD	NAI	2	['LYS A 54', 'LYS G 242']	['K89', 'K277']
P11413	Glucose-6-phosphate 1-dehydrogenase	G6PD	2BH9	NAP	4	['LYS A 47', 'LYS A 403', 'LYS A 366', 'LYS A 238']	['K47', 'K403', 'K366', 'K238']
P11586	C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	1DIA	NAP	1	['LYS A 198']	['K198']
P11766	Alcohol dehydrogenase class-3	ADH5	1MA0	NAD	1	['LYS A 227']	['K228']
P12268	Inosine-5'-monophosphate dehydrogenase 2	IMPDH2	1NFB	NAD	0	['None']	[]
P13995	Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, mitochondrial	MTHFD2	5TC4	NAD	0	['None']	[]
P14061	Estradiol 17-beta-dehydrogenase 1	HSD17B1	1A27	NAP	2	['LYS A 195', 'LYS A 159']	['K196', 'K160']
P15121	Aldose reductase	AKR1B1	1ADS	NAP	3	['LYS A 21', 'LYS A 262', 'LYS A 77']	['K22', 'K263', 'K78']
P15428	15-hydroxyprostaglandin dehydrogenase [NAD(+)]	HPGD	2GDZ	NAD	1	['LYS A 155']	['K155']
P16152	Carbonyl reductase [NADPH] 1	CBR1	1WMA	NDP	2	['LYS A 14', 'LYS A 197']	['K15', 'K198']
P16435	NADPH--cytochrome P450 reductase	POR	3QE2	NAP	1	['LYS A 605']	['K602']
P17516	Aldo-keto reductase family 1 member C4	AKR1C4	2FVL	NAP	2	['LYS A 270', 'LYS A 84']	['K270', 'K84']
P21695	Glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic	GPD1	1WPQ	NAD	2	['LYS A 296', 'LYS A 120']	['K296', 'K120']
P23368	NAD-dependent malic enzyme, mitochondrial	ME2	1DO8	NAD	2	['LYS B1346', 'LYS B1156']	['K346', 'K156']
P23526	Adenosylhomocysteinase	AHCY	1A7A	NAD	2	['LYS B 426', 'LYS A 186']	['K426', 'K186']
P28845	Corticosteroid 11-beta-dehydrogenase isozyme 1	HSD11B1	1XU7	NDP	2	['LYS A 44', 'LYS A 187']	['K44', 'K187']
P28907	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1	CD38	2I65	NAD	1	['LYS A 129']	['K129']
P30043	Flavin reductase	BLVRB	1HDO	NAP	0	['None']	[]
P30838	Aldehyde dehydrogenase, dimeric NADP-preferring	ALDH3A1	4L2O	NAD	1	['LYS A 137']	['K138']
P31937	3-hydroxyisobutyrate dehydrogenase, mitochondrial	HIBADH	2I9P	NAD	2	['LYS A 284', 'LYS A 209']	['K284', 'K209']
P32322	Pyrroline-5-carboxylate reductase 1, mitochondrial	PYCR1	2IZZ	NAD	1	['LYS A 71']	['K71']
P35270	Sepiapterin reductase	SPR	1Z6Z	NAP	1	['LYS A 171']	['K174']
P40394	Alcohol dehydrogenase class 4 mu/sigma chain	ADH7	1AGN	NAD	1	['LYS A 228']	['K240']
P40926	Malate dehydrogenase, mitochondrial	MDH2	2DFD	NAD	0	['None']	[]
P42330	Aldo-keto reductase family 1 member C3	AKR1C3	1RY0	NAP	2	['LYS A 270', 'LYS A 84']	['K270', 'K84']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
P47895	Aldehyde dehydrogenase family 1 member A3	ALDH1A3	5FHZ	NAD	3	['LYS A 204', 'LYS A 364', 'LYS C 360']	['K204', 'K364', 'K360']
P48735	Isocitrate dehydrogenase [NADP], mitochondrial	IDH2	4JA8	NDP	1	['LYS A 112']	['K112']
P49327	Fatty acid synthase	FASN	4PIV	NDP	1	['LYS A1995']	['K1995']
P49366	Deoxyhypusine synthase	DHPS	1DHS	NAD	0	['None']	[]
P49419	Alpha-aminoacidic semialdehyde dehydrogenase	ALDH7A1	2J6L	NAI	1	['LYS A 191']	['K190']
P50213	Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	IDH3A	5YVT	NAI	2	['LYS A 69', 'LYS A 173']	['K96', 'K200']
P51659	Peroxisomal multifunctional enzyme type 2	HSD17B4	1ZBQ	NAD	1	['LYS A 168']	['K168']
P51857	3-oxo-5-beta-steroid 4-dehydrogenase	AKR1D1	3BUR	NAP	2	['LYS A 273', 'LYS A 87']	['K273', 'K87']
P52209	6-phosphogluconate dehydrogenase, decarboxylating	PGD	2JKV	NAP	3	['LYS B 261', 'LYS B 184', 'LYS B 38']	['K261', 'K184', 'K38']
P52895	Aldo-keto reductase family 1 member C2	AKR1C2	1IHI	NAP	2	['LYS A 270', 'LYS A 84']	['K270', 'K84']
P53004	Biliverdin reductase A	BLVRA	2H63	NAP	0	['None']	[]
P56545	C-terminal-binding protein 2	CTBP2	2OME	NAD	0	['None']	[]
Q00796	Sorbitol dehydrogenase	SORD	1PL6	NAD	0	['None']	[]
Q04828	Aldo-keto reductase family 1 member C1	AKR1C1	1MRQ	NAP	2	['LYS A 270', 'LYS A 84']	['K270', 'K84']
Q08257	Quinone oxidoreductase	CRYZ	1YB5	NAP	1	['LYS B 324']	['K324']
Q13303	Voltage-gated potassium channel subunit beta-2	KCNAB2	1ZSX	NAP	2	['LYS A 240', 'LYS A 104']	['K254', 'K118']
Q13363	C-terminal-binding protein 1	CTBP1	1MX3	NAD	0	['None']	[]
Q13423	NAD(P) transhydrogenase, mitochondrial	NNT	1DJL	NAP	1	['LYS A 999']	['K1042']
Q13630	GDP-L-fucose synthase	TSTA3	4E5Y	NDP	2	['LYS A 147', 'LYS B 44']	['K147', 'K44']
Q14376	UDP-glucose 4-epimerase	GALE	1EK5	NAD	2	['LYS A 92', 'LYS A 161']	['K92', 'K161']
Q14894	Ketimine reductase mu-crystallin	CRYM	2I99	NDP	0	['None']	[]
Q14914	Prostaglandin reductase 1	PTGR1	2Y05	NAP	2	['LYS A 178', 'LYS A 194']	['K178', 'K194']
Q16698	2,4-dienoyl-CoA reductase, mitochondrial	DECR1	1W6U	NAP	2	['LYS A 214', 'LYS A 92']	['K214', 'K92']
Q16836	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	HADH	1F0Y	NAD	2	['LYS A 293', 'LYS A 115']	['K305', 'K127']
Q16881	Thioredoxin reductase 1, cytoplasmic	TXNRD1	2J3N	NAP	1	['LYS A 315']	['K465']
Q53FA7	Quinone oxidoreductase PIG3	TP53I3	2OBY	NAP	3	['LYS A 177', 'LYS A 325', 'LYS A 193']	['K177', 'K325', 'K193']
Q658P3	Metalloreductase STEAP3	STEAP3	2VQ3	NAP	0	['None']	[]
Q99714	3-hydroxyacyl-CoA dehydrogenase type-2	HSD17B10	2O23	NAD	1	['LYS A 172']	['K172']
Q6UWP2	Dehydrogenase/reductase SDR family member 11	DHRS11	1XG5	NAP	2	['LYS A 208', 'LYS A 170']	['K208', 'K170']
Q6YN16	Hydroxysteroid dehydrogenase-like protein 2	HSDL2	3KVO	NAP	3	['LYS A 42', 'LYS A 172', 'LYS A 116']	['K42', 'K172', 'K116']
Q7Z4W1	L-xylulose reductase	DCXR	1PR9	NAP	2	['LYS A 17', 'LYS A 153']	['K17', 'K153']
Q8IXJ6	NAD-dependent protein deacetylase sirtuin-2	SIRT2	4RMG	NAD	1	['LYS A 287']	['K287']
Q8N335	Glycerol-3-phosphate dehydrogenase 1-like protein	GPD1L	2PLA	NAD	2	['LYS A 298', 'LYS A 122']	['K298', 'K122']
Q8N4P3	Guanosine-3',5'-bis (diphosphate) 3'-pyrophosphohydrolase MESH1	HDDC3	5VXA	NDP	3	['LYS A 25', 'LYS A 101', 'LYS A 97']	['K25', 'K101', 'K97']
Q8N4Q0	Prostaglandin reductase 3	ZADH2	2C0C	NAP	1	['LYS A 200']	['K209']

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
Q8N4T8	Carbonyl reductase family member 4	CBR4	4CQM	NAP	1	['LYS B 152']	['K152']
Q8N8N7	Prostaglandin reductase 2	PTGR2	2ZB7; 2ZB8	NDP	3	['LYS A 192', 'LYS A 209']; ['LYS A 340']	['K192', 'K209']; ['K340']
Q8NBZ7	UDP-glucuronic acid decarboxylase 1	UXS1	4GLL	NAD	2	['LYS A 235', 'LYS B 174']	['K235', 'K174']
Q8WVW3	Reticulon-4-interacting protein 1, mitochondrial	RTN4IP1	2VN8	NDP	1	['LYS A 256']	['K256']
Q92506	Estradiol 17-beta-dehydrogenase 8	HSD17B8	4CQL	NAD	2	['LYS A 209', 'LYS A 173']	['K209', 'K173']
Q96EB6	NAD-dependent protein deacetylase sirtuin-1	SIRT1	4I5I	NAD	0	['None']	[]
Q96HN2	Adenosylhomocysteinase 3	AHCYL2	3GVP	NAD	1	['LYS C 605']	['K605']
Q96T66	Nicotinamide/nicotinic acid mononucleotide adenylyltransferase 3	NMNAT3	1NUU	NAD	0	['None']	[]
Q9BPX1	17-beta-hydroxysteroid dehydrogenase 14	HSD17B14	5HS6	NAD	1	['LYS A 158']	['K158']
Q9BTZ2	Dehydrogenase/reductase SDR family member 4	DHRS4	3O4R	NAP	1	['LYS A 186']	['K186']
Q9BUT1	3-hydroxybutyrate dehydrogenase type 2	BDH2	2AG5	NAD	1	['LYS A 151']	['K151']
Q9HAN9	Nicotinamide/nicotinic acid mononucleotide adenylyltransferase 1	NMNAT1	1KQN	NAD	1	['LYS A 57']	['K57']
Q9HBL8	NmrA-like family domain-containing protein 1	NMRAL1	2EXX	NAP	3	['LYS B 92', 'LYS A 41', 'LYS A 133']	['K92', 'K41', 'K133']
Q9NRG7	Epimerase family protein SDR39U1	SDR39U1	4B4O	NDP	0	['None']	[]
Q9NTG7	NAD-dependent protein deacetylase sirtuin-3, mitochondrial	SIRT3	4BV3	NAD	0	['None']	[]
Q9NUI1	Peroxisomal 2,4-dienoyl-CoA reductase	DECR2	4FC6	NAP	1	['LYS A 182']	['K182']
Q9NXA8	NAD-dependent protein deacylase sirtuin-5, mitochondrial	SIRT5	3RIY	NAD	0	['None']	[]
Q9NZL9	Methionine adenosyltransferase 2 subunit beta	MAT2B	2YDX	NAP	1	['LYS A 163']	['K163']
Q9P2T1	GMP reductase 2	GMPR2	2C6Q	NDP	1	['LYS A 78']	['K78']
Q9UBK8	Methionine synthase reductase	MTRR	2QTZ	NAP	1	['LYS A 291']	['K318']
Q9UBQ7	Glyoxylate reductase/hydroxypyruvate reductase	GRHPR	2GCG	NDP	0	['None']	[]
Q9UDR5	Alpha-aminoadipic semialdehyde synthase, mitochondrial	AASS	5L78	NAD	0	['None']	[]
Q9Y2S2	Lambda-crystallin homolog	CRYL1	3F3S	NAD	2	['LYS A 104', 'LYS A 102']	['K104', 'K102']
O14686	Histone-lysine N-methyltransferase 2D	KMT2D	4Z4P	SAH	0	['None']	[]
O14717	tRNA (cytosine (38)-C(5))-methyltransferase	TRDMT1	1G55	SAH	0	['None']	[]
O14744	Protein arginine N-methyltransferase 5	PRMT5	4X61	SAM	2	['LYS A 333', 'LYS A 393']	['K333', 'K393']
O43148	mRNA cap guanine-N7 methyltransferase	RNMT	3BGV	SAH	1	['LYS A 16']	['K180']
O43159	Ribosomal RNA-processing protein 8	RRP8	2ZFU	SAH	0	['None']	[]
O60678	Protein arginine N-methyltransferase 3	PRMT3	2FYT	SAH	1	['LYS A 329']	['K312']
O95050	Indolethylamine N-methyltransferase	INMT	2A14	SAH	0	['None']	[]
O95568	Histidine protein methyltransferase 1 homolog	METTL18	4RFQ	SAM	0	['None']	[]
O96028	Histone-lysine N-methyltransferase NSD2	NSD2	5LSU	SAM	0	['None']	[]
P11086	Phenylethanolamine N-methyltransferase	PNMT	1HNN	SAH	0	['None']	[]
P21964	Catechol O-methyltransferase	COMT	3A7E	SAM	1	['LYS A 144']	['K194']
P22061	Protein-L-isoaspartate	PCMT1	111N	SAH	0	['None']	[]
P26358	DNA (cytosine-5)-methyltransferase 1	DNMT1	4WXX	SAH	0	['None']	[]

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
P31153	S-adenosylmethionine synthase isoform type-2	MAT2A	4KTT	SAM	2	['LYS B 289', 'LYS A 181']	['K289', 'K181']
P35520	Cystathionine beta-synthase	CBS	4UUU	SAM	0	['None']	[]
P40261	Nicotinamide N-methyltransferase	NNMT	2IIP	SAH	0	['None']	[]
P42898	Methylenetetrahydrofolate reductase	MTHFR	6FCX	SAH	0	['None']	[]
P46597	Acetylserotonin O-methyltransferase	ASMT	4A6D	SAM	0	['None']	[]
P50135	Histamine N-methyltransferase	HNMT	1JQD	SAH	0	['None']	[]
P51580	Thiopurine S-methyltransferase	TPMT	2BZG	SAH	0	['None']	[]
Q00266	S-adenosylmethionine synthase isoform type-1	MAT1A	2OBV	SAM	1	['LYS A 181']	['K181']
Q03164	Histone-lysine N-methyltransferase 2A	KMT2A	2W5Y	SAH	0	['None']	[]
Q13395	Probable methyltransferase TARBP1	TARBP1	2HA8	SAH	0	['None']	[]
Q14353	Guanidinoacetate N-methyltransferase	GAMT	3ORH	SAH	0	['None']	[]
Q15910	Histone-lysine N-methyltransferase EZH2	EZH2	5HYN	SAH	0	['None']	[]
Q4FZB7	Histone-lysine N-methyltransferase KMT5B	KMT5B	3S8P	SAM	0	['None']	[]
Q53H47	Histone-lysine N-methyltransferase SETMAR	SETMAR	3BO5	SAH	1	['LYS A 271']	['K285']
Q5T280	Putative methyltransferase C9orf114	SPOUT1	4RG1	SAH	0	['None']	[]
Q5T8I9	Small RNA 2'-O-methyltransferase	HENMT1	4XCX	SAH	1	['LYS A 83']	['K83']
Q5VVY1	Alpha N-terminal protein methyltransferase 1B	METTL11 B	5UBB	SAM	0	['None']	[]
Q5VZV1	Protein-lysine methyltransferase METTL21C	METTL21 C	4MTL	SAH	0	['None']	[]
Q7L0Y3	tRNA methyltransferase 10 homolog C	TRMT10C	5NFJ	SAM	0	['None']	[]
Q7L2J0	7SK snRNA methylphosphate capping enzyme	MEPCE	3G07	SAM	0	['None']	[]
Q7Z624	Calmodulin-lysine N-methyltransferase	CAMKMT	4PWY	SAH	0	['None']	[]
Q86TU7	Histone-lysine N-methyltransferase setd3	SETD3	3SMT	SAM	0	['None']	[]
Q86U44	N6-adenosine-methyltransferase catalytic subunit	METTL3	5IL1	SAM	1	['LYS A 513']	['K513']
Q86VU5	Catechol O-methyltransferase domain-containing protein 1	COMTD1	2AVD	SAM	0	['None']	[]
Q86W50	U6 small nuclear RNA (adenine-(43)-N(6))-methyltransferase	METTL16	2H00	SAH	0	['None']	[]
Q86X55	Histone-arginine methyltransferase CARM1	CARM1	2Y1X	SAH	0	['None']	[]
Q86Y97	Histone-lysine N-methyltransferase KMT5C	KMT5C	3RQ4	SAM	0	['None']	[]
Q8N1G2	Cap-specific mRNA (nucleoside-2'-O)-methyltransferase 1	CMTR1	4N49	SAM	1	['LYS A 404']	['K404']
Q8N6R0	Methyltransferase-like protein 13	METTL13	5WCJ	SAH	0	['None']	[]
Q8NEZ4	Histone-lysine N-methyltransferase 2C	KMT2C	5F59	SAH	0	['None']	[]
Q8TBK2	N-lysine methyltransferase SETD6	SETD6	3RC0	SAM	0	['None']	[]
Q8TBZ6	tRNA methyltransferase 10 homolog A	TRMT10A	4FMW	SAH	0	['None']	[]
Q8TEK3	Histone-lysine N-methyltransferase, H3 lysine-79 specific	DOT1L	1NW3	SAM	0	['None']	[]
Q8WTS6	Histone-lysine N-methyltransferase SETD7	SETD7	3CBM	SAH	1	['LYS A 294']	['K294']
Q8WXB1	Protein N-lysine methyltransferase METTL21A	METTL21 A	4LEC	SAH	0	['None']	[]
Q92979	Ribosomal RNA small subunit methyltransferase NEP1	EMG1	5FAI	SAH	0	['None']	[]
Q96AZ1	EEF1A lysine methyltransferase 3	EEF1AKM T3	4QPN	SAH	0	['None']	[]

Uniprot IDs	Protein Name	Gene Name	PDB Code	Lig. ID	Num Lys	Lysine Chain Residue# (from PDB file)	Uniprot Residue Number
Q96CB9	5-methylcytosine rRNA methyltransferase NSUN4	NSUN4	4FP9	SAM	1	['LYS A 187']	['K187']
Q96KQ7	Histone-lysine N-methyltransferase EHMT2	EHMT2	2O8J	SAH	1	['LYS B1162']	['K1162']
Q96L73	Histone-lysine N-methyltransferase, H3 lysine-36 and H4 lysine-20 specific	NSD1	3OOI	SAM	0	['None']	[]
Q96LA8	Protein arginine N-methyltransferase 6	PRMT6	4HC4	SAH	0	['None']	[]
Q96RS0	Trimethylguanosine synthase	TGS1	3EGI	ADP	1	['LYS A 724']	['K724']
P0DPD6	Endothelin-converting enzyme 2	ECE2	2PXX	SAH	1	['LYS A 130']	['K130']
Q9BV86	N-terminal Xaa-Pro-Lys N-methyltransferase 1	NTMT1	2EX4	SAH	0	['None']	[]
Q9BVS5	tRNA (adenine(58)-N(1))-methyltransferase	TRMT61B	2B25	SAM	0	['None']	[]
Q9BYW2	Histone-lysine N-methyltransferase SETD2	SETD2	4H12	SAH	0	['None']	[]
Q9BZ95	Histone-lysine N-methyltransferase NSD3	NSD3	4YZ8	SAM	0	['None']	[]
Q9H511	Histone-lysine N-methyltransferase SUV39H2	SUV39H2	2R3A	SAM	1	['LYS A 189']	['K300']
Q9H7B4	Histone-lysine N-methyltransferase SMYD3	SMYD3	5EX3	SAH	1	['LYS A 135']	['K135']
Q9H867	Protein-lysine methyltransferase METTL21D	VCPKMT	4LG1	SAM	1	['LYS A 125']	['K125']
Q9H9B1	Histone-lysine N-methyltransferase EHMT1	EHMT1	4I51	SAH	1	['LYS A1219']	['K1250']
Q9NQR1	N-lysine methyltransferase KMT5A	KMT5A	2BQZ	SAH	1	['LYS A 226']	['K267']
Q9NR22	Protein arginine N-methyltransferase 8	PRMT8	5DST	SAH	0	['None']	[]
Q9NR48	Histone-lysine N-methyltransferase ASH1L	ASH1L	4YPA	SAM	1	['LYS B2150']	['K2155']
Q9NRG4	N-lysine methyltransferase SMYD2	SMYD2	3RIB	SAH	1	['LYS B 17']	['K17']
Q96P11	Probable 28S rRNA (cytosine-C(5))-methyltransferase	NSUN5	2B9E	SAM	1	['LYS A 240']	['K240']
Q9UBP6	tRNA (guanine-N(7)-)-methyltransferase	METTL1	3CKK	SAM	0	['None']	[]
Q9UI43	rRNA methyltransferase 2, mitochondrial	MRM2	2NYU	SAM	1	['LYS A 144']	['K194']
Q9UIC8	Leucine carboxyl methyltransferase 1	LCMT1	3IEI	SAH	1	['LYS A 37']	['K37']
Q9UNQ2	Probable dimethyladenosine transferase	DIMT1	1ZQ9	SAM	0	['None']	[]
Q9Y6K1	DNA (cytosine-5)-methyltransferase 3A	DNMT3A	2QRV	SAH	0	['None']	[]
Q96FX7	tRNA (adenine(58)-N(1))-methyltransferase catalytic subunit TRMT61A	TRMT61A	5CCB	SAH	0	['None']	[]
Q9NQT8	Kinesin-like protein KIF13B	KIF13B	3GBJ	ADP	1	['LYS A 109']	['K109']
Q7L0Q8	Rho-related GTP-binding protein RhoU	RHOU	2Q3H	GDP	1	['LYS A 62']	['K62']
Q6IQ22	Ras-related protein Rab-12	RAB12	2IL1	GDP	2	['LYS A 151', 'LYS A 252']	['K55', 'K156']
P52298	Nuclear cap-binding protein subunit 2	NCBP2	1H2U	GDP	0	['None']	[]

Publishing Agreement

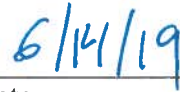
It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature



Date