

Designing Active Site-Directed Covalent Probes for Tyrosine Phosphatases

Suk ho Hong

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  
under the Executive Committee  
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2022

© 2022

Suk ho Hong

All Rights Reserved

## **Abstract**

Designing Active Site-Directed Covalent Probes for Tyrosine Phosphatases

Suk ho Hong

Tyrosine phosphorylation is an important post-translational modification in cells that modulates key cellular pathways. Tyrosine phosphatases are the class of enzymes that remove this modification from proteins, yet we know relatively little about how they are regulated in various signaling contexts. Activity-based probes that successfully target active sites of tyrosine phosphatases and report on their activities can fill in this gap. We show the assessment of various thiol-reactive groups for their ability to target catalytic cysteine residues with specificity. Then we construct and screen a library of fragment-like compounds for their on-target and off-target reactivities. We also discuss theoretical considerations for screening covalent inhibitors for their kinetic parameters and show this using our experimental data. Lastly, we augment compounds selected from the library to enable click chemistry for reporter group attachment for use on the whole proteome, ultimately through mass spectrometry-based proteomics methods. We show enrichment of target proteins. These target-centric design efforts will yield new insights into the general development processes of activity-based probes or inhibitors.

# Table of Contents

List of Charts, Graphs, Illustrations .....	iv
Acknowledgments.....	viii
Dedication .....	x
Chapter 1: Introduction .....	1
1.1. Protein post-translational modifications .....	1
1.2. Significance of protein phosphorylation .....	2
1.3. Protein tyrosine phosphatases .....	3
1.3.1 Function and classification of protein phosphatases.....	3
1.3.2 Structure and classification of protein tyrosine phosphatases .....	6
1.3.3 Catalytic mechanism of protein tyrosine phosphatases .....	9
1.3.4 Regulation of protein tyrosine phosphatases .....	12
1.4. Activity-based profiling .....	13
1.4.1 Activity-based protein profiling and activity-based probes.....	13
1.4.2 ABPP for phosphatases and design considerations .....	16
1.5. Overview of dissertation .....	18
Chapter 2: Evolution of vinylsulfonate probes for PTPs .....	20
2.1. Background .....	20
2.2. PVSN is highly reactive towards non-catalytic cysteines.....	22
2.3. Methylcoumaryl vinylsulfonate, a substrate-based design with higher reactivity.....	26
2.4. More vinylsulfonate compounds.....	30

2.5. Tagged PVSN can label proteins, but indiscriminately .....	32
2.6. Discussion .....	34
2.7. Methods.....	36
Chapter 3: Designing active site-directed probes .....	42
3.1. Background .....	42
3.2. Biochemical assays of various thiol-reactive groups.....	43
3.3. Screening a fragment-like library .....	48
3.4. Intrinsic reactivity assays.....	54
3.5. Two-dimensional analysis of the library.....	55
3.6. Further analyses and identifying false leads .....	57
3.7. Computational docking.....	60
3.8. Discussion .....	61
3.9. Methods.....	62
Chapter 4: Inactivation kinetics of irreversible inhibitors .....	66
4.1. Background .....	66
4.2. Approximations.....	67
4.3. Analytical solution of the irreversible inhibition system.....	70
4.4. Using different models to fit inactivation assay data.....	73
4.5. Analyzing library screening data for kinetic parameters .....	81
4.6. Discussion .....	88
4.7. Methods.....	88
Chapter 5: Utilizing active site-directed covalent probes in proteomics .....	90

5.1. Background .....	90
5.2. Design and synthesis of tagged probes .....	91
5.3. Gel-based ABPP methods .....	93
5.4. Mass spectrometry-based ABPP methods .....	98
5.5. Discussion .....	101
5.6. Methods.....	102
Chapter 6: Conclusion and future outlooks .....	111
6.1. Overview .....	111
6.2. Towards an ideal ABP .....	112
6.3. Future directions .....	113
References.....	116

## List of Charts, Graphs, Illustrations

Figure 1.1: Structures of phosphorylated amino acids.....	4
Figure 1.2: Classification of human PTPs and domain architectures of structures. ....	7
Figure 1.3: Structural motifs of a classical PTP bound to a ligand.....	8
Figure 1.4: Catalytic mechanism of protein tyrosine phosphatase. ....	10
Figure 1.5: Redox regulation of catalytic Cys residue.....	11
Figure 1.6: Regulation mechanisms of PTPs.....	12
Figure 1.7: Reported activity-based probes and their reaction mechanisms. ....	17
Figure 2.1: Reaction of PTP1B and PVSN as determined by intact protein MS.....	25
Figure 2.2: Example spectra for PVSN labeling site determination. ....	26
Figure 2.3 Commonly used substrates and their kinetic properties. ....	27
Figure 2.4: Inhibitory activities of PVSN, PVS, and substrate-inspired compounds. ....	29
Figure 2.5: PVSN and MUVSN labeling over time as monitored by intact protein MS.....	30
Figure 2.6: PTP inhibition and labeling by vinylsulfonate compounds.....	32
Figure 2.7: Reaction of purified SHP1/2 with azPVSN over time. ....	34
Figure 3.1: Thiol-reactive groups assessed for reactivity. ....	44
Figure 3.2: Inhibition of PTPs by phenyl compounds. ....	46
Figure 3.3: Degree of labeling compared with inhibition data. ....	47
Figure 3.4: Compound labeling sites for PTP1B as determined from peptide MS. ....	48
Figure 3.5: Fragment-like library.....	50
Figure 3.6: Screening the library for PTP inhibition. ....	52
Figure 3.7: Trends in individual PTPs. ....	53
Figure 3.8: Intrinsic reactivities of the library compounds.....	55

Figure 3.9: 2D analysis of IC <sub>50</sub> and intrinsic reactivity. ....	56
Figure 3.10: SAR of phenyl chloroacetamide with substituents.....	57
Figure 3.11: MS analyses of library compounds. ....	58
Figure 3.12: Aggregating compounds with and without detergent.....	58
Figure 3.13: Inhibition of compounds in the presence of a competitor. ....	59
Figure 4.1: Two ways of fitting for k <sub>obs</sub> , and the assessed compounds. ....	74
Figure 4.2: Secondary plot method for PTP1B and phenyl chloroacetamide.....	76
Figure 4.3: Secondary plot method for 3 PTPs and phenyl bromoacetamide. ....	78
Figure 4.4: Secondary plot method for YopH with 4 compounds.....	79
Figure 4.5: Global fit method for YopH with 4 compounds.....	80
Figure 4.6: Fitting library screen data to obtain k <sub>obs</sub> . ....	82
Figure 4.7: k <sub>inact</sub> , K <sub>I</sub> and covalent efficiency for individual PTPs. ....	84
Figure 4.8: Obtained constants plotted against IC <sub>50</sub> for each PTP. ....	85
Figure 4.9: k <sub>inact</sub> for each PTP compared to small molecule thiol reaction constant. ....	86
Figure 4.10: Example of a noisy data fitting.....	89
Figure 4.11: Example of a less noisy data fitting well without y <sub>0</sub> constraint. ....	89
Figure 5.1: Chasing with a broadly reactive dye conjugate.....	95
Figure 5.2: Dye chasing after direct cell treatment.....	96
Figure 5.3: In-gel ABPP using synthesized probes. ....	97
Figure 5.4: Enrichment of proteins from intact cells treated with probes. ....	101
Figure 5.5: Comparison of hexynyl-tagged probes with ethynyl- and propargylamide-tagged probes.....	104
Figure 5.6: Checking for background signals of click chemistry. ....	106



Figure 5.7: Optimization of bead washing protocol to minimize nonspecific binding. ....	108
Figure 5.8: An example of a sample blot during proteomics sample preparation. ....	108
Figure 5.9: Verification of MS proteomics protocol. ....	109
Figure 6.1: Potential improvements to functional parts of probes. ....	114
Scheme 1.1: Effects of tyrosine phosphorylation. ....	3
Scheme 1.2: A visual explanation of activity-based protein profiling. ....	14
Scheme 2.1: Workflow of intact protein MS experiments. ....	23
Scheme 2.2: Workflow of PTP inhibition assay with pNPP. ....	24
Scheme 2.3: Structures and leaving group pK <sub>a</sub> of more vinylsulfonate compounds. ....	31
Scheme 2.4: Synthesis and click chemistry of azPVSN. ....	33
Scheme 2.5: Balancing probe reactivity with affinity for target active sites. ....	35
Scheme 5.1 ABP structure and representative synthesis pathways. ....	93
Scheme 5.2: Workflow for MS proteomics-ABPP. ....	99
Table 3.1: Glide docking scores of the highest-ranking compounds. ....	60
Table 4.1: Kinetic parameters of YopH with 4 compounds as determined from different fit methods. ....	81
Table 4.2: All values of k <sub>inact</sub> , K <sub>I</sub> and covalent efficiency for individual PTPs. ....	86

## List of Abbreviations

SH2	Src Homology 2
PTP1B	Protein Tyrosine Phosphatase 1B
HePTP	Hematopoietic Protein Tyrosine Phosphatase
ERK	Extracellular signal-Regulated Kinase
VH1	<i>Vaccinia</i> virus H1 protein
Cdc14/25	Cell division cycle 14/25
PTEN	Phosphatase and TENsin homolog
LMWPTP	Low Molecular Weight Protein Tyrosine Phosphatase
SHP1/2	Src Homology 2 domain Phosphatase 1/2
PTPN11	Protein Tyrosine Phosphatase Non-receptor type 11
CD45	Cluster of Differentiation 45
PI3K	Phosphatidylinositol-3-Kinase
YopH	<i>Yersinia</i> outer protein H
Sts-1/2	Suppressor of T-cell signaling 1/2
HEK293	Human Embryonic Kidney 293
Cys, C	Cysteine
Ser, S	Serine
Thr	Threonine
His, H	Histidine
pTyr, pY	Phosphotyrosine

## Acknowledgments

I am ever grateful for all the goodwill and generosity I have received over the course of my program. Even though I cannot list everyone that I want to thank, they will always have a place in my heart.

First, I want to thank my advisor, Prof. Neel Shah, for helping me take on my doctorate research and supporting me throughout the program. I met him through a bit of luck and coincidence – and I am glad that I was able to work with him. I look forward to seeing the lab that he will be building on and the research that will continue to be produced by it. I also thank members of my advisory committee, Prof. Wei Min and Prof. Brent Stockwell, for their insightful comments and encouragement at every checkpoint of my program.

Moving to the other side of the world to pursue higher education can be daunting, and the people I met are what kept me pushing forward. Especially, the years I spent at Columbia would not have been as fun and exciting if it were not for the people of the Shah Lab. Building a lab together from scratch was an invaluable experience to have, and I had countless moments that were fun and unforgettable with them. Allyson, who was around since the beginning of the lab, has been such a reliable and caring friend. Sarah inspired me to be more passionate about research. Also, many fellow graduate students, both in and out of this department, have been great friends, and gave me genuine support during times of need. I thank all the collaborators I have worked with as well as staff scientists who helped me with various instruments. As I had a rather unconventional

timeline during my graduate program, special thanks go out to the Chemistry Department staff, Directors of Graduate Studies, and people over at GSAS for helping me navigate through it.

Looking back on my upbringing and academic career, there were so many people who have taught me science and related disciplines and influenced me to be a passionate scientist. I would like to acknowledge every one of them. I was blessed to get as many opportunities as I had, and I hope to make them proud.

Friends – wherever they are – helped me make the most out of this stage of life, and I sincerely thank them for their indispensable existence. Most importantly, I thank my mom, dad, and sister for their truly unconditional support; I could always count on them for anything. Without a doubt, they gave me the foundations for the person and scientist I was shaped out to be now.

## **Dedication**

To my Family. *Veritas omnia vincit*

# Chapter 1: Introduction

## 1.1. Protein post-translational modifications

Post-translational modification (PTM) is defined as the chemical change that happens to proteins after their synthesis, which is known as translation.<sup>1</sup> This encompasses bond formation to add small functional groups, such as phosphoryl, glycosyl or acetyl, or larger structures such as lipids or even proteins such as ubiquitin, as well as bond cleavage, as in proteolysis or autoprocesing.<sup>2</sup>

PTMs can happen at any time during the lifecycle of a protein, and depending on when they happen, many different aspects of protein characteristics may be altered. PTMs during earlier stages mainly alter protein folding, conformation stability and transport destination, and those during later stages influence the activity and biological function of the targets, for example by switching protein conformation. Some PTMs tag the protein for degradation, ending the life of the protein.<sup>3</sup> This functional variation introduced through PTMs is significant, as through this process, the proteome attains a much higher complexity than what is genetically encoded in the genome, allowing another layer of functional diversity.

More than 400 PTMs have been identified so far, through methods including those utilizing antibodies as in Western blots or immunoprecipitation and mass spectrometry (MS) methods, as have hundreds of thousands of modification sites. While newer proteome-wide methods have facilitated the detection of PTMs, our ability to understand their biological function lags behind.<sup>1</sup> Still, tools such as quantitative MS, peptide synthesis with modified amino acids, and PTM-specific antibodies have allowed us to study many PTMs and link them to different biological functions or significant diseases. The majority of PTMs are reversible with specialized enzymes

for both catalysis of modification and undoing of it, and they are thought to serve as an information processing route in biological systems, sometimes in combination.<sup>4</sup>

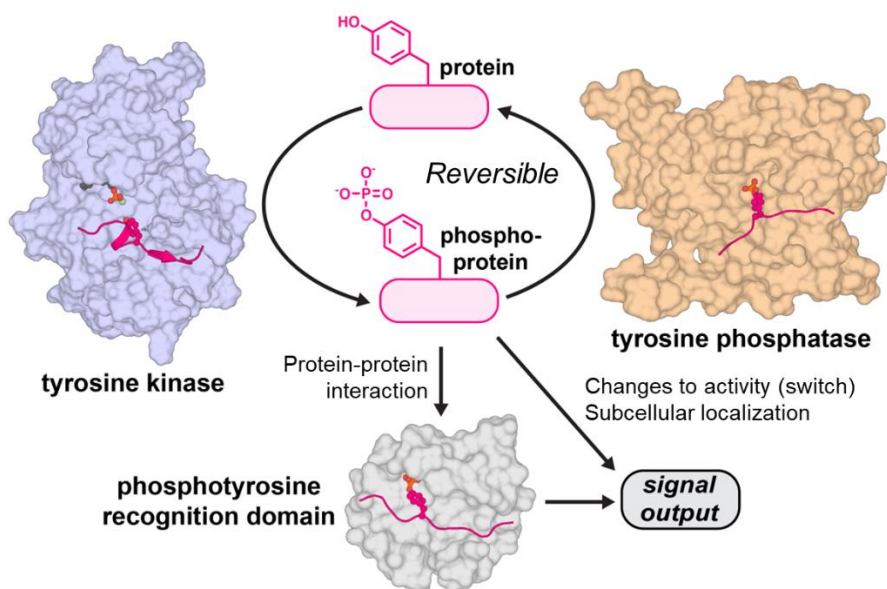
## **1.2. Significance of protein phosphorylation**

Phosphorylation has been proven to be a very significant modification in cellular processes, as it is a reversible PTM that modulates enzyme activity, interactions with other proteins, stability and localization. For example, phosphorylation of the activation loop in protein kinases typically stabilizes the active conformation and promotes their activity.<sup>5</sup> Phosphorylated residues can be recognized by specialized domains such as SH2 domains, that are embedded within other proteins, which will lead to differentiated protein-protein interactions (Scheme 1.1).

In addition to being a functionally significant PTM, phosphorylation is also the most common PTM in eukaryotes. Databases have hundreds of thousands of identified phosphorylation sites, and up to 30 % of proteins in a mammalian cell are estimated to be phosphorylated at a given time.<sup>6</sup> Therefore, the balance of protein phosphorylation, controlled by protein kinases that catalyze phosphorylation and phosphatases that catalyze dephosphorylation, is very significant in understanding signal transduction pathways in cells, and knowing states of phosphorylation in involved proteins lead to insights about what “state” a cell is in at a particular moment. Indeed, propagation of most signals that are generated from outside of a cell (e.g., hormones or growth factors) and within the cell, whether metabolic or oxidative stress, involve phosphorylation. Pathways that are regulated using phosphorylation are significant to all aspects of cell biology, including proliferation, metabolism, progression of cell cycles, transcription and translation. As disruptions to these crucial pathways can cause changes to cell states and lifecycle, many diseases

of importance, such as cancer and diabetes, have been identified to be accompanied by aberrant protein phosphorylation.

Notably, phosphorylation of other molecules in an organism, such as lipid<sup>7</sup>, oligonucleotide<sup>8</sup> and sugar<sup>9</sup>, also plays important roles in a cell. Take note of the fact that adenosine triphosphate (ATP), the energy “currency” molecule in cells<sup>10</sup>, provides free energy through its hydrolysis, which may produce inorganic phosphate, or result in phosphorylation of a biomolecule.



**Scheme 1.1: Effects of tyrosine phosphorylation.** Tyrosine kinases and phosphatases catalyze the addition and removal of phosphoryl groups on tyrosine residues, respectively. They result in profound changes in the properties of the substrate protein.

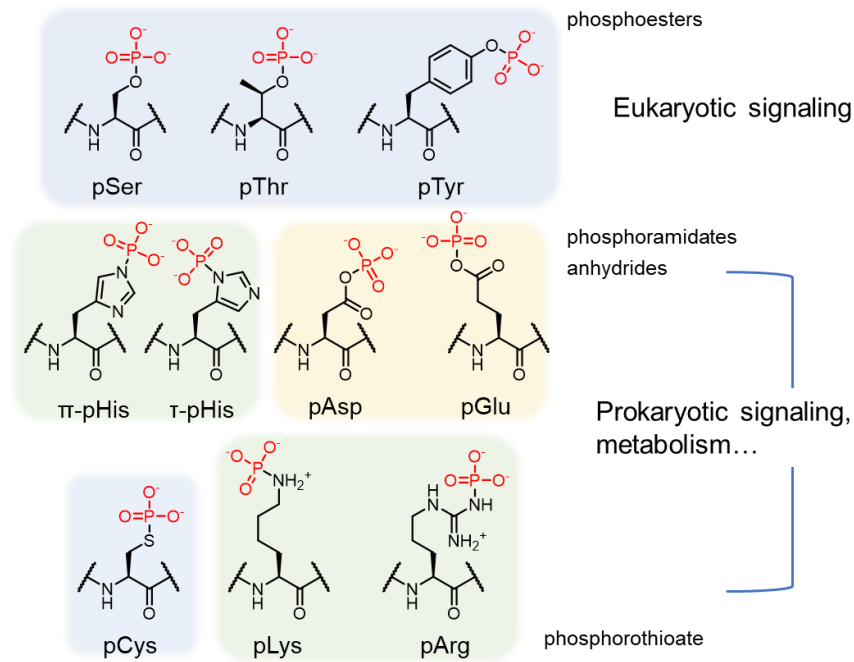
### 1.3. Protein tyrosine phosphatases

#### 1.3.1 Function and classification of protein phosphatases

Kinases, as stated above, catalyze the transfer of a phosphoryl group from a high-energy donor molecule, usually ATP, to an acceptor substrate; they are distinct from a similar family of



enzymes called phosphorylases, which catalyze the addition of an inorganic phosphate group to a substrate. Phosphatases carry out the reverse process of catalyzing the cleavage of phosphoric acid ester into phosphate and alcohol. Naturally, protein kinases and protein phosphatases catalyze the phosphorylation and dephosphorylation of protein substrates, and they are further classified by which amino acid residues they recognize as substrates. While there are 9 amino acids identified to be substrates of protein phosphorylation (Figure 1.1), three are notable for eukaryotes: serine, threonine, and tyrosine, with the first two being the most common.



**Figure 1.1: Structures of phosphorylated amino acids.** More have been reported, such as pyrophosphoserine (ppSer).

Traditionally, kinases have been studied more extensively than phosphatases. For example, many pathways have first been understood through the actions of the kinases involved - a good example is mitogen-activated protein kinases (MAPKs) and the MAPK/ERK pathway, which has

been extensively researched, due to its far-reaching impacts on cell proliferation, survival and differentiation, and consequent prominence in many diseases including cancer. This pathway involves multiple checkpoints, which are different kinases serving as regulation sites. Another example is receptor tyrosine kinases, which serve as starting points for many pathways originating from an external signal. On the other hand, originally protein phosphatases were seen as “housekeeping” enzymes that tend to dephosphorylate phosphorylated proteins, rather indiscriminately<sup>11</sup>. However, phosphatases have gradually emerged in the recent decades as regulators of many signaling pathways, and thus as potential drug targets as well<sup>12,13</sup>.

Protein phosphatases are classified according to the residues they take as substrates. Protein serine/threonine phosphatases dephosphorylate serine and threonine residues of a protein, and protein tyrosine phosphatases dephosphorylate tyrosines. Some of them recognize both serine/threonine and tyrosine residues as substrates.

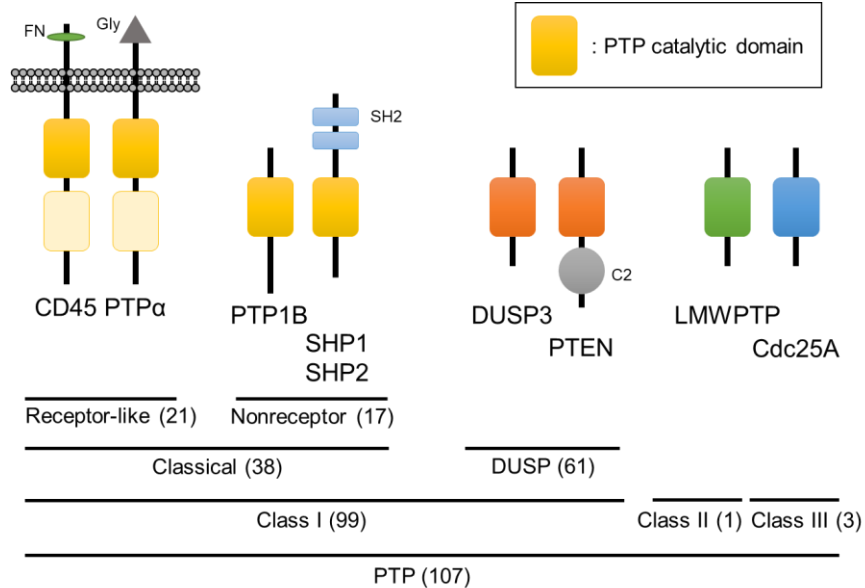
Protein tyrosine phosphatases (PTPs), along with protein tyrosine kinases, are involved in the regulation of many pathways, such as the aforementioned MAPK/ERK pathway. One interesting characteristic of PTPs is that they share a common motif in the active site – (H/V)CX<sub>5</sub>R(S/T) – whereas they are different from serine/threonine phosphatases in their structure and mechanism: serine/threonine phosphatases are metalloenzymes that have a bimetallic center, relying on metal ions such as Fe<sup>3+</sup> or Mn<sup>2+</sup> for their catalytic activity. This is in contrast to the protein tyrosine kinases, which are largely homologous in structure and sequence to the protein serine/threonine kinases.

More than 98% of all phosphorylation sites are Ser and Thr, and accordingly, serine/threonine kinases are more numerous than tyrosine kinases. However, there are over 100 PTPs in humans, compared to only about 40 protein serine/threonine phosphatases<sup>14</sup>. This suggests

that individual PTPs have different functions, and there perhaps is substrate specificity for each PTP. Indeed, even though they share a structural motif and the same catalytic mechanisms, they differ in their biological roles.<sup>15</sup> Interestingly, even with highly conserved active sites, individual PTPs display a wide range of catalytic efficiencies; for example, PTP1B has about tenfold higher efficiency than that of HePTP.<sup>15</sup> Establishing the roles of each PTP has been a long-standing goal in this topic.<sup>16</sup> Genomic approaches such as ribonucleic acid (RNA) interference and knockout analysis have been used to fill this gap, showing that phosphatases are important to cell survival or serve as tumor suppressors.<sup>17</sup> But there are limitations to those methods, including the fact that phosphatases are often regulated post-translationally, which those methods cannot report on.<sup>18</sup> Research in this field would benefit significantly from a family-wide tool to profile the functional states of PTPs; such a tool could allow researchers to assess the roles of individual PTPs in particular signaling events or pathways, and potentially help identify therapeutic targets for disease treatments.

### **1.3.2 Structure and classification of protein tyrosine phosphatases**

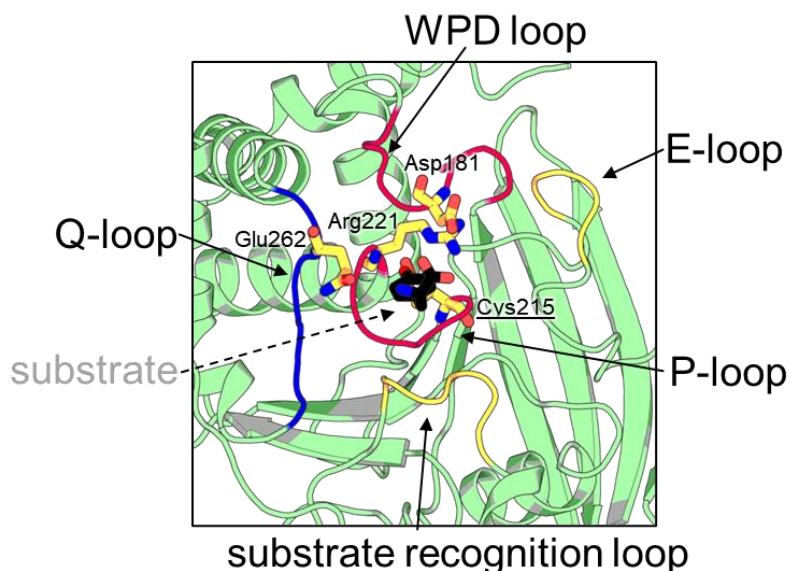
The PTP superfamily, with 107 genes in the human genome, can be further separated into several categories (Figure 1.2). There are 99 class I PTPs, 1 class II PTP, 3 class III PTPs, and 4 class IV PTPs. Class I PTPs form the major group, that contains 38 classical PTPs, which are specific to phosphotyrosine residues and consist of 21 transmembrane, receptor-like PTPs and 17 cytosolic, non-receptor PTPs. They also contain 61 dual specificity phosphatases (DUSPs, or DSPs), also known as VH1-like. Class II PTP is a single gene encoding LMWPTP, and class III PTPs comprise Cdc25A, Cdc25B and Cdc25C. Class IV PTPs are unique in that their catalytic residue is an aspartic acid.<sup>19</sup>



**Figure 1.2: Classification of human PTPs and domain architectures of structures.**

As introduced, PTPs possess a common motif also known as the PTP signature motif in its active site. The structural similarity of class I PTPs does not end at just the signature motif. The tertiary structures share many features such as the active site cleft including the P-loop, and a functionally important surface loop called the WPD loop, as well as other loops called Q-loop and E-loop (Figure 1.3).<sup>11</sup> In addition, classical PTPs have a substrate recognition loop that DUSPs lack, and this may explain the difference in their compatibility with substrates. Moreover, PTPs are notable for their modular structures (Figure 1.2). At least 79 PTPs contain one or more additional domains: receptor-like PTPs have transmembrane regions, and PTPs may have another pseudophosphatase domain or other extensions that modulate their function or targeting abilities. For example, SHP2 has two SH2 domains that regulate the catalytic activity by assuming different conformations in response to phosphoprotein binding.

Substrate specificity is conferred by both the auxiliary domains that may provide protein-protein interaction or subcellular localization,<sup>20</sup> as well as individual residues in the catalytic domain. Several individual residues have been linked to substrate specificity in PTPs, mostly through crystal structures;<sup>21–23</sup> the dynamics and plasticity of various loop structures were also shown to be important in substrate recognition.<sup>24,25</sup>



**Figure 1.3: Structural motifs of a classical PTP bound to a ligand.** Conserved residues are shown in sticks, and numbered as in PTP1B. (PDB: 1PTY)

The reason why tyrosine phosphatases have been regarded as “undruggable” by researchers at first was that the active site pocket preferred negatively charged groups in its active site, which led to poor bioavailability of drug candidate molecules, and that the conserved motif is so well conserved, making it hard to find a specific structure for a single target.<sup>13,26</sup> This is in stark contrast to protein tyrosine kinases, for which several different targeting strategies exist, due to its structural

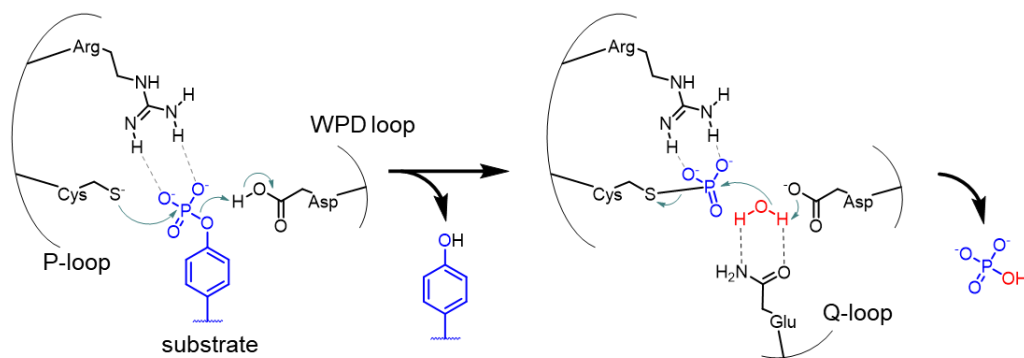
features such as the ATP binding pocket; for this reason, there are more than 70 FDA-approved drugs targeting tyrosine kinases, while there still is none for tyrosine phosphatases.<sup>27,28</sup>

While tyrosine kinases have been investigated more extensively than tyrosine phosphatases for various reasons such as PTPs being purified and cloned later than kinases<sup>19</sup>, some PTPs that have been better characterized have been shown to have important physiological implications. For example, PTPN11, encoding SHP2, is a proto-oncogene that was found to be mutated in cancer patients; germ-line mutations in this gene cause ~50 % of Noonan syndrome and almost all cases of LEOPARD syndrome, both of which are genetic disorders.<sup>29</sup> PTP1B is well known for its involvement in the insulin signaling pathway, and as such has come under spotlight as a potential therapeutic target for diabetes.<sup>30</sup> CD45, a receptor-type PTP, is involved in T cell signaling, dysregulation of which may lead to autoimmune diseases.<sup>31</sup> Several other PTPs, such as PRL3 and Cdc25s, have been identified for the correlation between their expression levels and cancers.<sup>32,33</sup>

### **1.3.3 Catalytic mechanism of protein tyrosine phosphatases**

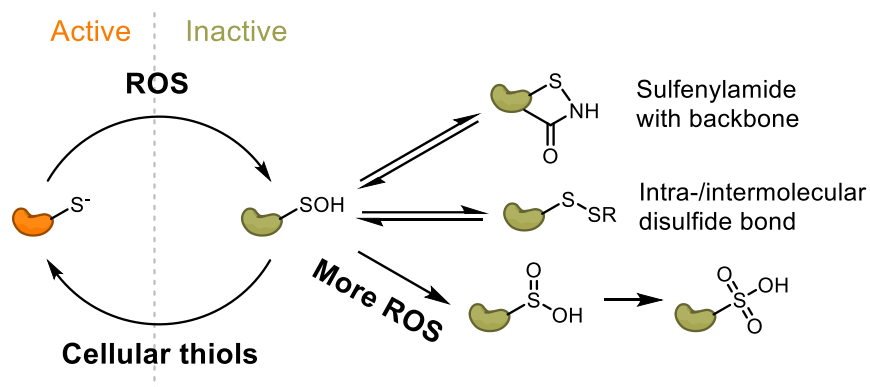
As all PTPs share catalytically important residues, their catalysis mechanism is very much conserved across members of the family (Figure 1.4), and it has been extensively studied using structural information, site-directed mutagenesis, and various kinetic analyses.<sup>34-36</sup> In the first step after substrate binding, the thiolate on the active site Cys of the enzyme forms an S-P bond by nucleophilic attack on the phosphorus of the pTyr. During this step, the transition state is stabilized by the conserved Asp residue on the WPD loop serving as a general acid, spreading the negative charge over its carboxylic acid group. After this step, the enzyme has the phosphoryl group as an E-P complex, and this is hydrolyzed, facilitated by the aspartate group; a water molecule sees nucleophilic attack by the carboxylate, and in turn, the oxygen attacks phosphorus and the thiolate

group is set free. Research shows that aside from the two crucial residues, Arg in the active site stabilizes the transition states with its guanidinium group, and Gln on the Q-loop aids the positioning of a water molecule and the Asp on the WPD loop.



**Figure 1.4: Catalytic mechanism of protein tyrosine phosphatase.**

As the active site Cys is central to the catalytic mechanism, any modification of it disables the activity of the enzyme. This opens the door for various regulatory mechanisms for PTPs. A well-known mechanism is through oxidation: oxidation of thiol groups produces sulfenic acid (Cys-SOH), which is reversible, and further oxidation takes the functional group into irreversible states of oxidation, namely sulfinic acid (Cys-SO<sub>2</sub>H) and sulfonic acid (Cys-SO<sub>3</sub>H) (Figure 1.5).



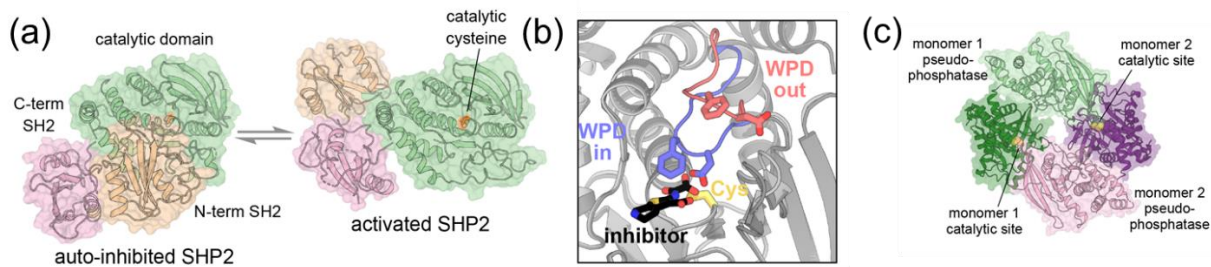
**Figure 1.5: Redox regulation of catalytic Cys residue.**

With methods to monitor PTP oxidation, it seems that reactive oxygen species (ROS) are functioning as intracellular secondary messengers in many signaling pathways.<sup>37</sup> What is less clear is whether this is actively used in the cell as a way of controlling the activities of PTPs, or if thiol oxidation is yet another side effect of ROS generation in cells. Interestingly, there is a chemical reaction that prevents active site Cys from getting irreversibly oxidized. Sulfenic acid has been shown to react with an adjacent backbone amide or a thiol group to form a cyclic sulfenic amide or disulfide bond, respectively, states that are reversible by reduction and cannot be oxidized further. This may suggest that oxidation is being put to a regulatory use in cells, or at least presents therapeutic targets.<sup>38</sup> Furthermore, some sulfenic amide formation has been shown to induce a substantial conformational change in PTP1B, changing its substrate binding ability.<sup>39</sup> It has also been suggested that *S*-nitrosylation, another known form of PTM for cysteines, potentially protects the active site cysteines from irreversible oxidation.<sup>40</sup>



### 1.3.4 Regulation of protein tyrosine phosphatases

Besides oxidation, several regulatory mechanisms are observed in PTPs.<sup>38</sup> One category of regulation mechanisms where substrate binding capabilities are altered to control the activity. As mentioned earlier, SHP2 possesses a mechanism where it switches conformation between an active one where access to the catalytic domain is not blocked and an inactive one where the active site is occluded by its N-SH2 domain, preventing catalysis (Figure 1.6a). This has made possible the development of allosteric inhibitors of SHP2, such as SHP099, that lock SHP2 in an inactive state, without perturbing the other PTPs. Indeed, SHP099 was demonstrated to suppress Ras-ERK1/2 signaling, showing that PTPs can be used as targets to control downstream signaling.<sup>41</sup> This allostery is recognized as a promising strategy for targeting phosphatases in therapeutics, given the hard-to-drug nature of the PTP active site; 6 allosteric inhibitors of SHP2 are currently in clinical trials, with more as combination drugs.<sup>42,43</sup>



**Figure 1.6: Regulation mechanisms of PTPs.** (a) Allosteric regulation of SHP2. (b) WPD loop of PTP1B, shown in both “in” and “out” states. (c) Dimerization of receptor PTP  $\gamma$ . (PDB: 2SHP, 6CRF, 5K9V, 5K9W and 2NLK)

There are several additional examples of allosteric regulation in PTPs. The WPD loop of PTP1B is known to be in two different conformations of “open” and “closed”, typically staying open in apo states and closed in bound states (Figure 1.6b).<sup>44</sup> This loop positioning is also known

to change depending on oxidation states.<sup>45</sup> Importantly, several allosteric inhibitors that fix PTP1B in the open and thus inactive state have been identified.<sup>46</sup> Research suggests that MKP3 (DUSP6) undergoes a conformation change upon ERK2 binding, which increases the catalytic efficiency by realigning the active site and the acid loop.<sup>47</sup> Several PTPs oligomerize which affects the activity too (Figure 1.6c): a receptor-like PTP $\alpha$  was found to dimerize in a ligand-induced way, after which the active sites become occluded and thus activity is lost;<sup>48</sup> PTEN, a well-known tumor suppressor due to its specificity for phosphatidylinositol-3,4,5-trisphosphate involved in PI3K/Akt pathway, was shown to homodimerize.<sup>49</sup> Some cancer-linked mutations to PTEN were shown to constrain the catalytic activity of wild-type PTEN as well in a dominant-negative matter, due to the dimerization between those two types of PTEN.<sup>50</sup>

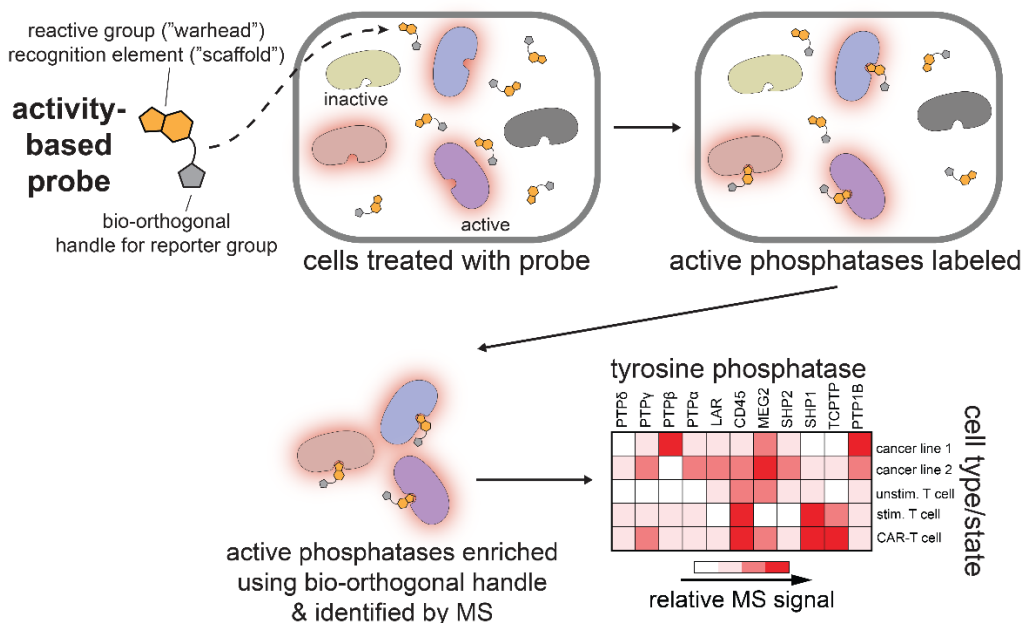
Other ways of regulation exist, including controlled expression, subcellular localization, and alternative splicing; PTPs may also be phosphorylated themselves, the effects of which may not be fully known yet.<sup>51,52</sup> While regulation methods for a small number of PTPs are relatively well studied, much is unknown about the vast majority of the PTPs.

## **1.4. Activity-based profiling**

### **1.4.1 Activity-based protein profiling and activity-based probes**

Activity-based protein profiling (ABPP) is a method to profile enzyme activity, using a small molecule activity-based probe (ABP) that labels the target enzyme. The important components of the probes are the reactive group, which is responsible for bond formation between the probe and the target, and the detectable group, which allows for recovery or identification of the labeled proteins. In addition, the probe has to have a structure that allows the probe to have an

affinity for the target enzymes and prefers to bind the active enzyme over inactive or modified ones (Scheme 1.2).<sup>53</sup>



**Scheme 1.2: A visual explanation of activity-based protein profiling.**

The reactive group, often called the “warhead,” is what allows the probe to covalently label the target, through a mechanism that is related to the catalytic activity of the target. Often this is an electrophilic group that results in a nucleophilic attack by the nucleophile in the active site. The reactive group needs to have the right reactivity, as it must be reactive enough to label the target site, but not so reactive that it also labels the off-target residues. The detectable group is usually either a fluorescent tag or an affinity tag. Fluorophores may be used for a gel-based ABPP or microscopy; they can be enhanced by making them activatable, which is useful for *in vivo* applications. An affinity tag, such as biotin, can be used in blotting, or for enrichment using resin-based reagents for subsequent MS analysis. Furthermore, a two-step approach is possible, where

a chemical group is in the place of an affinity tag, which is later reacted with a reporter group, through bioorthogonal click chemistry such as copper-catalyzed cycloaddition reaction.<sup>54,55</sup>

One key advantage of ABPP is that it allows researchers to monitor the activity levels of the target enzymes, instead of just abundance at the transcription or the translation level. With this technique, enzymatic activity of the target may be queried across various cell types, cell states, or after treatment with different inhibitor compounds. Since the advent of tandem mass spectrometry, the combination of ABPP with MS has given researchers a powerful tool to identify hundreds of enzymes of the same family at once.<sup>55</sup>

A mode known as competitive ABPP can also be employed on inhibitor screens, where a biological system is treated with candidate compounds and compared to the untreated system via ABPP, to look for a reduction in the signal. This can be used even for enzymes that are poorly characterized, as it does not require prior knowledge about the catalytic mechanism. It has been applied to high-throughput screening (HTS).<sup>56</sup>

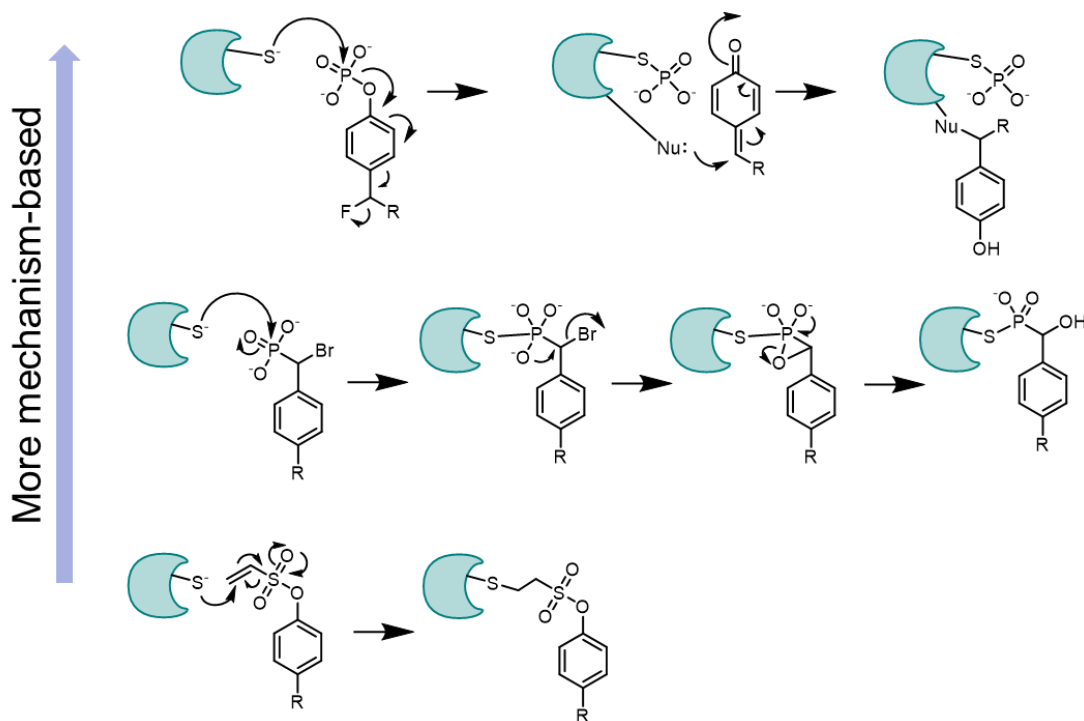
ABPP has been successfully applied to a number of enzyme families, yielding valuable discoveries. For serine hydrolases, ABPP has identified a new member of the family without sequence homology, and led to the development of multiple inhibitors of the family.<sup>57</sup> ABPP helped uncover several cysteine proteases that contribute to tumor growth, provided a diagnostic tool for lysosomal storage diseases by targeting glycosidases, and was used for inhibitor screening and *in vivo* imaging of several enzymes, including kinases and metalloproteases.<sup>58-62</sup> Most families with viable ABPs have been probed for their activity profiles in different cancer contexts, providing new understandings of cancer metabolism and signaling pathways.<sup>63</sup> The enzyme families that present the easiest target for ABP are those that researchers know sufficiently about the catalytic mechanism and structure. Especially, if an enzyme forms an enzyme-substrate

intermediate in the catalytic process, this can be exploited to create ABPs.<sup>53</sup> By this logic, tyrosine phosphatases should also be a good target for potential ABP. Indeed, several activity-based probes for PTPs have been reported so far;<sup>64</sup> however none of those molecules have been successfully applied to a mass spectrometry proteomics work.

#### **1.4.2 ABPP for phosphatases and design considerations**

We ranked previously reported ABPs by the similarity of their labeling reaction mechanism to the substrate dephosphorylation mechanism of the PTPs (Figure 1.7). A school of thought is to use a suicide substrate. Also successfully applied to targets that do not form an enzyme-substrate intermediate, suicide substrates will label the enzyme in an indirect way after catalysis.<sup>65</sup> In this case, a quinone methide-generating probe, such as 4-difluoromethylphenyl phosphate (DFPP) derivatives, has been shown to work with PTPs.<sup>66,67</sup> These will undergo dephosphorylation as a normal substrate would, but the resulting quinone methide is quickly attacked by a nearby nucleophile in the active site to generate a covalent adduct. However, these are not specific for PTPs, and also there is always a concern that the created reactive species will react at a place that is distant from the active site, or more importantly, to another protein in the vicinity and not the protein of interest.<sup>53</sup> Another broad family of probes has been  $\alpha$ -halobenzylphosphonates: closely resembling the native substrate pTyr, their reaction also begins with the active site thiolate attacking the phosphorus atom for an S-P bond formation. Phosphonodifluoromethyl phenylalanine (F<sub>2</sub>Pmp) has been reported as a non-hydrolyzable inhibitor for PTPs,<sup>26</sup> and subsequently covalent bond-forming derivatives were developed, including a tripeptide of L-phosphono-bromomethylphenylalanine (BrPmp). A bromo variety,  $\alpha$ -bromobenzylphosphonate ( $\alpha$ -BBP), with a biotin tag, has been put to use as an activity-based probe for PTPs, as well as with

rhodamine for visualization.<sup>68</sup> This was possible as they did not react with other phosphatases, such as serine/threonine phosphatases.



**Figure 1.7: Reported activity-based probes and their reaction mechanisms.** They are ranked by their resemblance of substrate dephosphorylation mechanism.

A third option is a covalent inhibitor probe that will react with the active site cysteine, but in a less mechanism-based way than those introduced above. In this area, as there are many thiol-reactive groups in the chemical space, many classes of compounds have been identified as inhibiting PTPs, either through direct bond formation, or irreversible oxidation of the active site cysteine.<sup>64</sup> However, many of these reports only go as far as demonstrating that the compounds inhibit PTPs. In many cases these molecules are very small molecules with a functional group to react with the thiols, so without the degree of similarity in their mechanisms to substrate dephosphorylation mechanism that the probes mentioned previously have, it can be expected that

they will have very high off-target reactivity as well. Indeed, it seems that the only thing that separates the active site cysteines from all other cysteines is the  $pK_a$ , which the active site cysteines have lower values.<sup>69</sup> In terms of accessibility, active site cysteines are actually at a disadvantage, as molecules that lack a scaffold to speak of, such as iodoacetamide (IAA), are known to prefer the surface-exposed residues to the ones in pockets.<sup>70</sup> Still, aryl vinyl sulfones and sulfonates have been reported as an active site-directed probe for PTPs.<sup>71</sup>

Therefore, it seems imperative that binding affinity is introduced through a scaffold structure - also known as a recognition element - towards the active site pockets of PTPs. In this aspect, PTPs having a highly conserved active site is a plus, as the goal of activity-based profiling is more of a family-wide probing of PTPs, as opposed to drug discovery efforts.

## **1.5. Overview of dissertation**

In this dissertation I present a series of works to develop active site-directed probes for tyrosine phosphatases that may be used as activity-based probes. It starts with previously reported vinylsulfonate and vinyl sulfone molecules, which were characterized extensively; on top of that, new variants were synthesized. The result was that these molecules were too reactive to be successful ABPs. Next, I turn to a wide range of thiol-reactive groups to evaluate their reactivities, and then sample a few of them for a systematic analysis of different scaffold structure fragments. PTP-specific assays were used to assess the chemical space for potency and target specificity, and PTP-independent assays aided the contextualization and interpretation of the data. This process identified trends in individual PTPs, that thiazoles are preferred by many PTPs, and even uncovered some structure-activity relationships. Encouraged by this, I employed an in-depth analysis of kinetic parameters, and showed that affinity and reactivity aspects of each probe

reaction can be partly extracted from PTP inhibition data. Finally, several candidate molecules were tagged and taken to both in-gel and MS-based proteomics experiments, where successful enrichment of PTPs were shown.



## Chapter 2: Evolution of vinylsulfonate probes for PTPs

### 2.1. Background

While many PTPs play an important role in cell signaling, and their dysregulation has been shown to lead to diseases of clinical significance, a majority of them have not been scrutinized for their substrate specificities or regulation. A method that will allow a more detailed look into how each PTP behaves during different cell states is activity-based protein profiling (ABPP). Many families of enzymes, including kinases and proteases, have been investigated for their activity or regulation across cell states or signaling events through this method, especially when employed together with tandem mass spectrometry (MS) proteomics. As explained in the previous chapter, several types of probes have been reported for PTPs. Quinone methide-generating probes, which are dephosphorylated the same way as natural substrates, have seen some success with the attachment of peptide recognition elements, but have an inherent issue of generated reactive species not necessarily reacting with the target protein;  $\alpha$ -bromobenzylphosphonate probes were used both in conjunction with biotin and fluorophores, but suffer from low cell permeability and susceptibility to solvolysis.<sup>18</sup> Aryl vinyl sulfone and sulfonate probes, which do not have ionic charged groups and are cell-permeable, were shown to be active site-directed probes that present a potential for building upon with rational design, which is the basis of this chapter.<sup>18</sup> However, there has been no report on ABP with MS proteomics with PTPs to date. Indeed, the most challenging aspect of finding the right probe is striking the balance between the reactivity, mostly rooted in the reactive group – or “warhead” – choice, and the selectivity towards target proteins over the background, which tends to depend on the overall structure of the compound. Then, having cell permeability is necessary to probe a live biological system, such as cell culture.

Phenyl vinylsulfonate (PVSN), along with phenyl vinyl sulfone (PVS), was reported in 2008 as an active site-directed and mechanism-based probe for PTPs.<sup>71</sup> In their work, the formation of a covalent adduct with YopH was shown through both intact protein MS, peptide MS, and peptide MS/MS; and its kinetics were observed through inactivation assays, for example,  $K_I$  and  $k_{inact}$  of PVSN for YopH were measured to be  $0.29 \pm 0.06$  mM and  $0.061 \pm 0.002$  min<sup>-1</sup>, respectively. Crystal structures for PVSN and PVS adduct of YopH were also obtained, providing insight into the probe's orientation in the active site pocket, and some *in vitro* experiments were done, where PVSN was shown to label His-tagged HePTP transformed into the proteome and also increase global tyrosine phosphorylation levels in COS-7 cells treated with the compound - suggesting cell permeability.

One notable condition of the experiments was that they were conducted at pH 6. This provides for the differentiation of the active site thiol, as it will mostly stay in a thiolate state due to its lower  $pK_a$ <sup>69</sup> originating from its local environment.<sup>72</sup> Indeed, some of the results implicitly show that the covalent labeling by PVSN mostly happens on the active site cysteine; however, even at pH 6, none of the results in the original study convincingly show that PVSN labeling is selective towards the active site Cys.

As explained in the previous section, this may be because at this stage researchers were focused on identifying what inhibits PTPs and how, without too much concern going into the off-target reactivity. The reader will find that some of the conditions used in the original work, such as the millimolar range concentration of PVSN, are actually on the extreme side for a highly reactive group as vinylsulfonate. Nonetheless, this work proposes an interesting point, that PVSN is a pTyr mimetic, supported by the observations that: crystal structure shows the phenyl group situated in the “pTyr binding pocket,” methyl vinylsulfonate did not show an appreciable inhibition

of YopH even after prolonged incubation, and mutation of the general acid Asp results in a slower inhibition of YopH.

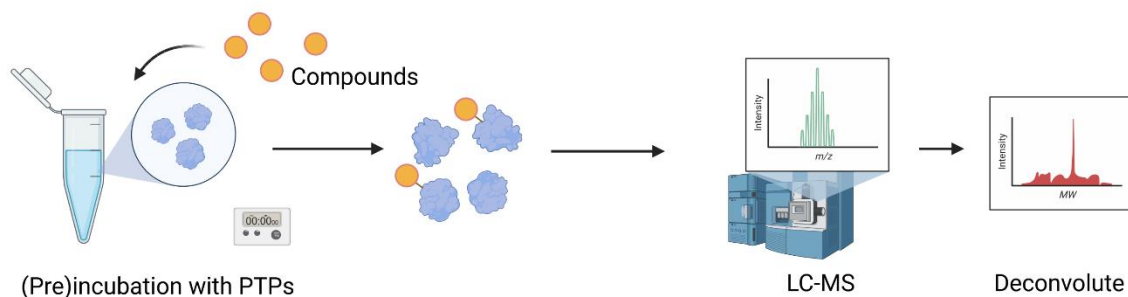
So even though PVSN may be on the lower end of how mechanism-based different activity-based probes are, it is possible that the inhibitor benefits from the active site structure designed to catalyze the dephosphorylation of pTyr, such as the general acid Asp, or the stabilization of sulfonate group oxygen atoms from the P-loop side chains and backbone amides. This leads us to a seminal concept with which I embarked on this journey: if PVSN is a mimetic of pTyr substrates, cues may be taken from other known substrates as well.

In this chapter, we showed that PVSN is highly reactive with multiple Cys sites on PTPs at physiological pH. Then, a new coumarin-based compound was conceived, and we show that it has an even higher reactivity. As in the original paper, azide-tagged PVSN was synthesized and shown to label proteins covalently; lastly, more of the vinylsulfonate derivatives were synthesized and characterized similarly.

## **2.2. PVSN is highly reactive towards non-catalytic cysteines**

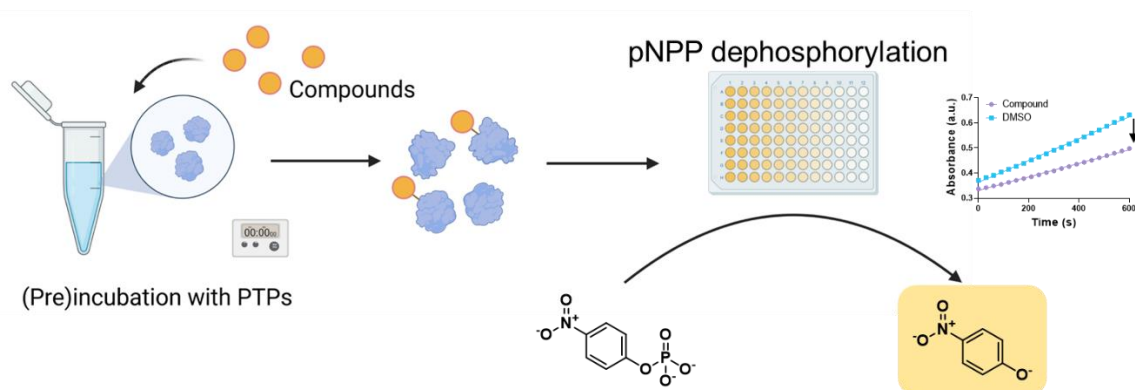
To monitor the reaction of PVSN with proteins, a liquid chromatography-mass spectrometry (LC-MS) was developed. Here, purified PTPs were incubated with probe compounds – in this case PVSN – and the resulting mixture was analyzed on MS. While the mixture did not contain any significant contaminant or side product, the LC portion helped to get cleaner protein signals. After integrating the  $m/z$  spectra accumulated over the protein elution window, deconvolution of the charge envelope with MaxEnt1 algorithm revealed molecular weights of different species, corresponding to proteins covalently labeled with various numbers of candidate

molecules. Then the composition of total proteins in different degrees of labeling could be obtained. (Scheme 2.1, see also Figure 2.1)



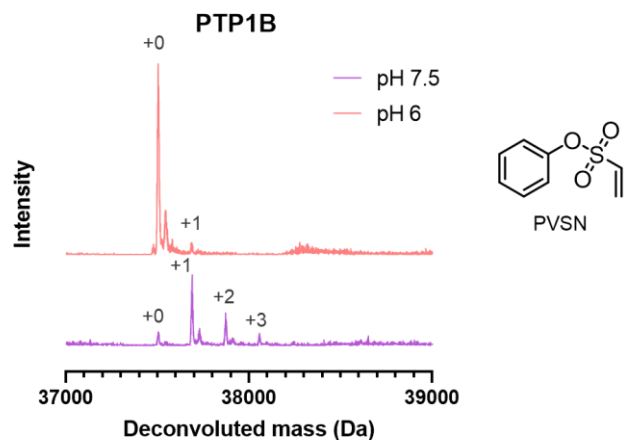
**Scheme 2.1: Workflow of intact protein MS experiments.**

After testing PVSN with several PTPs, it was immediately clear that, given the existence of an E+2I (protein covalently linked to two probe molecules) species or higher, at least some of the off-target residues are getting labeled. The finding was further supported by inhibition assay: after the preincubation of compounds with purified PTPs, the remaining activity of the enzyme was measured by monitoring the catalysis of pNPP (4-nitrophenyl phosphate) dephosphorylation, which results in pNP (4-nitrophenol), the deprotonated form of which has a strong absorbance around 405 nm (Scheme 2.2). Inhibition data can be used as a proxy for labeling in the active site, the absence of inhibition gives away that the active site cysteine is not labeled at all, and so some other residues must be labeled, to give the adduct species in the MS experiments.



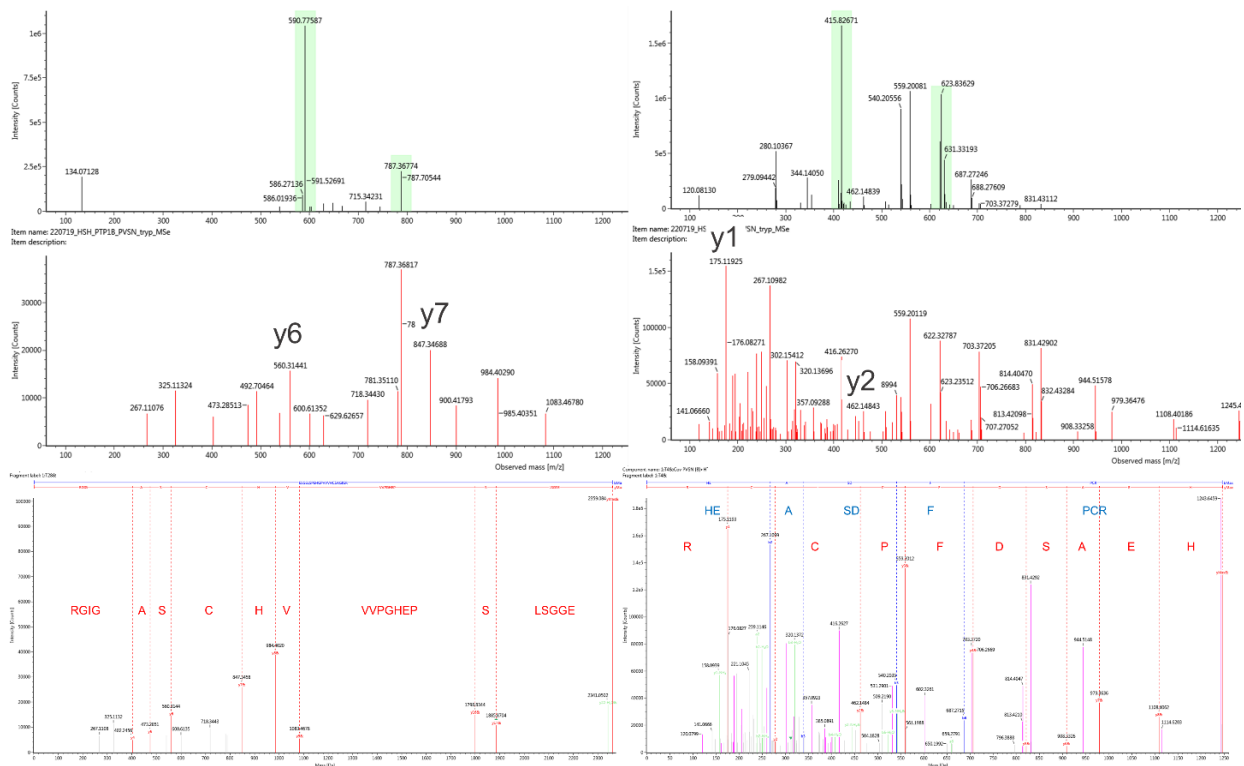
**Scheme 2.2: Workflow of PTP inhibition assay with pNPP.**

The result showed that for certain PTPs such as CD45, the same condition does not inhibit the enzyme at all (see Figure 2.4). We also tested PVS, and it had a lower reactivity than PVSN. This meant a few things: first, while PVSN labeling may be somewhat mechanism-based, it does not necessitate the existence of catalytic structures in the active site pocket. In fact, simple Michael addition reactions between PVSN and any surface thiol residue seem to be possible, and in some cases favored. Second, this makes it very unlikely for PVSN to be successfully used as an activity-based probe because if the molecule can label some off-target residues in the same protein, it will likely react with other off-target residues across the proteome as well. It is important to note that all these experiments were carried out at pH 7.5, the physiological pH that the ABP will have to function in if it is to be used in *in vitro* cell-treated experiments. At a higher pH, all thiols are more in the deprotonated state, which is the reactive species; therefore, the reactivity needs to be tuned at the same time as the scaffold is improved to provide affinity for the active sites of PTPs. When the reaction of PVSN with PTPs was observed in pH 6, it was confirmed that the reactivity is greatly reduced, with the majority of PTP species still staying unreacted. (Figure 2.1)



**Figure 2.1: Reaction of PTP1B and PVSN as determined by intact protein MS.** PTP1B was incubated with 100  $\mu$ M PVSN (structure on the right) at pH 6 and pH 7.5 for 1 h before being analyzed on LC-MS. Numbers on each peak denote the corresponding number of PVSN per enzyme reacted.

We subsequently identified the labeling sites on PTP1B and YopH, by tryptic digest MS. PTP samples were preincubated with compounds as with intact protein MS experiments, then digested with trypsin. The resulting peptide samples were analyzed by tandem mass spectrometry, specifically with data-independent acquisition (DIA) over the full  $m/z$  range, also known as MS<sup>E</sup> for Waters instruments. This showed that noncatalytic Cys residues are also labeled, and for some residues at a higher rate than the catalytic Cys, confirming the observations made in the intact protein MS assay and the inhibition assay. (Figure 2.2)

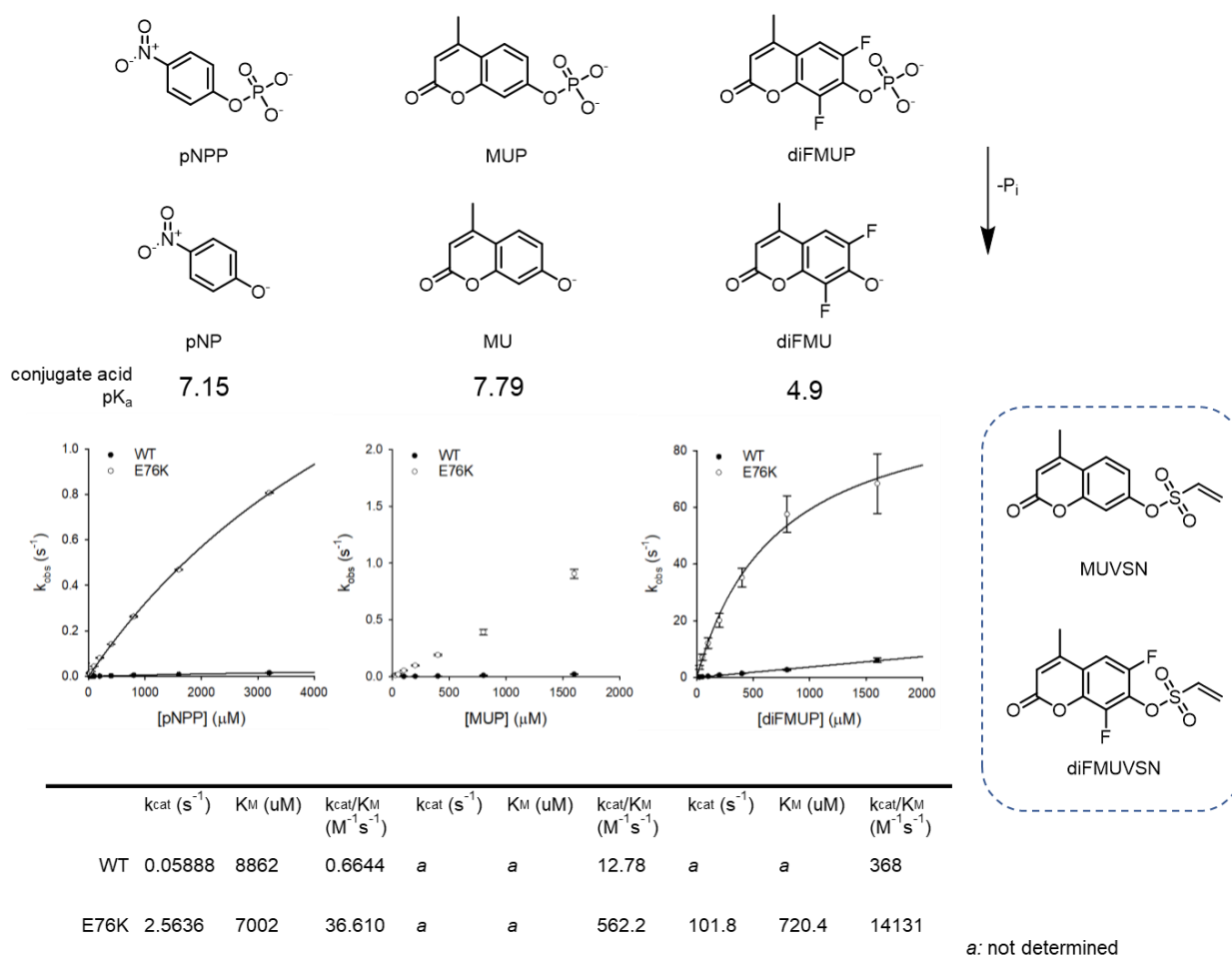


**Figure 2.2: Example spectra for PVSN labeling site determination.** Both peptides containing the active site Cys (left) and those containing other Cys (right) show labeling by PVSN. On top are low-energy spectra where green highlights the peptide  $m/z$ , then are high-energy spectra where fragment ions surrounding Cys-PVSN are marked, and at the bottom are the fragment ion assignments.

### 2.3. Methylcoumaryl vinylsulfonate, a substrate-based design with higher reactivity

While pNPP is a common substrate for kinetic analysis of any phosphatases, because the product generated from its phosphorylation is measured by its absorbance, there are limits to how low the substrate concentration can go. There are several fluorogenic substrates used to get around this with much more sensitive fluorescence readings: 6,8-difluoro-4-methylumbelliferyl phosphate (diFMUP) is a commonly used example of it. Michaelis-Menten analyses of pNPP and diFMUP using full-length SHP2 gain-of-function mutant E76K, which stays in the open conformation, showed that  $k_{cat}$  was about 40 times higher for diFMUP than for pNPP, while  $K_M$  was lower, about

tenfold (Figure 2.3). This is confirmed by previous reports of the same analysis.<sup>73,74</sup> That the  $k_{cat}$  of diFMUP substrate would be higher than that of pNPP is expected, due to the stabilization of negative charge state of diFMU – also evidenced in the lower  $pK_a$  of methylumbelliferone (7.79) than that of phenol (10.0), which would be only further enhanced by the two fluorine substitutions ( $pK_a$  of the leaving group is 4.9). However, the scaffold also provides for better affinity, resulting in a lower  $K_M$  value. Catalytic efficiency  $k_{cat}/K_M$  for MUP is between those of pNPP and diFMUP for both wild-type and E76K, likely a result of both factors.



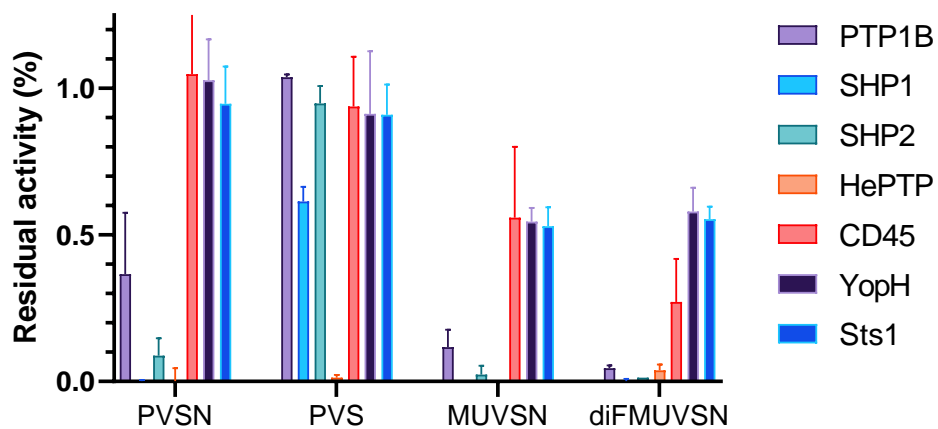
**Figure 2.3 Commonly used substrates and their kinetic properties.** Three substrates and their leaving groups with  $pK_a$ ,  $K_M$  and  $k_{cat}$  for SHP2 FL wild-type and E76K mutant as obtained from



Michaelis-Menten assays, secondary plots of which are shown. Structures of vinylsulfonate analogs of MUP and diFMUP.

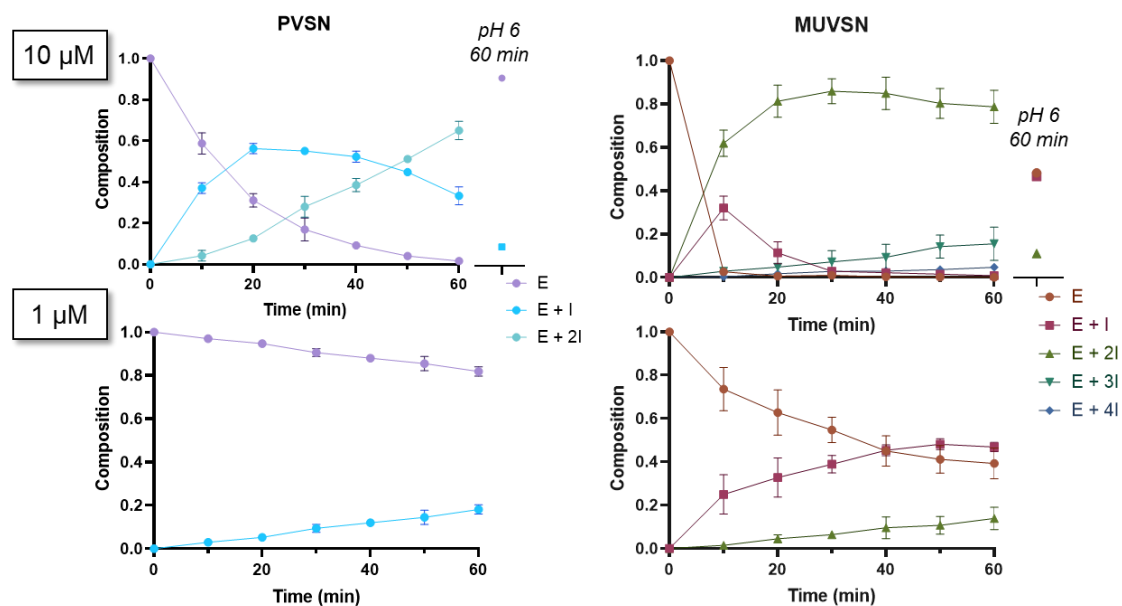
Based on the lower  $K_M$  for diFMUP over pNPP, we reasoned that covalent probes based on an umbelliferone scaffold might be better than those based on a phenyl scaffold (Figure 2.3). Simply, if we regard PVSN as an analog of pNPP minus the nitro group, we can imagine an umbelliferone molecule with a vinylsulfonate attached to the 4-position. We were able to synthesize both molecules with and without the difluoro substitutions, producing 7-(4-methylcoumaryl) vinylsulfonate (abbreviated as MUVSN for convenience) and 7-(6,8-difluoro-4-methylcoumaryl) vinylsulfonate (diFMUVSN). Due to the small amount of the diFMUP analog, more assays were performed on the MUP analog molecule. The results showed that MUVSN was more reactive than PVSN both towards the active site of PTPs and generally towards the Cys residues. The most dramatic illustration of this came from the intact protein MS labeling experiment of SHP2 catalytic domain with a high concentration of vinylsulfonate compounds (1 mM): while PVSN also showed adduct formations of higher than 6 - the number of Cys residues in the catalytic domain of SHP2 - going up to 8 labeled sites at 30 minutes, MUVSN displayed the highest adduct species of 31 at a shorter time of 20 minutes (data not shown). It is known that groups containing vinyl sulfone can label other amino acid residues such as lysine and histidine. This was corroborated by both our fragment  $MS^E$  data showing PVSN label those residues, and our data in the inhibition of Sts-1/2, phosphatases with His as catalytic residue instead of Cys, with various thiol-reactive compounds. Thus, this result is expected in terms of higher adduct species being formed, but still striking in showing how different the reactivity is between the phenyl and the methycoumaryl scaffold of the same vinylsulfonate group of compounds. While diFMUVSN reaction with PTPs was harder to characterize on MS due to the fragmentation of the covalent

adduct, it showed impressive reactivities for the active site Cys, inhibiting CD45 and YopH, the ones harder to inhibit, up to ~40 % and ~90 %, ranking among the most reactive covalent probes among the ones tested in latter chapters. (Figure 2.4)



**Figure 2.4: Inhibitory activities of PVSN, PVS, and substrate-inspired compounds.** All preincubation was done with 100  $\mu$ M compounds for 1 h. Sts-1 is a ubiquitin-associated protein that belongs in the histidine phosphatase superfamily, meaning that it displays phosphatase activity with a catalytic His residue.

By going to much lower concentrations, as low as 1  $\mu$ M, we were able to compare the reaction progress of PVSN and MUVSN by injecting the incubation mixture onto the MS at different timepoints. This showed how, as the reaction progressed, species of the higher labeling degree emerged and lower ones disappeared accordingly. At pH 6, the progress was far slower for both molecules (Figure 2.5).

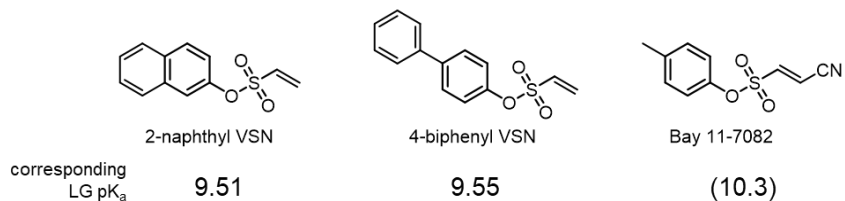


**Figure 2.5: PVSN and MUVSN labeling over time as monitored by intact protein MS.** The reaction was carried out at pH 7.5; for comparison, datapoint for 60 min at pH 6 (10  $\mu$ M) is also shown to the right.

## 2.4. More vinylsulfonate compounds

Data such as intact protein MS that showed higher off-target reactivity for MUVSN made suggested that going from PVSN to MUVSN had not only scaffold effects, increasing the binding affinity which was the intention, but also drove up the intrinsic reactivity of the vinylsulfonate group. We sought to try other scaffold structures with a similar bulk but without the heteroatoms of the umbelliferone structure. For a quick reference, it has been noted through studies on phosphate substrates, that the  $pK_a$  of 2-naphthol (9.51) is similar to that of phenol, and is higher than that of 4-nitrophenol (7.15) and methylumbelliferone, which is the leaving group for pNPP, even though 2-naphthyl phosphate behaves similarly with pNPP and MUP in terms of its  $k_{cat}$  and  $K_M$  when tested with Cdc25A and PTP1, a rat homolog of PTP1B<sup>35,73</sup>. This means that 2-naphthyl phosphate is still a better substrate than phenyl phosphate despite having a similar leaving group

$pK_a$ . Thus, we hypothesized that using 2-naphthyl analog of PVSN may show the preference for double-ringed substrates over their single-ringed counterparts more clearly, with the leaving group  $pK_a$  being similar; while the group never leaves in this case, it could still speak for the stabilization of the transition state. (Scheme 2.3)

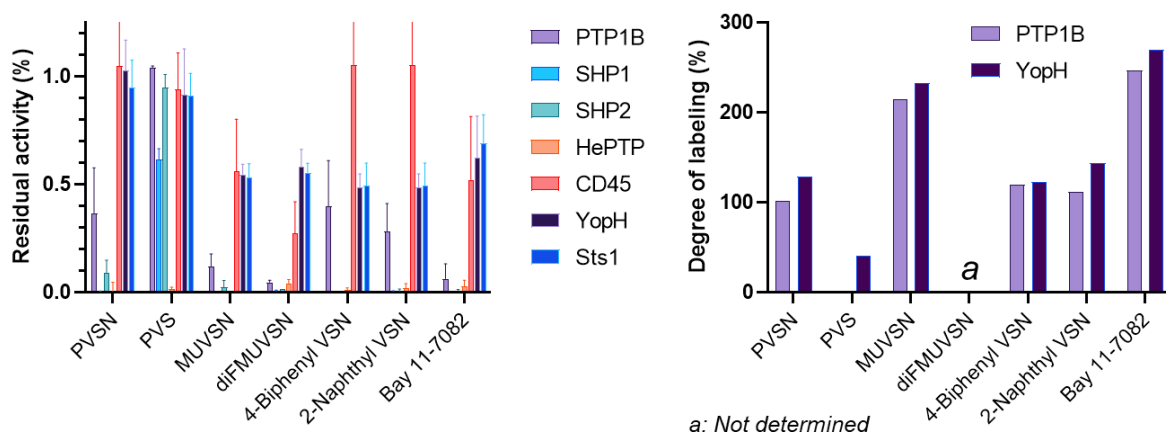


**Scheme 2.3: Structures and leaving group  $pK_a$  of more vinylsulfonate compounds.**

Accordingly, vinylsulfonates connected to 2-naphthyl and 4-biphenyl structures were synthesized, as representatives for double-ring structures that are either connected by a C-C bond or fused. The result showed that 2-naphthyl vinylsulfonate and 4-biphenyl vinylsulfonate had a comparable level of reactivity, with some differences in the individual PTP level, the naphthyl compound being slightly more reactive with PTP1B and CD45. On average, these compounds were on the same level of reactivity as MUVSN, but slightly lower. With the MS experiment, a similar trend was observed, with the two compounds coming in between PVSN and MUVSN for the average degree of labeling. (Figure 2.6)

Another vinylsulfonate compound that has been reported as a PTP inhibitor is Bay 11-7082,<sup>75</sup> which differs from PVSN mainly by the (E)-2-cyano group on the vinyl. While it is best known as an nuclear factor  $\kappa$ B inhibitor,<sup>76</sup> it was reported to inhibit many other classes of enzymes such as ubiquitin-specific proteases.<sup>77</sup> We tested Bay 11-7082 in both substrate inhibition and intact protein MS assays and found it to be extremely reactive, only matched by our own

diFMUVSN molecule (Figure 2.6). Therefore, it is no surprise that Bay 11-7082 is a broad-spectrum inhibitor for many classes of enzymes that require a Cys group in catalysis. We also found that covalent adducts of Bay 11-7082 manifest in two different added molecular weights: one corresponding to Michael addition mechanism, and another lighter modification, which corresponds to a substitution reaction with 4-methylbenzenesulfonic acid as the leaving group that has been proposed before.<sup>78</sup> The existence of a second, dissociative pathway renders this compound less useful for ABP purposes.

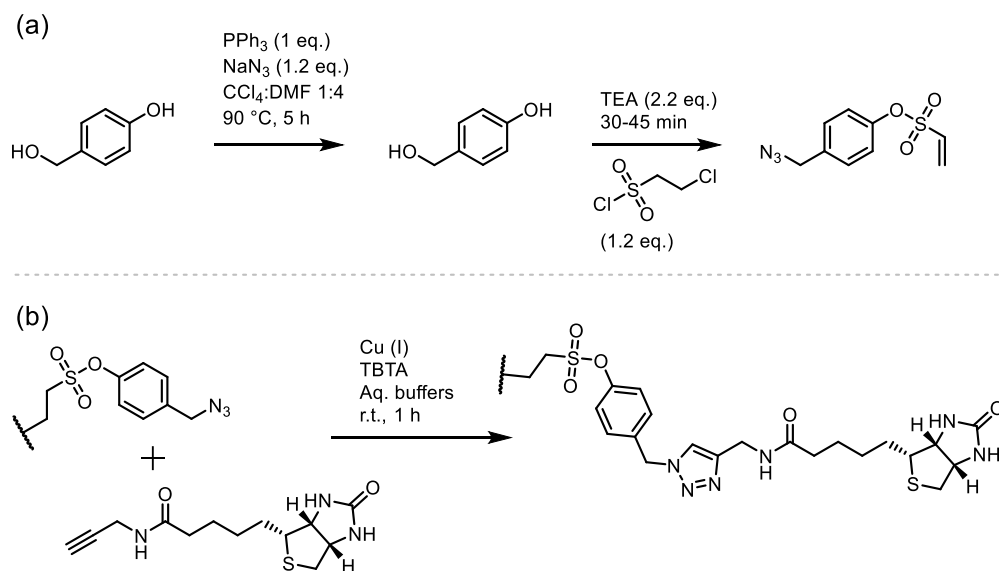


**Figure 2.6: PTP inhibition and labeling by vinylsulfonate compounds.**

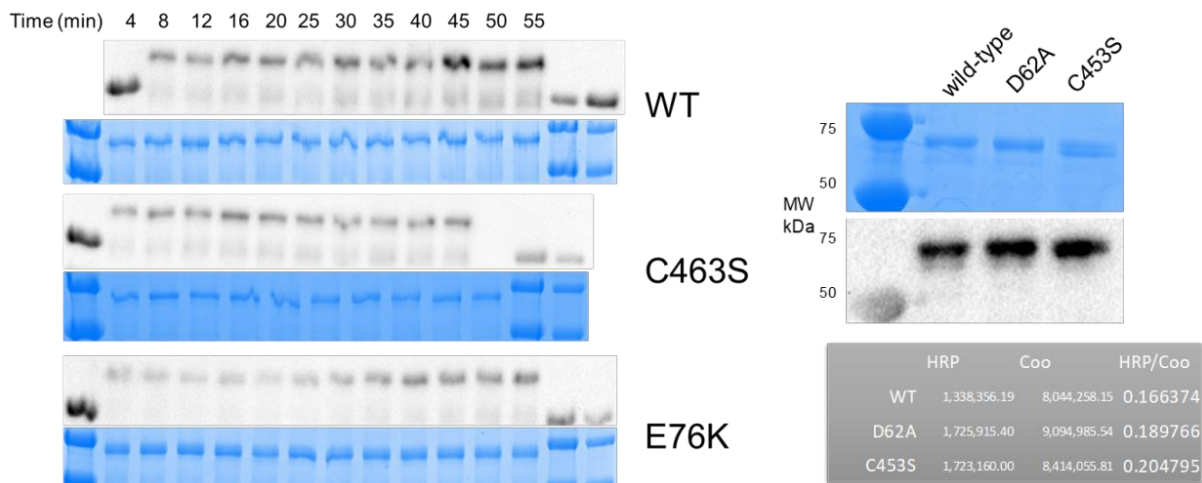
## 2.5. Tagged PVSN can label proteins, but indiscriminately

As the original reference for the PVSN molecule also provided a tagged variant that can report on labeling<sup>71</sup>, we synthesized the same 4-azidomethylphenyl vinylsulfonate (azPVSN) molecule, although through a slightly different synthesis path from the reported one (Scheme 2.4a). Once protein samples are treated with azPVSN, they can be linked to biotin groups through copper-catalyzed azide-alkyne cycloaddition (CuAAC) click chemistry with alkyne-biotin (Scheme 2.4b); standard gel electrophoresis followed by transfer to membranes allows immunoblotting with

streptavidin-horseradish peroxidase (HRP) conjugate for chemiluminescence reporting. We used this probe with purified enzymes to determine both the reaction progress over time and whether the catalytic residue was being preferentially labeled. Specifically, we used SHP2 full-length construct, with the wild-type and also catalytic dead C459S mutant as well as gain-of-function mutant E76K. The result showed all three groups being labeled with azPVS<sub>N</sub>, but with no noticeable difference in the rate of labeling over time (Figure 2.7). This is rather surprising, given that at least for the C459S mutant, there is one less Cys than 10 in the wild-type, for the molecule to react with: while it further backs up our observation in the previous section that the catalytic Cys is not necessarily the primary site of labeling for PVS<sub>N</sub>, this suggests that labeling at that site is so small that it is not possible to distinguish the signal difference resulting from the removal of one Cys residue there. Same experiment with SHP1 full-length constructs yielded similar results (Figure 2.7).



**Scheme 2.4: Synthesis and click chemistry of azPVS<sub>N</sub>.**



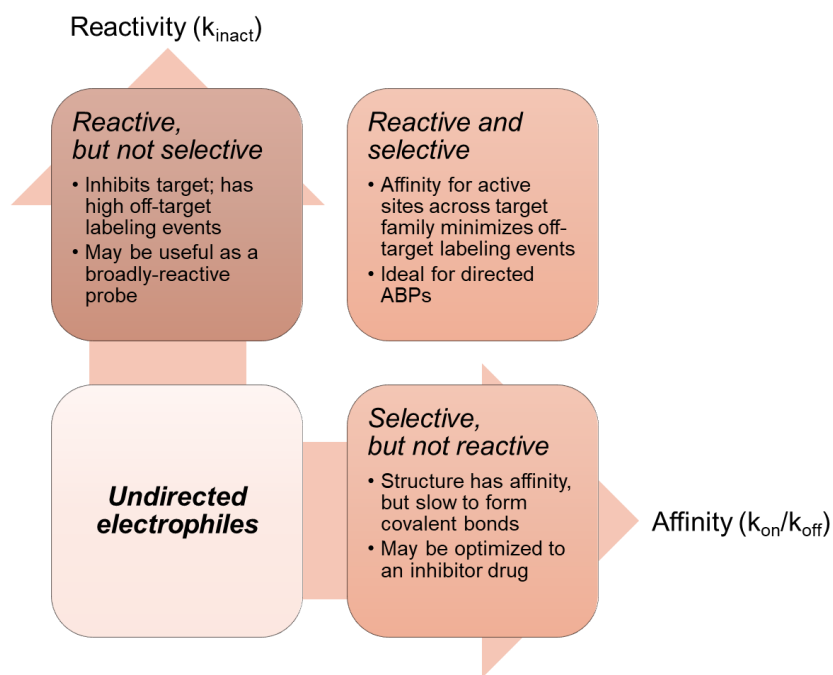
**Figure 2.7: Reaction of purified SHP1/2 with azPVSN over time.** SHP2 constructs were monitored for their labeling by azPVSN for 1 h (left); SHP1 constructs were reacted for 30 min and the bands were quantified in relation to the Coomassie stain image.

Intact protein MS using the SHP2 constructs with our compounds yielded an even more perplexing result, where the C459S mutant showed faster progress of adduct formation (data not shown). While we were not able to definitively determine why this happened, literature has reports of C459S mutant causing a conformation change, due to local charge changes that arise from the switch from cysteine to serine.<sup>79,80</sup> For SHP2, C459S mutant may be opening up more cysteines on the surface due to their stabilization of the open conformation.

## 2.6. Discussion

Based on the off-target reactivity of PVSN, it seems improbable that PVSN can be successfully employed as an ABP, as the molecule will label other residues of non-target proteins indiscriminately. Then we synthesized several new vinylsulfonate molecules with a bulkier scaffold structure and showed that they have increased reactivity. However, even at lower concentrations, the problem of off-target reactivity was not resolved. Lastly, we showed that the

tagged variant of PVSN labels wild-type and catalytic dead mutant of SHP1/2 at a similar rate, again supporting our observation that the catalytic Cys residue is not the primary residue to be labeled at physiological pH. Thus, to introduce specificity for the active site, one can imagine that the scaffold structure needs to be more complex to provide significant affinity for the active site. Once a higher affinity towards the active site has been achieved, with which probes will spend more time in the active site pocket, tuning down the warhead reactivity will be able to address the off-target reactivity problem. More moderately reactive groups can still covalently label the catalytic Cys with sufficient occupancy, but will mostly be unreactive towards surface Cys residues (Scheme 2.5).



**Scheme 2.5: Balancing probe reactivity with affinity for target active sites.**



## 2.7. Methods

*General information.* pH 7.5 Tris buffer was prepared with 50mM Tris and NaCl to adjust to an ionic strength of 150mM, and pH adjusted. Flash column chromatography was done on Selekt purification system (Biotage) using Sfar silica gel cartridges. All Coomassie Blue stained gels were imaged on ChemiDoc XRS+ gel imager (Bio-Rad).  $^1\text{H}$  and  $^{13}\text{C}$  NMR (nuclear magnetic resonance) spectra were collected at r.t. on Avance 400 MHz and 500 MHz spectrometers (Bruker). Compounds were purchased from Sigma and Fisher unless otherwise noted.

*Expression and purification of tyrosine phosphatases.* Catalytic domains of PTP1B, SHP1, SHP2, HePTP, YopH, Sts-1, and Sts-2 were prepared courtesy of Allyson Li. Expression vectors of pET-His6-SUMO-CD45, pET28-His6-TEV-SHP2-FL, and pET28-His6-TEV-SHP1-FL were prepared courtesy of Neel Shah. BL21(DE3) competent cells were transformed with each plasmid and colonies were isolated and transferred to 100 mL of LB supplemented with appropriate antibiotics at 37 °C. Once cells reached an optical density of 0.5 at 600 nm, they were moved to 2 L of TB and shaken (215 rpm) at 37 °C. After reaching optical density of 0.5, isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) was added for a final concentration of 5 mM to induce the expression of proteins, and the cultures were incubated at 18 °C overnight. Cells were harvested by centrifugation (4000 rpm at 4 °C for 30 min), resuspended in lysis buffer (pH 8, 50 mM Tris, 500 mM NaCl, 20 mM imidazole, 2 mM 2-mercaptoethanol (BME), 10 % glycerol), and lysed by sonication on ice. After separation of insoluble material by centrifugation (14000 rpm at 4 °C for 1 h), the supernatant was applied to a 5 mL HisTrap Ni-NTA column (Cytiva). The resin was washed with lysis buffer and wash buffer (pH 8.5, 50 mM Tris, 50 mM NaCl, 10 mM imidazole, 2 mM BME, 10 % glycerol). The protein was eluted with pH 8.5 buffer (50 mM Tris, 50 mM

NaCl, 500 mM imidazole, 2 mM BME, and 10% glycerol), through anion exchange on a 5 mL HiTrap Q column. The His6 tag of the collected fractions was cleaved by incubating with TEV or SUMO protease accordingly overnight. The reaction mixture was subsequently flowed through 2 mL of Ni-NTA resin (Thermo Fisher). The cleaved protein was collected in the flow-through and washes. Proteins were further purified using Superdex 75 16/600 gel filtration column (Cytiva) with SEC buffer (pH 7.5, 10 mM HEPES, 150 mM NaCl, 1 mM TCEP, 10 % glycerol); pure fractions were pooled then concentrated with centrifugal filters (MWCO 30k, Millipore). The resulting solution was aliquoted, and flash frozen in liquid N<sub>2</sub> for long-term storage at -80 °C.

*Intact protein LC-MS analysis.* Each enzyme was diluted in pH 7.5 Tris buffer so that concentration after addition of compound will be 500 nM. To 198 µL of enzyme solution, 2 µL of compound was added. When not specified, 100 µM was the final concentration of compounds. At a desired timepoint, mixture was injected onto BEH C8 column (Waters) on a UPLC-MS system (Xevo QToF, Waters). LC was carried out with gradient of 5 % to 95 % of acetonitrile (0.02 % formic acid) for 8.5 min. The protein typically eluted around 4 min; this peak on the chromatogram was integrated and subsequently deconvoluted using MaxEnt1 algorithm. Subsequently, peaks were chosen according to the theoretical MW of each adduct, within a range of 5 Da, and integrated for the signal intensity  $\pm$  1 Da. Total intensity of all detectable enzyme species was used as 100 % to normalize signal against.

*Substrate catalysis assay of phosphatase inhibition.* All assays were performed in pH 7.5 Tris buffer. All measurements were taken at 30 °C on a microplate reader (Synergy Neo2, Biotek). Each enzyme was diluted in pH 7.5 Tris buffer so that concentration after addition of compound

will be 250 (or sometimes 500) nM. For preincubation, 2  $\mu$ L aliquot of each compound stock in DMSO was dispensed into a well on a 96-well plate. A DMSO group was included to serve as an uninhibited control. Subsequently, 198  $\mu$ L enzyme solution was mixed in each well. Similarly, a well was prepared with DMSO and Tris buffer without enzyme to serve as a baseline control. At desired time points, 20  $\mu$ L aliquots of preincubation mixture were placed on a 96-well plate ( $n = 3$ ), and 180  $\mu$ L pNPP solution was added to each well. pNPP solution was prepared fresh with 100 mM DTT, and 0.1 % (w/v) BSA in Tris buffer for a final pNPP concentration of (typically) 20 mM during measurement. Absorbance at 405 nm was measured every 30 s for 10 min, and was fit with simple linear regression to obtain an initial rate from the slope, and the slope of the baseline control was subtracted from each group. The residual activity of each enzyme after incubation with compounds was calculated as the baseline-adjusted slope of the treated group divided by the adjusted slope of the untreated group.

*Michaelis-Menten assay of substrates.* General measurement method was similar to inhibition assay above. For fluorogenic substrates, fluorescence was read instead with  $\lambda_{\text{ex}} = 360$  nm and  $\lambda_{\text{em}} = 449$  nm for MUP and  $\lambda_{\text{ex}} = 358$  nm and  $\lambda_{\text{em}} = 455$  nm for diFMUP. Final enzyme concentration after adding substrate solution was 100 nM except for SHP2 FL E76K, which was adjusted to 10 nM. Each substrate solution was prepared in 8 steps of serial twofold dilution. All absorbance or fluorescence values were converted to product concentrations by a separate calibration curve prepared from the same 8 concentrations of purchased pNP, MU and diFMU.

*Peptide digest MS assay.* Samples were prepared in the same way as intact protein MS, but with double scale (400  $\mu$ L). Once preincubation was done, 10  $\mu$ L of 200 mM DTT solution was added

to quench the reaction, and the mixture was dried on a centrifugal vacuum concentrator (Labconco). Each sample was resuspended in 40  $\mu$ L RapiGest (Waters) solution and heated to 60  $^{\circ}$ C for 30 min. After cooling to r.t., 4  $\mu$ L of 150 mM IAA solution was added, and samples were shaken in the dark for 30 min. Then, 1  $\mu$ L of 0.4 g/L trypsin (sequencing grade, Promega) was added and incubation was done overnight at 37  $^{\circ}$ C. Samples were acidified to pH 3 with formic acid. Measurement was done by injecting onto HSS T3 peptide column (Waters) on the same UPLC-MS system. Peptides were resolved on a 90-min gradient of 0.5 % to 95 % acetonitrile with 0.02 % formic acid. MS acquisition was done on MS<sup>E</sup> mode with cone voltage ramp of 20-45 V for the high energy part. Resulting spectra were processed on UNIFI software version 1.9.4.053 (Waters) by searching within the single protein in each sample. For multiple samples in a group, all compound modifications were searched for all samples (C only, except for PVSN which were C, H and K), in addition to C carbamidomethylation, M oxidation and N-terminal acetylation (all as variable).

*General procedure for vinylsulfonate synthesis.* diFMU precursor was purchased from CarboSynth. 4 mmol of alcohol precursor was dissolved in 12 mL DCM on ice. 8 mmol TEA was added, then 4.8 mmol 2-chloroethanesulfonyl chloride was added. The mixture was stirred for 30 min. Then the solvent was removed on a rotary evaporator, mixture was redissolved in DCM, and white solid precipitate was removed, after which the solution was evaporated again. Product was obtained after purification with flash column chromatography using hexane and ethyl acetate as eluents.

*Synthesis of alkyne-biotin.* First, biotin-NHS was synthesized as follows. To a solution of biotin (4 mmol) in DMF (120 mL) was added 20 mmol TEA and 4.8 mmol di(*N*-succinimidyl) carbonate. The mixture was stirred for 6 h, and solid precipitate was washed with methanol. The resulting biotin-NHS was dissolved in 400 mL DMF, and to the solution 7 mmol propargylamine and 10 mmol TEA were added. After stirring for 6 h, the mixture was concentrated and purified with a silica gel column chromatography, eluting with 10:1 methanol:chloroform.

*Synthesis of azPVSN.* 4 mmol 4-hydroxybenzyl alcohol, 4 mmol triphenylphosphine and 4.8 mmol sodium azide were dissolved in 8 mL of 4:1 DMF:carbon tetrachloride. Mixture was stirred at 90 °C overnight. Extraction was performed by adding 50 mL water and 100 mL diethyl ether, and then washing the organic layer twice more with 50 mL water. Afterwards, the organic phase was cooled to 0 °C and dried with sodium sulfate. Eluting from a short silica column with 1:4 ethyl acetate:hexane and concentrating afforded 4-(azidomethyl)phenol. azPVSN was synthesized according to the general procedure.

*Blot assay of protein labeling with azPVSN.* 2.5 μM enzyme solutions were made with succinate buffer (pH 6, 50 mM sodium succinate, 1 mM EDTA, ionic strength adjusted to 150 mM with NaCl), for a total volume of 42.5 μL each. 0.43 μL of azPVSN 2 mM stock was added. After desired time, 1.25 μL of alkyne-biotin stock (5 mM in DMSO), 1.25 μL of 50 mM sodium ascorbate aqueous solution, 3.75 μL of TBTA stock (1.7 mM in DMSO) and 1.25 μL of copper sulfate stock (50 mM in water) was added. After 1 h of incubation, 50 μL sample loading buffer was added and standard SDS-PAGE protocol was followed. Transfer to membrane was done using Turbo-Blot (Bio-Rad), blocking was done for 1 h in 5 % BSA (w/v) in TBS-T, then the membrane

was incubated in 10 mL 1:1000 streptavidin-HRP solution (5 % BSA, TBS-T) for 1 h. After washing 3 times in TBS-T, enhanced chemiluminescence reagents (GE Healthcare) were added and chemiluminescence was imaged on a ChemiDoc XRS+ (Bio-Rad) gel imager. Another gel was loaded identically and stained with Coomassie Blue for loading control.

## Chapter 3: Designing active site-directed probes

### 3.1. Background

To date, many different inhibitors of protein tyrosine phosphatases have been reported, with some undergoing clinical trials.<sup>64</sup> While efforts to find covalent inhibitor drugs are still lagging, covalent inhibitors can serve as starting points for potential activity-based probes, as it is not the goal to make them specific towards a single PTP. Certainly, just because there is no need to make a probe specific to just one enzyme does not mean that we can pick an inhibitor and use it: in the previous chapter, we showed that inhibition through covalent labeling of the catalytic residue is only one part to a successful ABP, and selectivity for the catalytic residue, with low off-target reactivity, is necessary.

There are different modes of action for covalent inhibition of PTPs: irreversible oxidation of the catalytic cysteine residue, covalent bond formation in the active site, and covalent bond formation in an allosteric site. While in some cases exactly where and how inhibitors covalently modify PTPs have not been shown, in most cases the target is the active site cysteine, with a few exceptions such as the quinone methide-generating probe or a lysine-targeting molecule. Allosteric inhibitors also reacted with a Cys residue in other positions, which is not surprising as cysteine is the most common subject for any covalent targeting strategy.<sup>81</sup>

While there are different strategies for making probes that covalently react with Cys, most reported molecules are electrophiles that are subject to a nucleophilic attack by the thiol group in Cys. Another notable approach is using a thiol molecule to form a disulfide bond, which would be an oxidative reaction. A sizeable array of thiol-reactive groups have been analyzed in many previous studies, in a context unrelated to PTPs as targets.<sup>82-88</sup>

To design probes that are selectively directed towards PTPs, we need to find a structure with a reasonable specificity towards the target active sites. When we are designing probes like this in a target-centric fashion,<sup>60</sup> it may be useful to assess different thiol-reactive groups specifically in relation to our targets. For example, we already showed some rational design processes in the earlier chapter, taking cues from different substrate molecules and their bulkiness. Here we continue that effort, while also enlisting help from other inputs such as computational drug discovery tools.

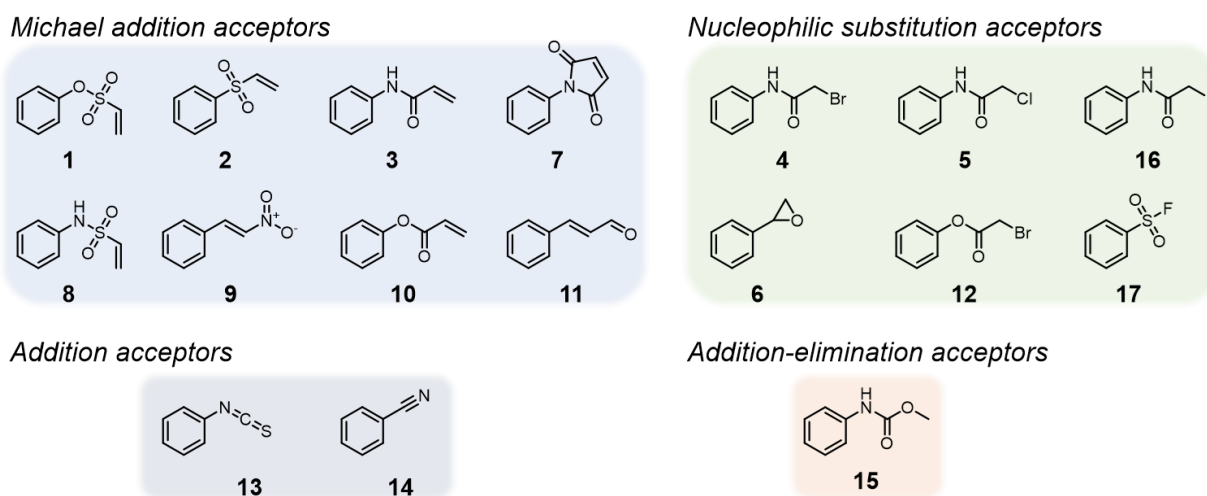
Fundamentally, as we are seeking a family-wide probe, it may not make sense to screen highly complex structures in a high-throughput screen looking for very potent inhibitors for individual PTPs. We reasoned that an approach similar to fragment-based drug discovery, which looks for useful leads that can be optimized down the line, is more suitable for our purposes.<sup>89</sup> Hence, we constructed a fragment library consisting of fewer, simpler scaffold structures that aim to sample from a large chemical space with a few warheads of varying reactivity. To accomplish this, we first assessed thiol-reactive groups with PTPs, then built a library of candidate compounds to assay.

### **3.2. Biochemical assays of various thiol-reactive groups**

Electrophilic functional groups were chosen based on their previous reports of thiol reactivity and PTP inhibition.<sup>64</sup> Many compounds were Michael acceptors such as vinylsulfonate (**1**), vinyl sulfone (**2**) and acrylamide (**3**), and nucleophilic substitution acceptors such as  $\alpha$ -haloacetamides (**4**, **5**, **16**) and epoxide (**6**), with some other addition acceptors and addition-elimination acceptors. (Figure 3.1) There is a big span of intrinsic reactivity towards thiols among the chosen molecules, and some of them are known to react with other functional groups as well.



We chose compounds that have a simple phenyl scaffold; while this is not expected to function as a recognition element specific to the target active sites, it has been suggested that the phenyl group can take advantage of binding in the active site pocket the same way a phosphotyrosine substrate would.<sup>71</sup> Indeed, methyl vinyl sulfone and methyl acrylate showed slower reactions with PTPs (data not shown). Compounds were mostly obtained from commercial sources.

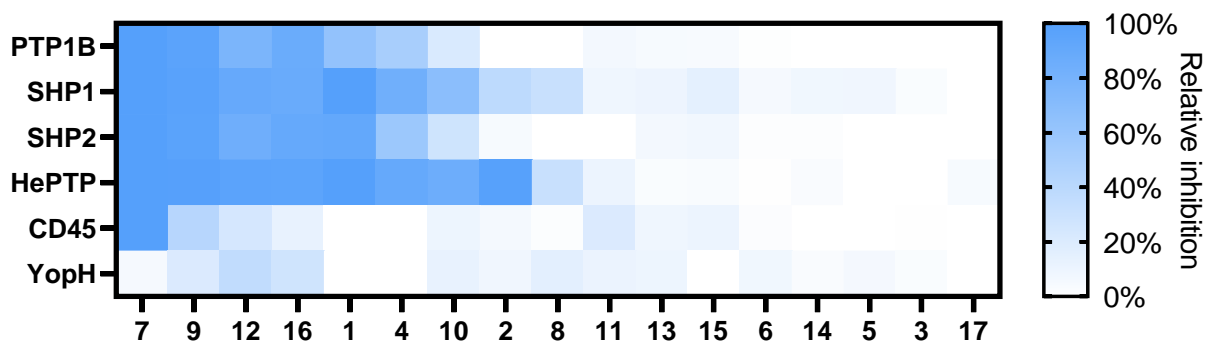


**Figure 3.1: Thiol-reactive groups assessed for reactivity.**

These compounds were then tested for their reactivity towards tyrosine phosphatases. Various PTPs were used: there are classical PTPs, namely PTP1B, SHP1, SHP2, HePTP, CD45 and a bacterial *Yersinia* PTP, YopH, which is similar to eukaryotic PTPs in structure and is one of the most efficient PTPs known,<sup>90</sup> frequently employed in biochemical assays of substrates and inhibitors for PTPs.<sup>91</sup> In most cases, the catalytic domain of the respective enzymes was expressed, purified and used in assays.

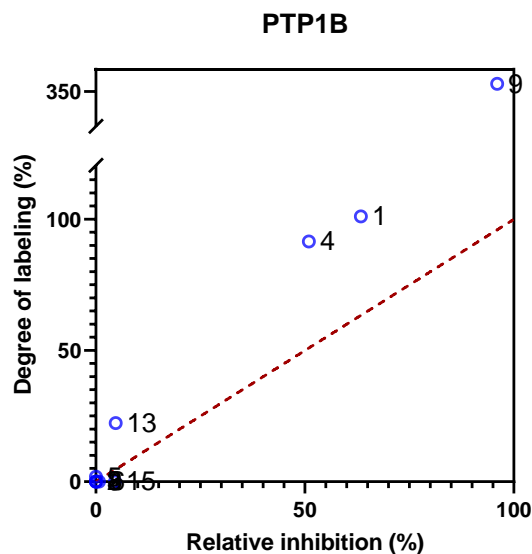
In a kinetic assay for the reaction of the candidates with the active site cysteine, several purified PTPs were preincubated with the compounds, after which the remaining activity of

enzymes was measured through catalysis of a colorigenic substrate dephosphorylation, as described in the previous chapter. The assay was optimized to maximize enzyme in solution to increase signal-to-noise over the window of the experiment and ensure linearity between enzyme concentration and activity from the product formation signal readout. Preincubation was done with 500 nM enzyme and 100  $\mu$ M of candidate compounds in Tris buffer for 1 h; the substrate reaction was carried out with 50 nM of preincubated enzyme and 20 mM pNPP. Due to this dilution of the mixture, any noncovalent interaction between the compound and enzyme will be attenuated in the substrate reaction step. On top of that, the excess of dithiothreitol (DTT) stops further inhibition by compounds during this stage. The resulting absorbance over time graph was linear and fit well with simple linear regression. This demonstrated that different PTPs have different inactivation rates with the same compounds, where in some cases compounds that inhibited one PTP, such as HePTP, completely, would not inhibit another PTP such as CD45 at all. As expected, there was a wide range in the degree of inhibition between the compounds; regardless, the result showed that compounds with relatively high reactivity inhibited PTPs across the board, save for YopH, which seemed unresponsive to most of the compounds (Figure 3.2). Some trends were confirmed: for carbonyl and sulfonyl-containing groups, those that have an *O*-bridge with the scaffold (**1**, **10**, **12**) were more reactive than their counterparts (**8**, **3**, **4**) with an *N*-bridge; between haloacetamides, the order of reactivity, **16** > **4** > **5**, was pronounced, as expected from their leaving groups I, Br and Cl.



**Figure 3.2: Inhibition of PTPs by phenyl compounds.** PTPs were treated with 100  $\mu$ M compounds for 1 h. Compounds are ranked by descending average inhibition from left to right.

Then, to assess active site specificity, the extent of labeling for each compound was tested with our intact protein MS method, after using the same conditions for preincubation. The result showed that with the more reactive groups, such as bromoacetamide or vinylsulfonate, species with multiple covalent labeling consistently appeared across several PTPs. In most cases, the total extent of labeling, calculated by a weighted average of how many probe molecules labeled each protein (Extent of labeling) =  $\frac{\sum n(\text{Signal of E+nI species})}{\sum (\text{Signal of E+nI species})}$ , was higher than the extent of labeling in the active site Cys, as inferred from the degree of relative inhibition (Figure 3.3). This reveals that other residues are also reacting with the compounds. An ideal probe that only labels the active site will fall on the plotted diagonal, as the two values will be the same. While most of the compounds tested are known to be relatively specific to Cys, some compounds, such as **9**, showed far higher adduct species than is possible with the labeling of just the Cys residues, supporting off-target labeling of other amino acid residues as well.

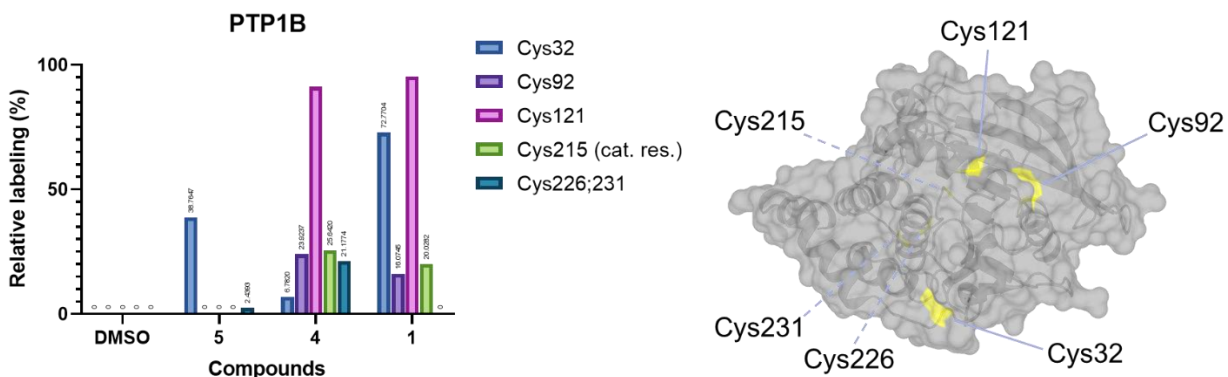


**Figure 3.3: Degree of labeling compared with inhibition data.** 10 compounds were compared for PTP1B; on the bottom left corner are 2, 3, 5, 6, and 8. The dashed line is a plot of  $y = x$ .

However, there were many compounds that did not show any significant labeling on intact protein MS, suggesting that those warheads may be improved upon to conserve the low off-target reactivity while bringing up the on-target reactivity. *O*-bridged compounds **10** and **12** showed fragmentation of adduct species, which made the results harder to quantify; in addition, this suggests that they are not suitable to be ABPs with their relatively weak ester linkage after labeling, which may even be susceptible to hydrolysis by esterases within the cell.

Lastly, to test if the inhibition was due to labeling at the active site, we performed two assays on a subset of the tested compounds. First, after incubation with compounds using the same conditions, a few PTPs were digested with trypsin and analyzed with MS<sup>E</sup>, which confirmed that there is labeling at the active site. Although there were moderate differences between the composition of labeled peptide in total peptide and the residual activity obtained from substrate catalysis, this may be attributed to limitations in quantifying MS data, as we did not control for

differences in tendencies to fragment, etc. between different peptides (Figure 3.4). As with PVSN, this also showed that compounds were labeling other Cys residues, which also concurs with intact protein MS results. Second, preincubation was carried out in the presence of a competing noncovalent inhibitor. This will be discussed in a latter section.



**Figure 3.4: Compound labeling sites for PTP1B as determined from peptide MS.** Each value represents the proportion of peptides detected that have the respective compound modifications. On the right, residues are represented on a PTP1B structure. (PDB: 1KAK)

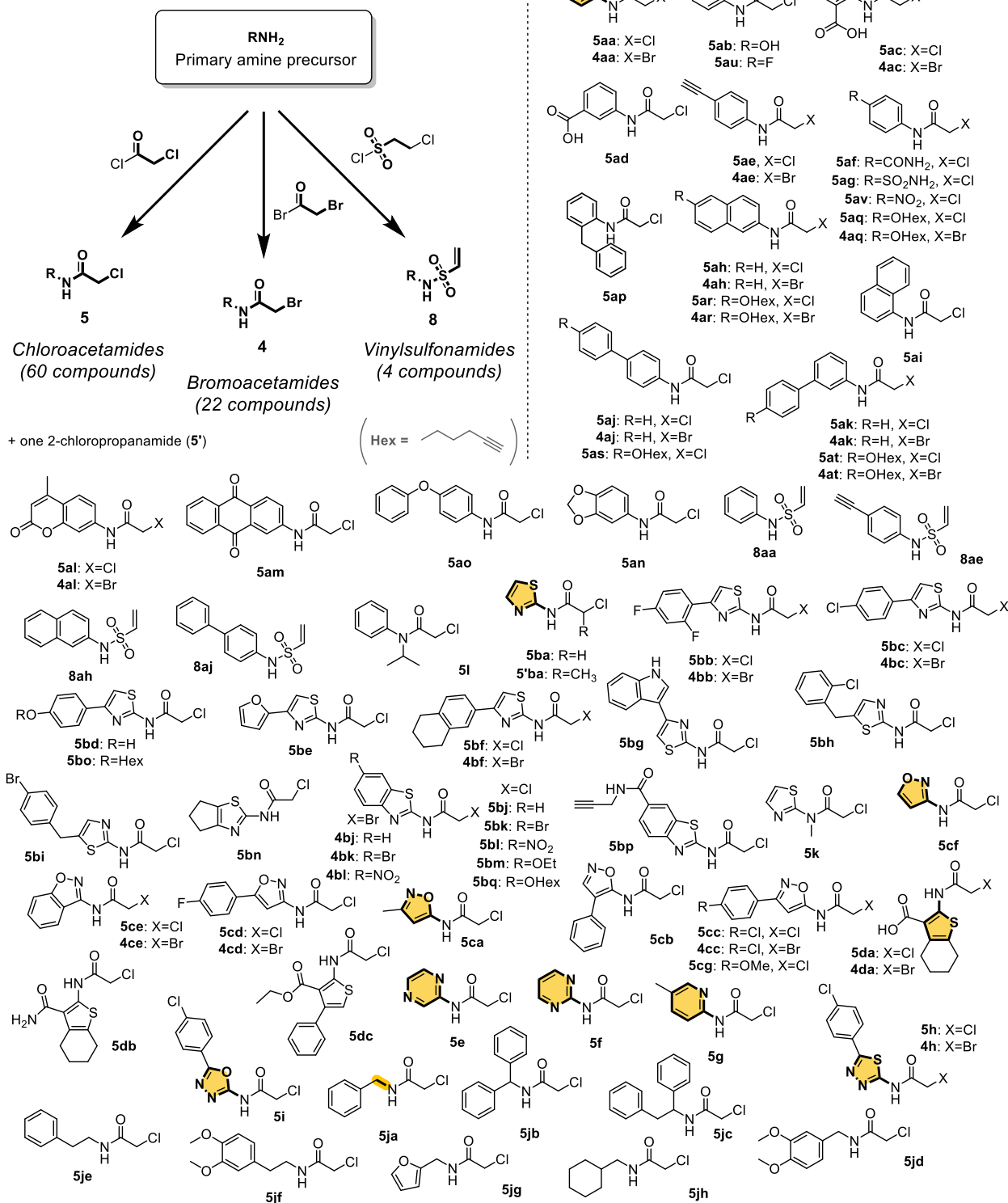
### 3.3. Screening a fragment-like library

As explained in the beginning of this chapter, we sought a scaffold structure that provides reasonable affinity towards the active site, so that the compound will still react with our target proteins even with a less reactive warhead. To understand how different compound structures affect enzyme inhibition, we tested a wider range of molecules, with various scaffold structures connected to a few reactive groups. During this process, we went through several redirections, starting from bromoacetamides (**4**) and some vinylsulfonamides (**8**), as an attenuated counterpart of **1**, until arriving at chloroacetamides (**5**) and in a few cases acrylamides (**3**). Chloroacetamides are known to have a reactivity on the higher side for therapeutic purposes, being more reactive

than acrylamides, which are commonly employed in covalent inhibitor drug molecules.<sup>85,88</sup> Ultimately, we arrived at the warheads with relatively low reactivity for PTP active sites.

As was the case for vinylsulfonates (**1**), we hypothesized that having a larger scaffold or additional functional groups as recognition elements would confer more affinity towards target active sites in the molecules, and thus gain faster reaction selectively towards the catalytic Cys residue. Therefore, we constructed this fragment library so that different structural motifs are represented: the scaffold part that is directly connected with the warhead is diverse, including phenyl rings (**a**), heteroaromatic rings such as thiazoles (**b**) and isoxazoles (**c**), and aliphatic chains (**j**). This can also be thought of as a rational design process, as the heteroaromatic rings were chosen from common recognition elements in medicinal chemistry<sup>92-95</sup>. Commercial availability was a factor in selection, and in cases compounds were to be synthesized, they were chosen mostly from available amine precursors for a one-step synthesis. In total, we selected about 90 compounds (Figure 3.5).

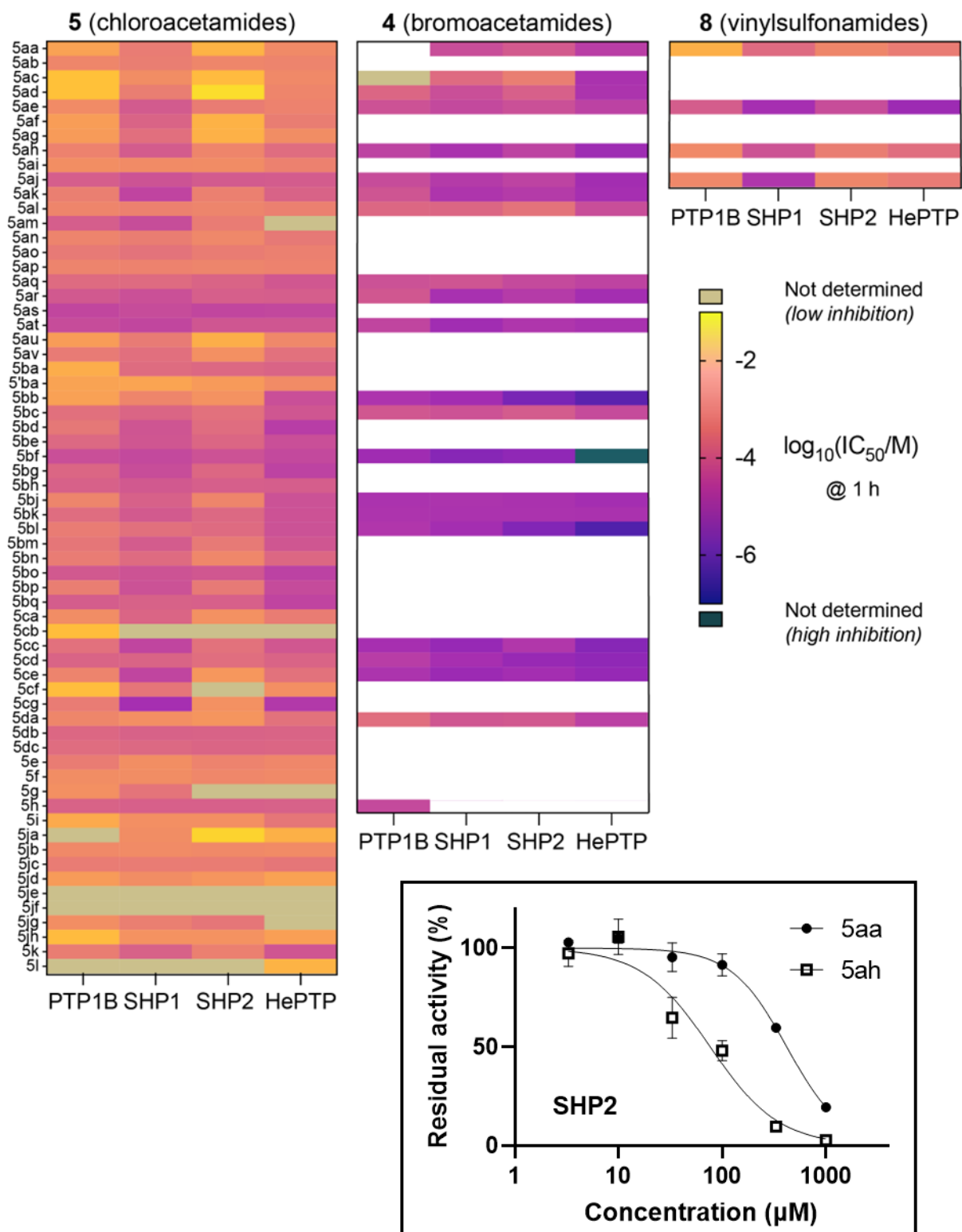
## Library Composition



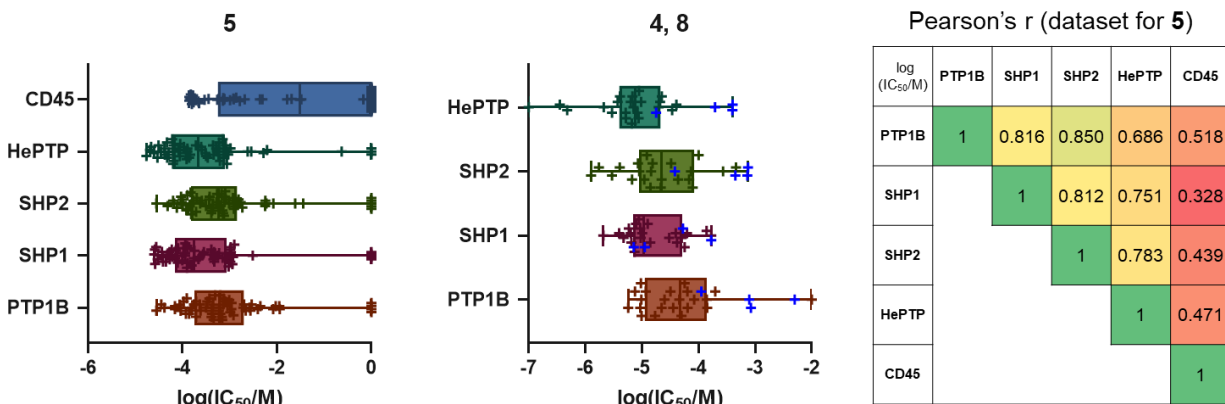
**Figure 3.5: Fragment-like library.** Structures for all compounds are shown; elementary structures for chemotypes are bolded and highlighted with yellow. Composition and synthesis pathways are shown for the three warheads.

For the inhibition assay, we tested across compound concentrations ranging from 3.3  $\mu\text{M}$  to 1 mM and a few incubation times ranging between 30 min and 4 h (experiments done with Sarah Xi), allowing us to look at the dose response of PTP inhibition and produce  $\text{IC}_{50}$  values for different time points (Figure 3.6). This way we were able to easily compare the performance of different compounds across PTPs. Moreover, by using all timepoints, we were able to fit the observed rate of reaction versus compound concentration graph, yielding the kinetic constants  $K_I$  and  $k_{\text{inact}}$ . We discuss this process in depth in the next chapter.





**Figure 3.6: Screening the library for PTP inhibition.** (inset) An example showing how data is fit to dose response curves; heatmap of  $\log(\text{IC}_{50})$  data for all compounds at 1 h. Compounds with same scaffold structure are placed in the same row.



**Figure 3.7: Trends in individual PTPs.** Distribution of IC<sub>50</sub> is shown for chloroacetamides (5), bromoacetamides (4) and vinylsulfonamides (8, marked in blue). Correlation data for 5 is shown on the right.

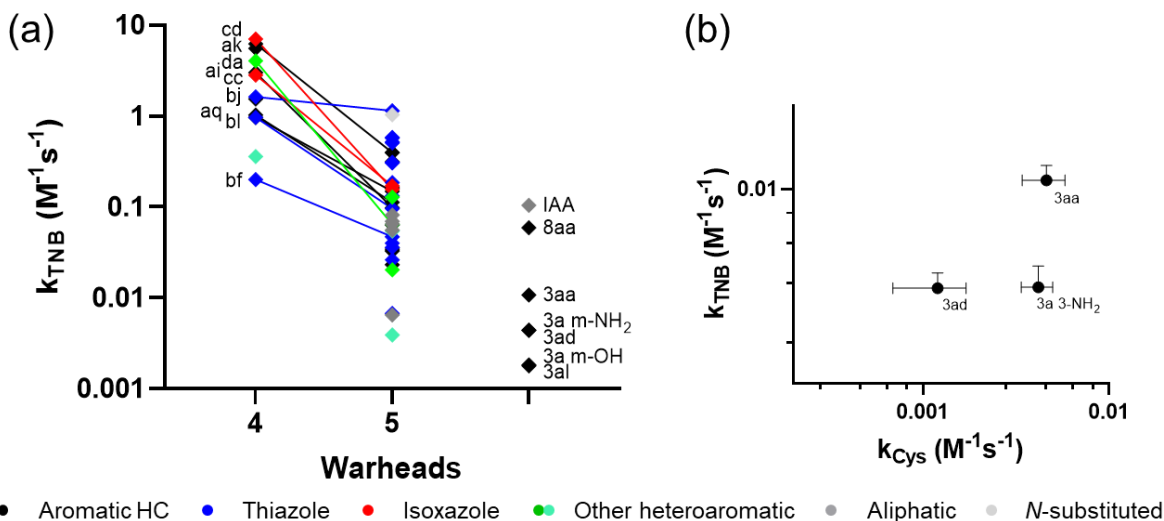
The first observation from our screen was that different phosphatases showed different intrinsic reactivities, on average, for the compounds in the library. For example, on average, and in most cases, HePTP was inactivated at a faster rate than others, and CD45 showed low inhibition with most of the tested compounds at the concentration of 100  $\mu$ M. There was also a correlation between the PTPs in terms of their responsiveness to different compounds; CD45 had a lower correlation with others in general, while PTP1B and SHP2 displayed the highest correlation. This correlation was a surprising result as SHP1 is closer to SHP2 in sequence and structure (Figure 3.7).<sup>16,20</sup> PTP1B and SHP2 pair had the highest correlation in the dataset for 4 and 8 as well ( $r = 0.895$ ). As for compound structures, thiazoles stood out in terms of their reactivity towards the active site. Isoxazoles displayed higher preference for SHP1 and HePTP on average. As expected, going from rudimentary structures (5aa, 5ba, 5ca and 5cf) to multi-ring structures generally resulted in lower IC<sub>50</sub> values. For a similar structure, compounds with a thiazole ring in place of a phenyl ring showed faster inactivation of the PTPs. Aliphatic structures generally fared worse;

benzyl (**5ja**) did not inhibit PTP1B, for example, but more phenyl substitutions improved the IC50 values (**5jb-c**). Phenethyl structures showed almost no reactivity (**5je-f**).

### 3.4. Intrinsic reactivity assays

Another factor to be considered about different scaffold structures is that it can also change the intrinsic reactivity of the probe molecule. To determine the reactivities of compounds independent of the enzyme active site, we followed a previously reported assay, measuring second-order rate constants of the reaction between compounds and the small molecule 2-nitro-5-thiobenzoate (TNB<sup>2-</sup>) (experiment by Sarah Xi).<sup>86</sup> To accompany compounds with higher reactivity, we used nonlinear regression instead to obtain higher quality fits for heteroskedastic data where significant deviation was occurring during later timepoints<sup>96</sup>. The result showed that the intrinsic reactivities have a large range, with the reaction rate constant ( $k_{\text{TNB}}$ ) ranging from  $0.0039 \pm 0.0004$  to  $1.14 \pm 0.01 \text{ M}^{-1}\text{s}^{-1}$  within chloroacetamides (**5**), and  $0.20 \pm 0.01$  to  $6.25 \pm 0.28 \text{ M}^{-1}\text{s}^{-1}$  within bromoacetamides (**4**) (Figure 3.8a). On average, compounds of the same scaffold had about an order of magnitude difference between **4** and **5**. Several acrylamides (**3**) were tested for their intrinsic reactivity as well, and they had a lower rate constant on average as expected.

To further support measurements from the above, we employed an NMR-based assay, modified from a report,<sup>97</sup> to monitor reactivity toward a *N*-acetyl-cysteine methyl ester on select compounds (**3aa**, **3ad** and 3-aminophenyl acrylamide). Due to the higher concentration of compounds used in this assay, a slow warhead such as **3** is ideal. While the rate constants for individual molecules varied by about tenfold between the two assays, with the aryl thiobenzoate reacting faster than alkyl cysteine derivative does as expected, the trend was mostly conserved (Figure 3.8b).

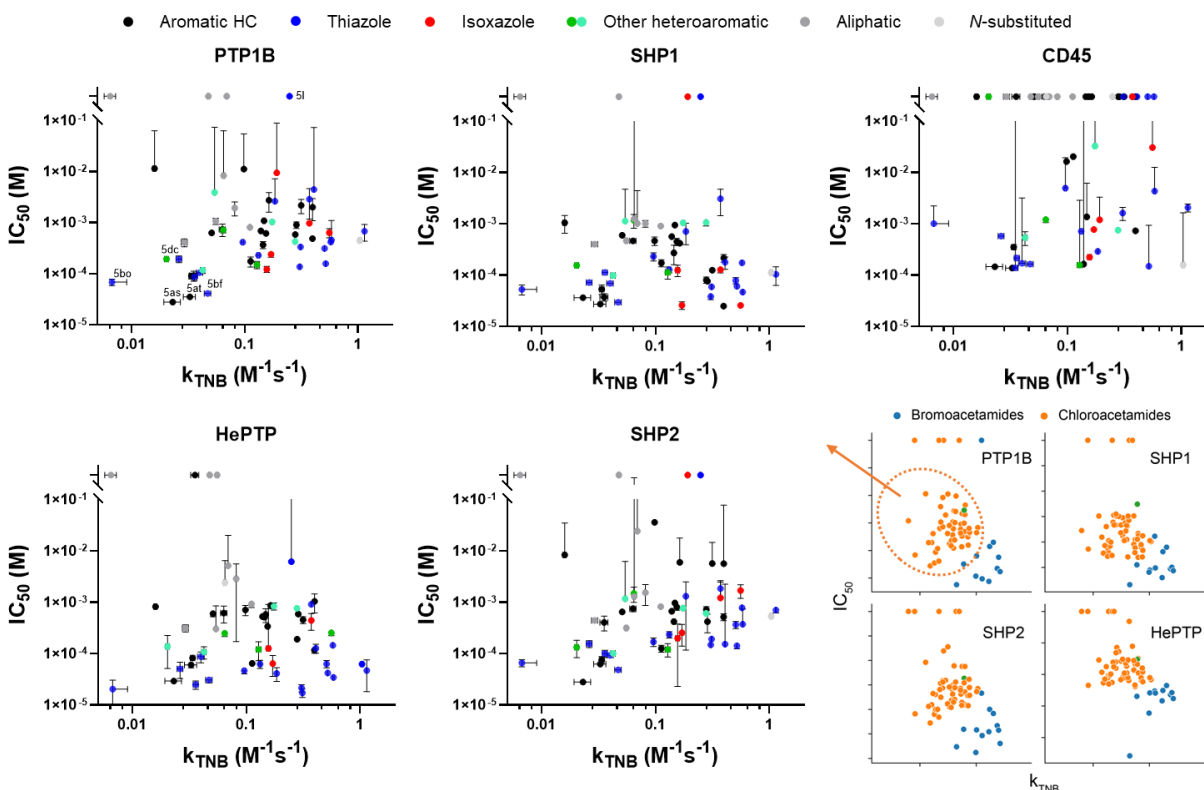


**Figure 3.8: Intrinsic reactivities of the library compounds.** (a) Rate constants for TNB reaction classified by warhead. Scaffold structures that are shared between chloroacetamides (5) and bromoacetamides (4) are connected and denoted with compound codes. (b) Rate constants for Cys reaction of acrylamides (3) compared with their TNB constants.

### 3.5. Two-dimensional analysis of the library

Next, we compared the intrinsic reactivity of each probe ( $k_{\text{TNB}}$ ) with its reactivity towards the active site Cys (represented by  $\text{IC}_{50}$ ) (Figure 3.9). Bromoacetamides (4) and chloroacetamides (5) clustered to themselves cleanly, with the population of 4 having lower  $\text{IC}_{50}$  and higher  $k_{\text{TNB}}$ , as expected. In regards to chemotypes, we analyzed compounds of 5. several trends stood out: first, although thiazoles (5b) were not noticeably more reactive with small molecule thiols than the phenyl-based structures (5a) (log mean -0.839 and -1.01 respectively; t-test showed nonsignificant  $p = 0.42$ ), they are more reactive towards the active site Cys, with a statistically significant difference for HePTP ( $p = 0.0003$ , for the compounds that have both data). This could mean that the thiazole ring is enhancing the reactivity of the probes selectively for the catalytic residue,

mainly through increasing the affinity towards the active site. In broader terms, there was not a significant correlation between the intrinsic reactivity and the  $IC_{50}$  values.



**Figure 3.9: 2D analysis of  $IC_{50}$  and intrinsic reactivity.** Bottom right: compounds of bromoacetamides (**4**, blue) and chloroacetamides (**5**, orange) clustered within its group on the plot of  $IC_{50}$  vs.  $k_{TNB}$ ; then **5** was analyzed further. Meaningful  $IC_{50}$  was not determined for groups with low inhibition; those are shown at the top.

We also noted some structure-activity relationship (SAR) trends: introducing a negatively charged functional group in the ortho position of a phenyl ring resulted in a lower intrinsic reactivity (**5ab-d**) than **5aa**. While this may be explained by resonance effects (also compare 2-carboxy **5ac** with 3-carboxy **5ad**), the reactivity towards the active site was not so different from the unsubstituted structures; **5ab** had lower  $IC_{50}$ s. (Figure 3.11) While active sites of PTPs are known to prefer negatively charged molecules in general, due to phosphate being a substrate, past

studies have shown that an *ortho*-carboxylic acid group can have hydrogen bonding interaction with the conserved general acid Asp residue as well as a salt bridge interaction with a conserved Lys residue close by.<sup>98</sup> This suggests that carboxylic acid, and similarly hydroxyl group on the *ortho* position may improve binding affinity, thereby offsetting the lower intrinsic reactivity.

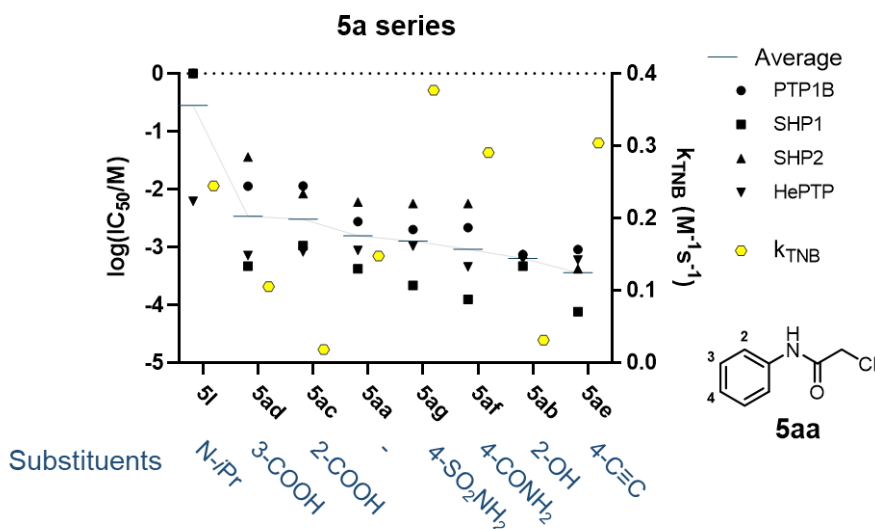
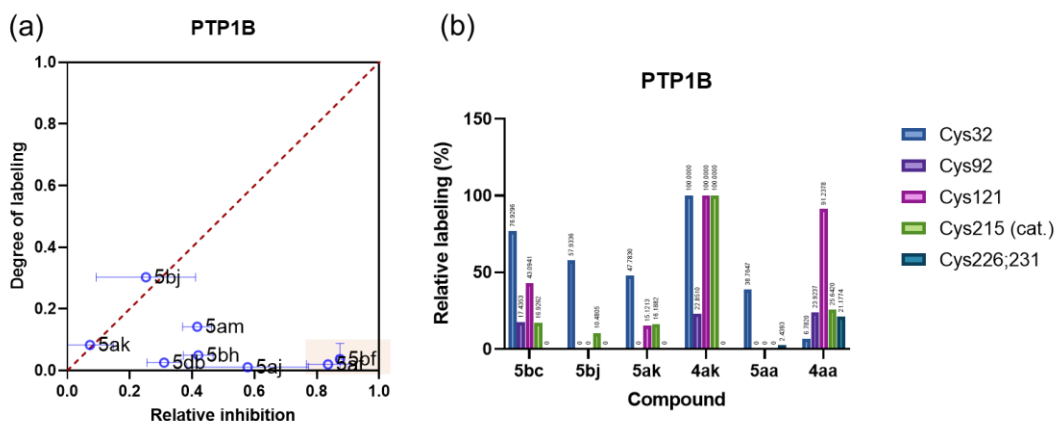


Figure 3.10: SAR of phenyl chloroacetamide with substituents.

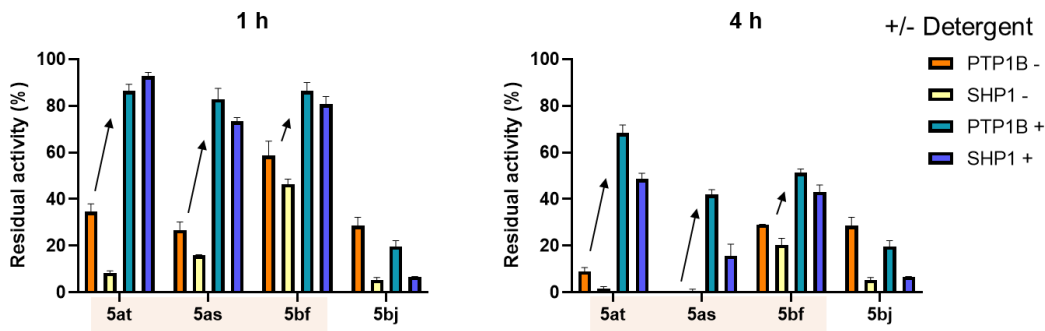
### 3.6. Further analyses and identifying false leads

We utilized intact protein and peptide MS data to further validate our inhibition assays. Many chloroacetamide compounds showed slower progress of labeling reactions in the intact protein MS than what was expected from inhibition data (Figure 3.11a). Fragment MS of select compounds showed that other Cys residues such as Cys32 were still favored over the catalytic residue (Figure 3.11b). We also employed additional experiments to test and filter out compounds that show high inhibition results for reasons other than orthosteric modification, which may lead to false hits or leads and hence are known as “nuisance inhibitors.”<sup>99</sup> Compounds **5as-t** and **5bf**

stood out in their low IC<sub>50</sub> values, but intact protein MS data above showed almost no adduct formation. Additionally, they had shown visual signs of aggregation when mixed with protein samples. Subsequent substrate catalysis experiments in the presence of 0.01 % Triton X-100 confirmed that there is significant inhibition due to the aggregation of protein in the sample (Figure 3.12). In this case, time-course data may be used for additional scrutiny, as aggregation effects are generally not time-dependent.

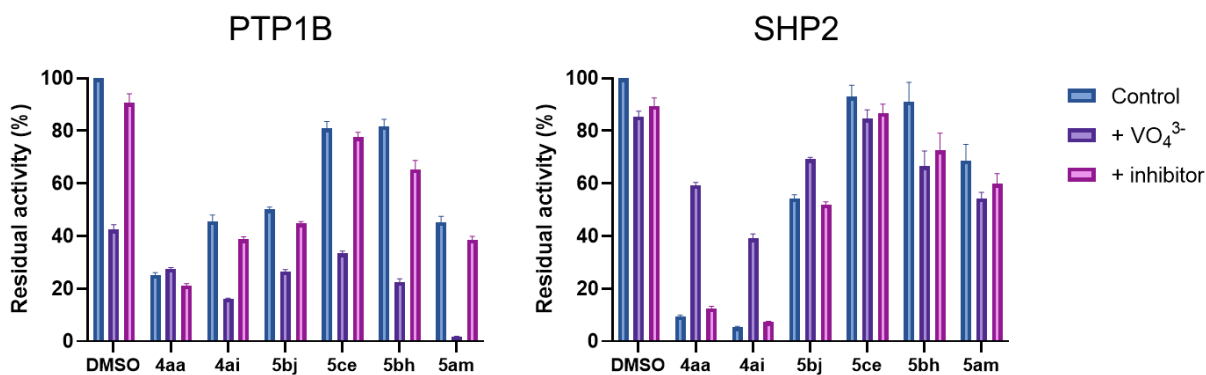


**Figure 3.11: MS analyses of library compounds.** (a) Degrees of labeling obtained from intact protein MS for select compounds were compared to the inhibition data. (b) Site-specific labeling data as determined from peptide MS.



**Figure 3.12: Aggregating compounds with and without detergent.** Three aggregation-inducing compounds (highlighted) compared with a normal compound, 5bj, in their inhibition performance with and without detergent during preincubation.

We again sought to confirm that most probes are mainly inhibiting target enzymes by covalent modification of the catalytic residue. Here, residual activity was measured from the same substrate dephosphorylation catalysis method, but with added noncovalent orthosteric inhibitors in the preincubation stage (Figure 3.13). A higher dilution was used when mixing with pNPP solution, to minimize the effects of noncovalent inhibition. Both orthovanadate, a commonly used phosphatase inhibitor, and drugs with high affinity for each enzyme – TCS-401 for PTP1B and GS-493 for SHP2 – were used. Decreases in inhibition with the presence of a competitor for the active site showed that compounds inhibit PTPs through covalent labeling of the active site Cys. Notably, even among orthosteric covalent inhibitors, the extent to which residual activity is rescued is different, as compounds with higher  $k_{on}$  and/or low  $k_{off}$  will perform better in the competition for active site occupancy. Thus, while the increase in activity with the addition of noncovalent inhibitors suggests orthosteric binding, as in the case of **4aa**, **4ai**, and **5bj** to a smaller extent, an orthosteric inhibitor can still be largely unaffected depending on the parameters.



**Figure 3.13: Inhibition of compounds in the presence of a competitor.** All activities are relative to that of the DMSO control group, after treatment with 100  $\mu$ M compounds for 2 h.



### 3.7. Computational docking

We sought to compare the experimental results with computational docking scores, to see if computational methods can predict affinity of a given scaffold for PTPs. For this, we used Glide in Schrödinger software suite for docking with 8 known substrates and 10 known inhibitors together with a 1K decoy set reported with Glide. Crystal structures of PTP1B with both open and closed WPD loop conformations were used. Result showed most inhibitors and substrates scoring better than decoys, demonstrating the ability to predict favorable interactions; bulkier molecules such as *O*-methylfluorescein phosphate and GS-493, while still placing in the top 10 %, did not perform as well as the other substrates and inhibitors. Interestingly, two decoy molecules topped the average docking scores, seemingly favored by the abundance of negatively charged functional groups; the rest of the top 15 molecules were 7 substrates followed by 6 oxalylaminobenzoic acid inhibitors<sup>98</sup> (Table 3.1). However, the top decoy, phenyl 2-sulfoacetic acid, did not display noncovalent inhibition when tested with SHP2. Another calculation protocol CovDock, which also accounts for covalent bond formation, did not provide a good estimate of experimentally obtained inhibition data (data not shown). From these observations, we concluded that computational docking is to be used more as a supplementary tool for docking poses or preliminary estimations, rather than for screening.

**Table 3.1: Glide docking scores of the highest-ranking compounds.** Lower score means that the docking pose is more stable.

Name	SMILES string	Average docking score
Decoy 460935	<chem>O=C(O)[C@@H](S(=O)(O)=O)C1=CC=CC=C1</chem>	-7.86
Decoy 385089	<chem>CC1=CC(C)=CC(C[C@@H](C(O)=O)[C@H](C(O)=O)C)=C1</chem>	-7.3

pTyr	<chem>O=C(NC)[C@H](CC1=CC=C(OP(O)(O)=O)C=C1)NC(C)=O</chem>	-7.27
diFMUP	<chem>O=C1C=C(C)C2=C(O1)C(F)=C(OP(O)(O)=O)C(F)=C2</chem>	-7.24
pCAP	<chem>O=C1C=C(C[C@@H](C(NC)=O)NC(C)=O)C2=C(O1)C=C(OP(O)(O)=O)C=C2</chem>	-7.23
4-MUP	<chem>O=C1C=C(C)C2=C(O1)C=C(OP(O)(O)=O)C=C2</chem>	-7.15
GpCAPG	<chem>CC(NCC(N[C@H](C(NCC(NC)=O)=O)CC(C1=C(O2)C=C(OP(O)(O)=O)C=C1)=CC2=O)=O)=O</chem>	-7.14
GpYGG	<chem>CC(NCC(N[C@H](C(NCC(NC)=O)=O)CC1=CC=C(OP(O)(O)=O)C=C1)=O)=O</chem>	-7.05
pNPP	<chem>O=P(O)(O)OC1=CC=C([N+])([O-])=O)C=C1</chem>	-7.04
OBA-d1	<chem>O=C(O)C1=C(C=CS1)NC(C(O)=O)=O</chem>	-7.01
OBA	<chem>O=C(O)C1=CC=CC=C1NC(C(O)=O)=O</chem>	-6.97
OBA-d3	<chem>O=C(O)C1=C(SC2=C1CCCC2)NC(C(O)=O)=O</chem>	-6.95
OBA-d5	<chem>O=C(O)C1=C(SC2=C1CCOC2)NC(C(O)=O)=O</chem>	-6.86
OBA-d2	<chem>O=C(O)C1=C(SC=C1)NC(C(O)=O)=O</chem>	-6.84
OBA-d4	<chem>O=C(O)C1=C(SC2=C1CCNC2)NC(C(O)=O)=O</chem>	-6.65

### 3.8. Discussion

We built a fragment library consisting of compounds with several different reactive groups and assessed their effectiveness as an inhibitor of PTPs. This was supplemented by the analysis of their reactivity with small-molecule thiols. This showed the difference in active site reactivity of each PTP with the compounds, with CD45 significantly slower to react. We were also able to compare among chemotypes, showing that thiazoles generally tend to be the most potent. Intrinsic reactivity of molecules differed by ~3 fold between compounds with the same structure except with different warheads; here thiazoles were not significantly more reactive than other groups on average. This suggests that a thiazole ring in the scaffold may in general make the molecule more specific towards PTP active sites. We were also able to deduce structure-activity relationships, such as a negatively charged substituent next to the covalent warhead being preferred. A more detailed analysis of the kinetic results may further guide our efforts by extracting affinity and reactivity effects, as will be shown in the next chapter.

### 3.9. Methods

*Substrate catalysis assay of phosphatase inhibition.* The procedure in Chapter 2 was followed. Determination of  $IC_{50}$  values was done by fitting the plot of residual enzyme activity over compound concentrations with nonlinear regression to a dose-response curve with fixed top and bottom and variable Hill slope in GraphPad Prism. For pNPP assay with detergent, Tris buffer with 0.01 % Triton X-100 was used only during preincubation. For pNPP assay with orthosteric inhibitors, same assay was carried out for three groups each time, including an enzyme only group. For the vanadate group, Tris buffer was mixed with sodium orthovanadate stock for a final 5 mM and pH was checked and adjusted to 7.5 if needed. For drugs, TCS-401 or GS-493 was diluted to 10  $\mu$ M and 5  $\mu$ M respectively in Tris. Then these buffers were used during the preincubation step. Enzyme mixture was prepared in a higher 1.2  $\mu$ M, and 5  $\mu$ L aliquots were taken and 295  $\mu$ L pNPP solution was added at each timepoint, for a 60-fold dilution.

*Mass spectrometry experiments.* The procedure in Chapter 2 was followed. For intact protein MS experiments in pH 6, Bis-Tris buffer (50 mM, ionic strength 150 mM) was used instead of pH 7.5 Tris buffer. For the calculation of labeling proportion from  $MS^E$ , peptides identified by m/z were additionally filtered according to our criteria of having >10 % primary ion matches. Then, all peptide response signal that contains a particular Cys modified by the corresponding compound was divided by detected peptide signal that contains the same Cys; those modified by the compound or carbamidomethylated were counted, as well as unmodified Cys in some rare cases, but peptides that were assigned to a different compound modification were discarded.

*Small molecule thiol reactivity assay using absorbance.* All assays were performed in PBS containing 20 mM sodium phosphate and 150 mM NaCl at pH 7.4. All measurements were taken at 37 °C. 50  $\mu$ M DTNB and 200  $\mu$ M TCEP were incubated in PBS for 5 min at 37 °C to obtain TNB<sup>2-</sup>. Separately, 200  $\mu$ M TCEP was incubated without DTNB in sodium phosphate buffer for 5 min at 37 °C. For each compound, TNB<sup>2-</sup> solution (73.5  $\mu$ L) was treated with 200  $\mu$ M compound (1.5  $\mu$ L of 10 mM stock in DMSO) in a 384-well plate (n = 3). Additionally, TCEP solution without TNB<sup>2-</sup> was treated with 200  $\mu$ M compound to control for background absorbance (n = 3). A DMSO group was included to serve as a negative control. UV absorbance measurements were acquired every 7.5 min for 7 h at 412 nm on a microplate reader. For each compound, the background absorbance without TNB<sup>2-</sup> was subtracted from the absorbance with TNB<sup>2-</sup>. [I] and [TNB<sup>2-</sup>] over time were determined from absorbance data using simple ratio to absorbance of the DMSO group at t = 0, taken as 100  $\mu$ M TNB<sup>2-</sup>. The data were fit to a second-order reaction equation using nonlinear regression in GraphPad Prism.

*Small molecule thiol reactivity assay using NMR.* Deuterated PBS was made using D<sub>2</sub>O, 20 mM Na<sub>3</sub>PO<sub>4</sub>, 50 mM NaCl and adjusting the observed pH to 7.6 using DCl. This corresponds to pD of ~8. Into an NMR tube, 400  $\mu$ L deuterated PBS, 100  $\mu$ L *N*-acetyl-cysteine methyl ester solution (120 mM in DMSO-*d*<sub>6</sub>) and 20  $\mu$ L 1,4-dioxane internal standard was added. 30  $\mu$ L of each compound stock (0.1 M) was mixed with 50  $\mu$ L DMSO-*d*<sub>6</sub>, and this mixture was added to the tube immediately before measurement. <sup>1</sup>H NMR spectra were taken every 5 min for 30 min. The integrals of the vinyl peaks and the internal standard was used to calculate the compound concentration. Resulting concentration over time was fit to a second-order reaction equation using nonlinear regression in GraphPad Prism.

*Library construction.* Compounds were purchased from Sigma and Enamine unless their synthesis is described here. All bromoacetamides and vinylsulfonamides were synthesized. Chloroacetamides **5ac-ae**, **5ah-al**, **5aq-at**, **5bb-bd**, **5bf**, **5bj-bl**, **5bp-bq**, **5cc-ce**, **5cg**, **5da**, **5h-i** were synthesized (**4al**, **5al**, **4da**, **5da**, **5as** by Andrew Johns); **5ai** and **5aj** were later purchased. Acrylamides **3ad** and **3al** were synthesized.

*General procedure for vinylsulfonamide synthesis.* 1 mmol of amine precursor was dissolved in 5 mL THF on ice. 1 mmol TEA was added, then 1.2 mmol 2-chloroethanesulfonyl chloride was added. The mixture was monitored by TLC. (Typically 1 h) Then the mixture was filtered, and then solvent was removed on a rotary evaporator. Product was obtained after purification with flash column chromatography using DCM and methanol as eluents.

*General procedure for amide synthesis.* 2 mmol of amine precursor was dissolved in 5 mL DMF on ice. 2.2 mmol potassium carbonate was added, followed by acyl chloride reagent – bromoacetyl bromide for **4**, chloroacetyl chloride for **5**, and acryloyl chloride for **3**. Mixture was allowed to come to r.t. while being stirred. Typically, reaction was complete within 2 h. When reaction was complete, adding 10 mL 1 M HCl and extracting with 10 mL ethyl acetate twice, followed by washing the organic layer with 10 mL brine, drying with sodium sulfate and concentrating, was enough to afford a pure product. If not, flash column chromatography was used. For the synthesis of tagged probes other than **2at** and **5at**, see Chapter 5.

*Docking studies.* Entire process was done using Schrödinger suite. Protein structures of PTP1B (PDB: 1KAK, 2HNP, 1SUG, 3CWE, 1G7G and 1G7F) were preprocessed, H-bond optimized, minimized and then solvents and ions were removed. For 3CWE, missing loop structure was filled with Prime. For all structures, both the original and those with catalytic Cys modified to a thiolate were used. Grid generation was done either from ligands included in the structure, or by using the gridbox generated from 1KAK. The catalytic thiol was set as rotatable. Ligands were generated from their formulae with LigPrep. Decoys were obtained from Schrödinger (1KLigandDecoys\_Epik). Then Glide was executed with no constraints and 1 pose for ligand. The resulting scores were averaged for each ligand for all grids.

## Chapter 4: Inactivation kinetics of irreversible inhibitors

### 4.1. Background

Enzyme kinetics models for noncovalent inhibitors are well known and taught in classrooms, in which inhibitors are classified into three groups – competitive, noncompetitive and uncompetitive – for their inhibitory mechanisms,<sup>100</sup> and empirical data can be easily fit into equations that are well compatible with Michaelis-Menten scheme. Thus, a simple linear regression that works for Michaelis-Menten kinetics will also elegantly reveal what mechanism an inhibitor uses when data with varying inhibitor concentrations are plotted the same way. The double reciprocal plot of substrate concentration with initial reaction rate (velocity), known as Lineweaver-Burke plot, is often cited as a graphical representation of the three inhibitory mechanisms, and other plots such as Eadie-Hofstee or Hanes-Woolf will produce a similar result, although extracting rate constants should now be done using nonlinear regression onto Michaelis-Menten equation itself, for more reliable estimates.

Irreversible inhibitors are most different in that enzyme activity will not be restored once inhibitor is removed. The result is that to determine kinetic constants, measurements must be made at different timepoints. A two-step mechanism, similar to Michaelis-Menten kinetics, is commonly used to model covalent inhibitors that have some specificity (sometimes called “quiescent”<sup>101</sup>):



Here, I discuss several models surrounding this particular system of reactions, how different approximations came about, and what precautions should be taken with them. Then I show our inactivation assay data and various fitting methods, and lastly, using what we have

learned, I show that it is possible to extract specific kinetic parameters from library screening data from the previous chapter, even with just three timepoints.

## 4.2. Approximations

For the above system, again in a similar fashion to Michaelis-Menten kinetics, two different assumptions can be made to derive equations. If we assume rapid (or, more accurately, instantaneous) equilibrium we have, as first used by Kitz & Wilson:<sup>102</sup>

$$k_{on}[E][I] = k_{off}[E \cdot I]$$

Applying conservation of enzyme species,  $[E]_0 = [E] + [E \cdot I] + [EI]$ , we get

$$k_{on}([E]_0 - [EI])[I] = (k_{on}[I] + k_{off})[E \cdot I]$$

$$[E \cdot I] = \frac{k_{on}([E]_0 - [EI])[I]}{k_{on}[I] + k_{off}}$$

Where, with the dissociation equilibrium constant  $K_i = \frac{k_{off}}{k_{on}}$ ,

$$[E \cdot I] = \frac{([E]_0 - [EI])[I]}{[I] + K_i}$$

If we assume a steady state model where the bound enzyme species are constant, as first used by Malcolm & Radda,<sup>103</sup> we have

$$k_{on}[E][I] = (k_{off} + k_{inact})[E \cdot I]$$

And the rest is the same, except only that  $K_i$  is swapped out for  $K_I$ , which is an equilibrium constant for binding, or potency:  $K_I = \frac{k_{off} + k_{inact}}{k_{on}}$ .

So far, the equations are identical to those in Michaelis-Menten kinetics, but they diverge in the last step:



$$\frac{d[EI]}{dt} = k_{inact}[E \cdot I] = \frac{k_{inact}([E]_0 - [EI])[I]}{[I] + K_{I\ or\ i}}$$

Where due to the fact that neither enzyme nor inhibitor is regenerated, the resulting rate that is actually observed in experiments is not a constant, while in Michaelis-Menten kinetics it is (under pseudo-first conditions):  $v = \frac{d[P]}{dt} = V_{max} \frac{[S]}{K_d\ (or\ M) + [S]}$ , where  $V_{max} = k_{cat}[E]_0$ . Therefore, an “observed” or “apparent” value of a first-order rate constant for the inactivation of the enzyme must be determined in each inhibitor concentration, which can be done with  $\varepsilon = [E] + [E \cdot I]$ :

$$\frac{d[EI]}{dt} = -\frac{d\varepsilon}{dt} = \frac{k_{inact}\varepsilon[I]}{[I] + K_{I\ or\ i}}$$

Thus, following first-order derivation,

$$\varepsilon = \varepsilon_0 e^{-\frac{k_{inact}[I]}{[I] + K_{I\ or\ i}}t}$$

Now we can see that the observed rate constant, which is the term next to time: ideally, inhibitor should be in excess so that pseudo-first order kinetic is followed with the constant of

$$k_{obs} = \frac{k_{inact}[I]}{[I] + K_{I\ or\ i}}$$

We must note a few points that are overlooked in most works that utilize this relationship. First, the steady-state approximation also assumes rapid equilibrium. It is easy to see that  $K_I$  approaches  $K_i$  as  $k_{inact}/k_{off}$  approaches zero. Past works have noted that  $K_i$  is a special case of  $K_I$  when that additional assumption holds, and using  $K_I$  is a general case where no assumption is made on the relative scale of  $k_{inact}$  and  $k_{off}$ .

As  $K_i$  and  $K_I$  are essentially counterparts for  $K_d$  and  $K_M$  in the Michaelis-Menten model, we refer to it for conditions of each assumption. For rapid equilibrium to hold,  $\frac{k_{inact}}{k_{off}} \ll 1$  needs to hold. Steady-state approximation requires that  $\frac{[E]_0}{[I]_0 + K_I} \ll 1$ . In real-world terms, this means that

either inhibitor concentration is much higher than enzyme concentration, or enzyme concentration is far less than  $K_I$ . This ensures equilibrium in the first step of the model is quickly reached: either inhibitor quickly saturates enzyme occupancy, or the inhibitor has such a low affinity that equilibrium is reached despite relatively low inhibitor concentration. Most experiment conditions in inactivation studies meet the first requirement. When the covalent bond formation is slow,  $K_I$  approaches  $K_i$ , meaning the empirical constant approaches the dissociation equilibrium constant.  $K_I$  itself has significance in that it is the inhibitor concentration at which enzymes have 50 % occupancy. From this point, we will refer to  $K_I$  in all cases.

Second, the determination of the observed rate constant depends on the measurement of  $\epsilon$ , the sum of free enzyme and bound but not reacted enzyme concentrations. Kitz & Wilson explicitly state that their method involves extensive dilution of the enzyme and inhibitor aliquot into solution containing the substrate, meaning that the bound species E·I will also display activity in the subsequent substrate reaction measurement; similar procedure is implied in the derivation process and experimental methods of Malcolm & Radda.

In our case, the vast majority of experiments was done in the same fashion, where the aliquot was diluted 10-20 times. Whether this dilution is sufficiently large for the approximation that all bound species will come unbound may depend on  $k_{on}$  and  $k_{off}$ , and was assessed in our experiments. In addition, in continuous enzymatic assays, where inhibitor and substrate are mixed in together with the enzyme, bound species E·I most likely cannot participate in the catalysis of the substrate reaction. Thus, care must be taken when using a secondary plot of  $k_{obs}$  vs [I] obtained from the product formation curve to get to the kinetic constants. When assuming rapid equilibrium,

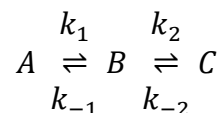
$-\frac{d[E]}{dt} = 0$ , as in an equilibrium, hence one cannot use this assumption when measuring [E]. Under steady-state approximation,  $-\frac{d[E]}{dt} = k_{inact}[E \cdot I] = \frac{k_{inact}[I][E]}{K_I}$ , so  $k_{obs}$  is linearly proportional to

[I]; in other words, it becomes a simple one-step bimolecular rate constant. This is inherent to the assumption that  $\frac{d[E \cdot I]}{dt} = 0$ , because then the consumption of free enzyme equals the formation of EI. In this case, only  $k_{\text{inact}}/K_I$  may be obtained. To get around this, one can fit [P] directly into the algebraic solution for [P], to get separate rate constants.<sup>104</sup>

With this information, we can pick out a few molecules that we may subject to more experiments. Through this, we will first obtain more reliable values of  $k_{\text{inact}}$  and  $K_I$  for those molecules from the additional data. Then we will discuss our efforts to extract those values for all compounds in our library, and how reliable those values are. This process will give us one more way to compare the candidate compounds for their reactivity and potency.

### 4.3. Analytical solution of the irreversible inhibition system

Finally, to assess the validity of these assumptions further, we can turn to another set of equations for the same model. Consider a series of two reversible unimolecular reactions:



Analytical solution for the system of ordinary differential equations in this chemical model has been shown many times before.<sup>105-108\*</sup> The concentration of A follows a double exponential decay:

$$[A] = [A]_0 \left( \left( \frac{k_1(a - k_2 - k_{-2})}{a(a - b)} \right) e^{-at} + \left( \frac{k_1(k_2 + k_{-2} - b)}{b(a - b)} \right) e^{-bt} \right) \quad (2)$$

---

\* Lente provides a more general solution for any starting concentrations of each species. Their expressions, however, slightly differ from the other works cited. There seems to be an error in Childs & Bardsley's  $b$ , or  $\lambda_3$ , due to them grouping out the term  $p$  out of square root, which, as Cornish-Bowden noted, results in  $\lambda_2$  and  $\lambda_3$  switching when the inequality between  $k_1 + k_{-1}$  and  $k_2$  is reversed.

where

$$a = \frac{1}{2}(p + q), b = \frac{1}{2}(p - q)$$

$$p = k_1 + k_{-1} + k_2 + k_{-2}, q = \sqrt{p^2 - 4(k_1k_2 + k_{-1}k_{-2} + k_1k_{-2})}$$

To apply this to irreversible inhibitor reaction (1), only two adaptations are needed. One is the pseudo-first order approximation of the first part of the reaction,  $k_1 = k_{on}[I]$ , which is a rather safe assumption, considering that inhibitor is in large excess in a typical assay, including our own<sup>†</sup>; another is irreversibility, meaning  $k_{-2} = 0$ . To our knowledge, there is no defined threshold of an equilibrium constant of the bond formation for such reaction to be called irreversible, but conventionally an inhibition is thought to be irreversible if the covalent bond is stable after thorough removal of inhibitor, for example through dialysis. As we are already working under the premise of irreversibility, this should not affect the quality of our model.

Then  $a$  and  $b$ , the terms in the exponents, which are the inverse of relaxation times, can be expressed by:

$$\frac{1}{\tau_1} = a = \frac{1}{2} \left( k_{on}[I] + k_{off} + k_{inact} + \sqrt{(k_{on}[I] + k_{off} + k_{inact})^2 - 4k_{on}k_{inact}[I]} \right)$$

$$\frac{1}{\tau_2} = b = \frac{1}{2} \left( k_{on}[I] + k_{off} + k_{inact} - \sqrt{(k_{on}[I] + k_{off} + k_{inact})^2 - 4k_{on}k_{inact}[I]} \right)$$

Most literature focuses on the relaxation times in this model. Cornish-Bowden suggested<sup>109</sup> that the apparent first-order kinetics that is observed by researchers is for the second relaxation time  $\tau_2$ , with the first relaxation time  $\tau_1$  being rather short. Childs & Bardsley proceeded to look at  $\varepsilon$ , and obtained the derivatives of  $\log(\varepsilon/E_0)$  over time, as researchers at the time were plotting some

---

<sup>†</sup> For a second-order solution with variable  $[I]$ , see <sup>106</sup>.

logarithmic form of activity over time, and use the slope obtained from linear regression as  $k_{\text{obs}}$ . Instead of following their arduous process of differentiation, we can refer to  $E_1 + E_2$  straight away:

$$[\varepsilon] = E_1 + E_2 = [E]_0 \left( \left( \frac{-k_1 k_2}{a(a-b)} \right) e^{-at} + \left( \frac{k_1 k_2}{b(a-b)} \right) e^{-bt} \right) \quad (3)$$

We note that the amplitude for the first decay is negative. Therefore, when  $\varepsilon$  is measured over time, at a certain time the curve passes an inflection point, after which it resembles a first-order decay. Next, we can assess when these features are noticeable. The amplitudes of the two exponents have a ratio of  $b$  to  $a$ . We know that as  $4k_{\text{on}}k_{\text{inact}}[I]$  approaches zero,  $a$  approaches  $p$  and  $b$  zero; when it approaches  $(k_{\text{on}}[I] + k_{\text{off}} + k_{\text{inact}})^2$ , both  $a$  and  $b$  converge to  $p/2$ . As  $k_{\text{on}}[I]$  is far larger than  $k_{\text{inact}}$  in most cases, which is the first case, we can see that the first exponent will be small in amplitude and fast in decay time, and so in those cases researchers will have no issue fitting  $k_{\text{obs}}$  to a single exponential decay without much regard for whether they are using  $[\varepsilon]/[E]_0$  or  $[E]/[E]_0$  (e.g. from continuous enzymatic assays).

We can see that when theoretical  $b$  is plotted against  $[I]$ , the curve resembles that of theoretical  $k_{\text{obs}}$  from steady-state approximation, and displays same the boundary conditions both when  $t = 0$  and  $t$  goes to infinity, although it is not the half-max height of  $K_I$ . As noted by Cornish-Bowden,  $b$  can be approximated into the steady-state equation with the assumption  $(k_{\text{on}}[I] + k_{\text{off}} + k_{\text{inact}})^2 \gg 4k_{\text{on}}k_{\text{inact}}[I]$ ; which corresponds to what we discussed above as a condition for the trajectory to be close to a single exponential decay. As in this case  $a \sim p$ ,  $b \approx \frac{k_{\text{on}}[I]k_{\text{inact}}}{k_{\text{on}}[I] + k_{\text{off}} + k_{\text{inact}}} = \frac{k_{\text{inact}}[I]}{[I] + K_I}$ . This approximation of relaxation time  $b$  into  $k_{\text{obs}}$  is reliable in all cases except when  $k_{\text{off}}$  is large compared to the other two constants, and considering that this is the same condition required for a good fit to single exponential decay, we can safely say that unless

one encounters conditions when it is necessary to use the exact algebraic solution, steady-state equation should provide a sufficiently accurate model.

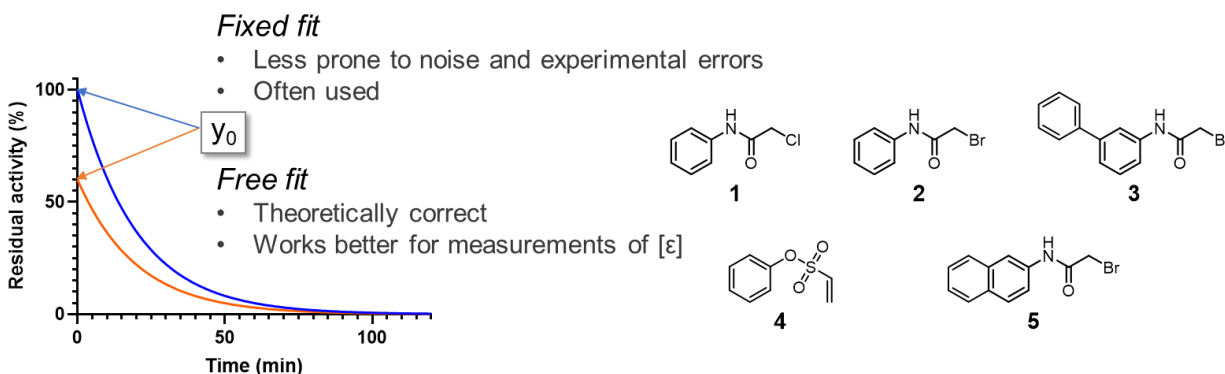
However, this only speaks to the relaxation time and not the amplitude of the single decay. Especially when  $[E]$  is being measured instead of  $[\varepsilon]$ , while same equation may be used,  $k_{\text{obs}}$  curve may have to start further down, given that the amplitude of the first decay is also positive. This is also easily understood in terms of chemistry: the scope of steady-state/rapid equilibrium assumptions begin after a certain point in time when the initial equilibrium is reached and  $[E \cdot I]$  is stable. Between true  $t = 0$  and this time, most of the change in system is in the proportion of E and E·I species, reflected in the first exponential term of the equations. Thus,  $[\varepsilon]$  will largely be constant during this time, but  $[E]$  will decrease rapidly to the equilibrium level. Thus, when one is measuring  $[E]$  instead of  $[\varepsilon]$ ,  $k_{\text{obs}}$  fitting may only work with the y-intercept unconstrained. Whether to tether regression to 100 % at  $t = 0$  is not explicitly mentioned in the works that first used the approximations, but at least Kitz & Wilson did not in their own analysis. This is also a natural consequence of doing a linear regression on some form of  $\log(\text{activity})$  – as it would intuitively only involve the measurement datapoints. On the other hand, when using nonlinear regression, many works in the literature describe fitting activity over time to exponential decay with a fixed amplitude at 1. We will discuss this in light of our own data, but given the theoretical background that we are fitting only for the second decay under the right assumptions, it seems reasonable to not fix the amplitude.

#### **4.4. Using different models to fit inactivation assay data**

Here, we compare several models for the fitting of our experimental data and discuss what conditions call for certain methods. We tested several enzymes such as YopH with several

compounds. While the concept of the experiment is the same as the pNPP reaction catalysis assay that we use throughout this thesis, the main difference is that aliquots are taken and measured much more frequently, every 15 minutes for 2 h. The highest concentration of the inhibitors tested also go above 1 mM that we use for other assays, up to 7 mM. This is to approach the plateau of the  $k_{\text{obs}}$  vs  $[I]$  curve, as much as other conditions of the experiment, such as solubility, permit.

Data was processed several different ways. First,  $k_{\text{obs}}$  was obtained from the plot of residual activity over time with nonlinear regression to a single exponent decay, both with and without the starting  $y$  (referred to as  $y_0$ ) set as 100 %. As discussed earlier, this is to account for the possibility that our measurement data is closer to  $[E]$ , in which case a free fit can be significantly better (Figure 4.1). Those values were plotted onto the secondary plot over time, and was fit to the steady-state approximation. Then, kinetic parameters were directly obtained from all plots of residual activity over time, by fitting to the algebraic equation for both  $[E]$  and  $[\varepsilon]$  globally across  $[I]$ . All nonlinear regression was done on GraphPad Prism. Fitting to algebraic solutions does not require special numerical solver software that support global data fitting across different  $[I]$ , and thus be much more accessible to researchers.

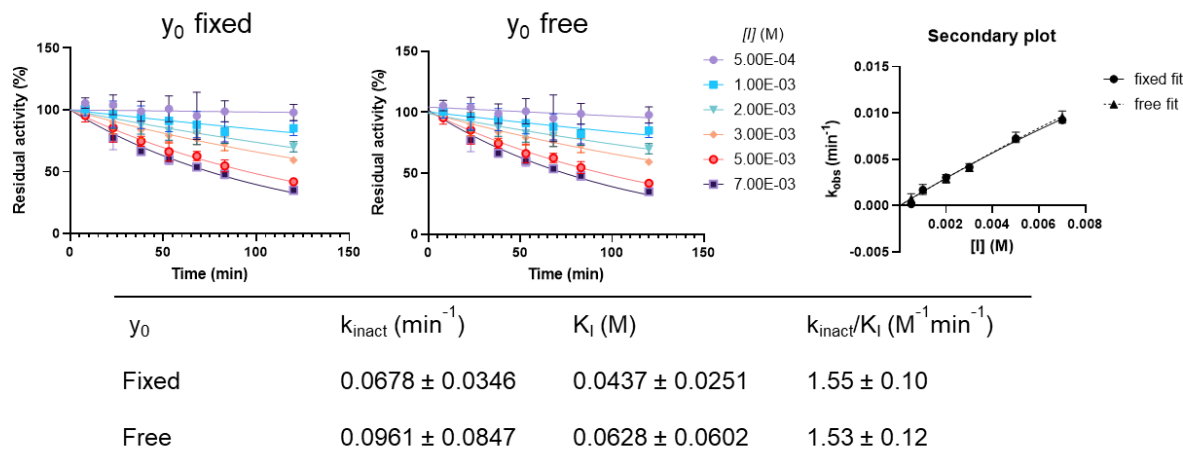


**Figure 4.1: Two ways of fitting for  $k_{\text{obs}}$ , and the assessed compounds.** Both curves are simulated for a  $k$  of 0.05, starting from 100 % for fixed and 60 % for free.

What we encountered is that it can be hard to find both values of  $K_I$  and  $k_{inact}$  because in some cases we could not get to  $[I]$  that is high enough for  $k_{obs}$  to start plateauing, due to limitations in either measurement intervals to capture the quick depletion of free enzyme, or solubility. As many researchers have pointed out, this happens when  $k_{inact}/K_I$  (covalent efficiency) is too high, for example for a highly potent inhibitor.<sup>99,110</sup> In this case, the compound behaves as a one-step inhibitor for all intents and purposes, and while we cannot have a reliable fit for individual parameters,  $k_{inact}/K_I$  is still obtained, as it only requires the linear part near the origin to be fitted. This allows us to compare compounds with their  $k_{inact}/K_I$ , giving us another metric besides  $IC_{50}$  at a given time to assess. This is especially beneficial in light of recent gradual shift away from using  $IC_{50}$  due to their dependence on time and enzyme concentration.<sup>99,111</sup>

In the case of phenyl chloroacetamide (**1**), a compound of relatively low reactivity, even the highest concentration tested did not inhibit PTP1B fully during the window of the experiment. The secondary plot showed some curvature to approximate  $K_I$  and  $k_{inact}$  individually. Comparing regressions for  $k_{obs}$  between free fit and  $y_0$  fixed at 100 %, we found that even without constraint, fitted  $y_0$  were very close to 100 %. (In some cases  $y_0$  was higher, e.g. 104 %, but this was still very close compared to other, poorer fits generated by the same method. When this method is used in practice to extract kinetic parameters,  $y_0$  should be confined to between 0 and 100 %, as those are the theoretical limits.) Therefore, both methods generated very similar values of  $k_{obs}$  (Figure 4.1).

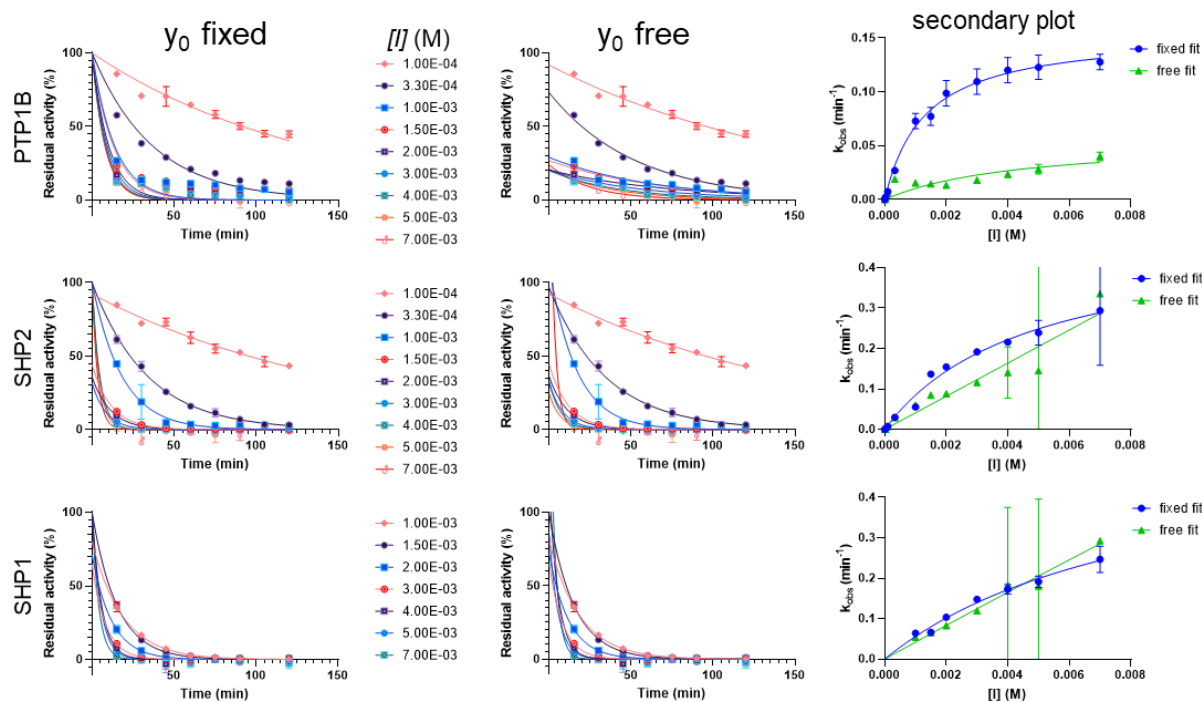




**Figure 4.2: Secondary plot method for PTP1B and phenyl chloroacetamide.**

For more reactive compounds, enzyme activity was quickly diminished in the earlier timepoints as well. Interestingly, when comparing enzymes for a fixed compound, we found that some enzymes show residual activity over time that visibly deviates from a single exponential decay from 100 %. In those cases, removing the constraint on the starting value improved the quality of fit significantly. The inhibition kinetics of phenyl bromoacetamide (**2**) for PTP1B showed this deviation, mostly due to higher [I] groups having lingering activity at later timepoints, whereas the same compound did not show as much deviation for SHP1 and SHP2 (Figure 4.3). We compared the fits of both cases with extra sum-of-squares F test. This test reports the p-value of null hypothesis that the simpler model is better. With the threshold of  $p = 0.05$ , for PTP1B, 11 out of 13 datasets of different [I] tested rejected the null hypothesis, meaning that freeing up  $y_0$  produced statistically significant improvements in fit for those. On the other hand, for SHP2, 6 out of 13 passed the test, and for SHP1, 4 out of 7 did. For comparison, with **1** and PTP1B, none of the datasets passed the tests, meaning that constrained fit produced fits of equal or better quality in all cases.

In general, we found that for higher [I] where residual activity starts off lower, free fit is preferred. SHP2 and SHP1 did not have  $k_{\text{obs}}$  values changing dramatically between two methods, and in the secondary plot, free fit values did not reach the plateau and so only  $k_{\text{inact}}/K_I$  was usable. For PTP1B,  $k_{\text{obs}}$  differed significantly, and this led to a huge difference in the reported  $k_{\text{inact}}$ . Letting  $y_0$  to be less than 100 % naturally results in a lower estimate for  $k_{\text{obs}}$ . However, better fit of activity over time into  $k_{\text{obs}}$  did not translate into having a better fit in the secondary plot of [I] vs  $k_{\text{obs}}$ ; this can be interpreted as experimental error being distributed between PTP1B, SHP2 and SHP1 all had lower R-squared values for free fit data in the secondary plot. This may be because if  $k_{\text{obs}}$  conforms more highly to the obtained activity data, any systematic error will carry over to the secondary plot. Along that line, we also found that the quality of data is even more important for regressions with the starting activity set free, as even slight errors can displace the theoretical curve much further than constrained fit does (Figure in methods section). However, when experimental data conformed closely to theoretical curve such as in the case of PTP1B with 3-biphenyl bromoacetamide (**3**), secondary plot fit was still better for the  $k_{\text{obs}}$  values obtained from free fit, although both cases had very high R-squared values of  $>0.98$  (Figure in methods section).



$y_0$ fixed	$k_{\text{inact}}$ ( $\text{min}^{-1}$ )	$K_I$ (mM)	$k_{\text{inact}}/K_I$ ( $\text{M}^{-1}\text{min}^{-1}$ )
PTP1B	$0.155 \pm 0.005$	$1.29 \pm 0.14$	$120. \pm 9$
SHP2	$0.483 \pm 0.083$	$4.71 \pm 1.50$	$102 \pm 16$
SHP1	$0.556 \pm 0.081$	$8.94 \pm 1.96$	$62.1 \pm 4.8$
$y_0$ free	$k_{\text{inact}}$ ( $\text{min}^{-1}$ )	$K_I$ (mM)	$k_{\text{inact}}/K_I$ ( $\text{M}^{-1}\text{min}^{-1}$ )
PTP1B	$0.0561 \pm 0.0127$	$4.37 \pm 1.88$	$12.8 \pm 2.8$
SHP2	<i>a</i>	<i>a</i>	$40.9$ <i>b</i>
SHP1	<i>a</i>	<i>a</i>	$42.0 \pm 18.0$

*a* Not determined  
*b* Error not determined

Figure 4.3: Secondary plot method for 3 PTPs and phenyl bromoacetamide.

To address compounds inhibiting almost all enzyme too soon in the experiment, which leaves fewer data points that are useful in the fitting process, we also used YopH, a PTP that is in general much slower to react with covalent probes. Four highly reactive compounds were selected. The results (Figure 4.4, Table 4.1) are also mixed, in that some compound-enzyme pairs deviate more from a simple exponential decay than others. Both **3** and PVSN (**4**) display this behavior,

and extra sum-of-squares F test showed a dramatic difference between those two compounds, where all 7 datasets for each compound passed the test, and phenyl bromoacetamide, for which none of the datasets passed the test; 2-naphthyl bromoacetamide (**5**) had 5 out of 7 sets pass. This also translated to the secondary plot, where only **3** and **4** had higher R-squared values for the free fit  $k_{obs}$  than for fixed fit  $k_{obs}$ .

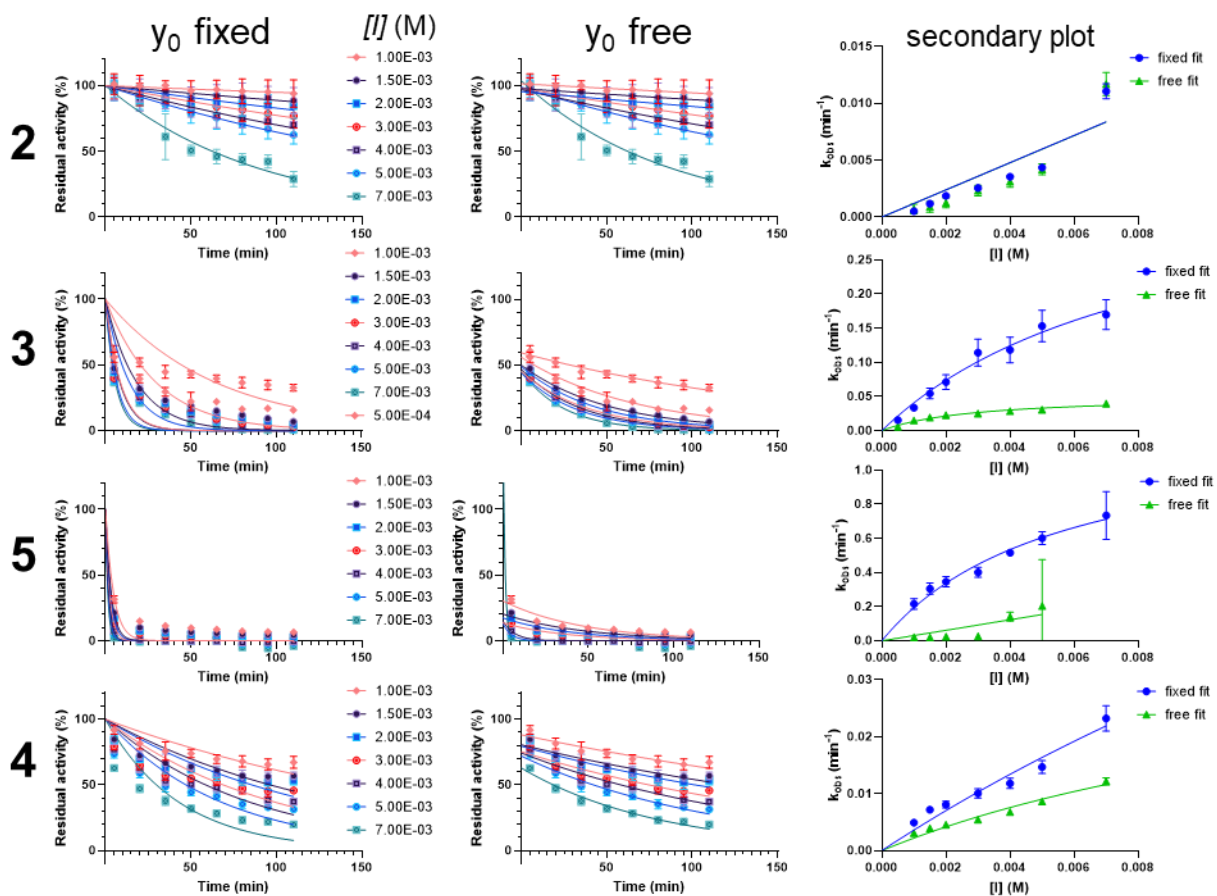


Figure 4.4: Secondary plot method for YopH with 4 compounds.

Because there was a large difference in  $k_{inact}$  obtained from two ways – for example, for **3** there is almost 8-fold difference – we also sought to compare those values to ones obtained from

directly fitting activity data into the algebraic solution for  $[\varepsilon]$  in equation (3) and  $[E]$  in equation (2) with constants shared globally across all datasets with different  $[I]$  (Figure 4.5, Table 4.1). We noticed two things: first, curves fitted to  $[\varepsilon]$  vs  $[E]$  roughly resemble curves fitted fixed and freely for  $k_{\text{obs}}$  respectively, as we expected. Second, with the parameters globally shared, the curves do not produce as good of a fit, in terms of individual R-squared, compared to the single exponential  $k_{\text{obs}}$  fits. While removing the shared constraint produces better fits for individual datasets, this fit method can only work with globally shared parameters. One notable case is when compounds have lower reactivity, such as with PTP1B with **1** or YopH with **2**: in those cases, both methods produced similar-looking fits, with very similar global R-squared values. But in other cases, generally data was fit better to  $[E]$ , and this may suggest that 10-fold dilution is not enough.

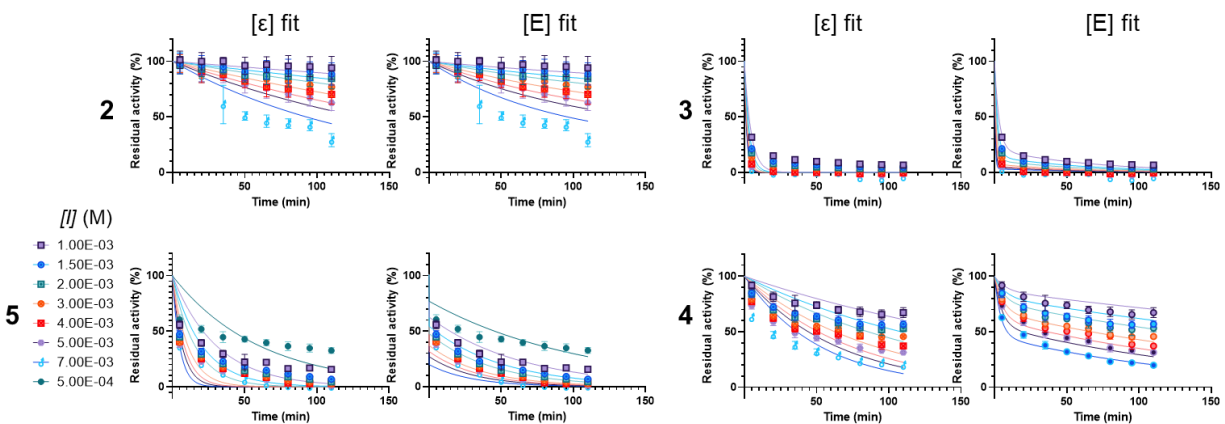


Figure 4.5: Global fit method for YopH with 4 compounds.

To summarize, in some cases where experimental data deviates more from a single decay, having  $k_{\text{obs}}$  fits without the constraint for the starting value produces noticeably better fits. However, that might also mean that the secondary plot has a worse fit for the steady state  $k_{\text{obs}}$  expression. While in theory there is no reason to constrain  $y_0$  except confining it to between 0 and

1, removing that also means the data is more prone to experimental errors. For less reactive compounds, even when  $y_0$  is not fixed the fit value tends to be very close to 100 %, and either method produces similar secondary plots and therefore  $k_{\text{inact}}$  and  $K_I$  values. For more reactive compounds, excellent experimental data is required to produce good fits that have, for each [I], decreasing  $y_0$  as [I] increases.

**Table 4.1: Kinetic parameters of 4 compounds with YopH as determined from different fit methods.** *N.D.* denotes not determined as secondary plot was in the linear regime.

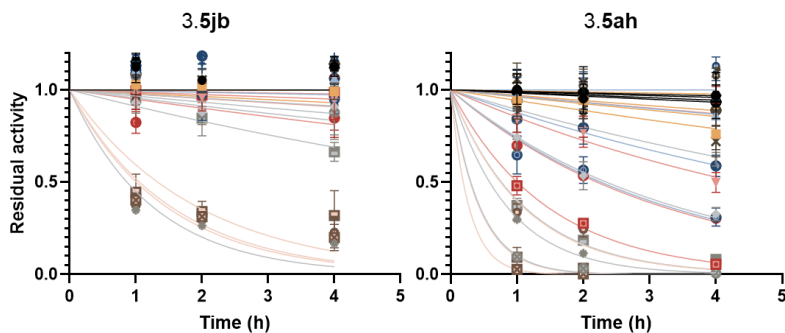
Compound	Global fit (blue shade: to $[\epsilon]$ , white: to $[E]$ )				Secondary fit (blue shade: fixed $y_0$ , white: free $y_0$ )	
	$k_{\text{on}}$ ( $\text{M}^{-1} \text{min}^{-1}$ )	$k_{\text{off}}$ ( $\text{min}^{-1}$ )	$k_{\text{inact}}$ ( $\text{min}^{-1}$ )	$K_I$ (M)	$k_{\text{inact}}$ ( $\text{min}^{-1}$ )	$K_I$ (M)
<b>2</b>	$1.22 \pm 0.17$	$0.00273 \pm 0.00286$	$4.34 \cdot 10^{-10}$	$0.00224 \pm 0.00625$	<i>N.D.</i>	<i>N.D.</i>
	1.40	235	777	725	<i>N.D.</i>	<i>N.D.</i>
<b>3</b>	$(1.81 \pm 0.05) \cdot 10^{15}$	$(3.18 \pm 0.20) \cdot 10^{12}$	$0.0427 \pm 0.0032$	$0.00176 \pm 0.00012$	$0.395 \pm 0.081$	$0.00873 \pm 0.00270$
	$2.27 \cdot 10^8$	$7.99 \cdot 10^6$	1.18	0.0352	$0.0519 \pm 0.0030$	$0.00285 \pm 0.00037$
<b>5</b>	$265 \pm 12$	$0.0610 \pm 0.0070$	$0.0183 \pm 0.0020$	$(2.99 \pm 0.23) \cdot 10^{-4}$	$1.28 \pm 0.18$	$0.00561 \pm 0.00135$
	$1.52 \cdot 10^{11}$	$5.21 \cdot 10^8$	$0.956 \pm 0.290$	$0.00342 \pm 0.00162$	<i>N.D.</i>	<i>N.D.</i>
<b>4</b>	$18.9 \pm 1.8$	$0.103 \pm 0.015$	$0.0139 \pm 0.0017$	$0.00618 \pm 0.00044$	$0.143 \pm 0.105$	$0.0388 \pm 0.0321$
	$6.79 \cdot 10^{10}$	$5.62 \cdot 10^8$	$0.0416 \pm 0.0092$	$0.00827 \pm 0.00251$	$0.0405 \pm 0.0122$	$0.0174 \pm 0.0067$

#### 4.5. Analyzing library screening data for kinetic parameters

Inhibition data for the library compounds in Chapter 3 was primarily obtained for the purposes of establishing  $\text{IC}_{50}$  values, and so measurement was done in only 3 timepoints. We sought to use this data to generate  $k_{\text{obs}}$  and subsequently  $k_{\text{inact}}$  and  $K_I$  values, and see how informative the numbers can be. For this purpose, we only used data for the chloroacetamides, and

filtered out datasets with no inhibition, defined by any pair that showed higher than 95 % residual activity at highest time and concentration, after which we were left with 54 to 57 groups; then we fit data into  $k_{obs}$  with  $y_0$  fixed, as with free  $y_0$ ,  $k_{obs}$  fluctuated especially for more reactive compounds, and the resulting secondary plot for  $k_{inact}$  and  $K_I$  as well. As inhibition rates for the analyzed compounds were relatively low, this should not cause too much deviation from true parameters, as discussed earlier.

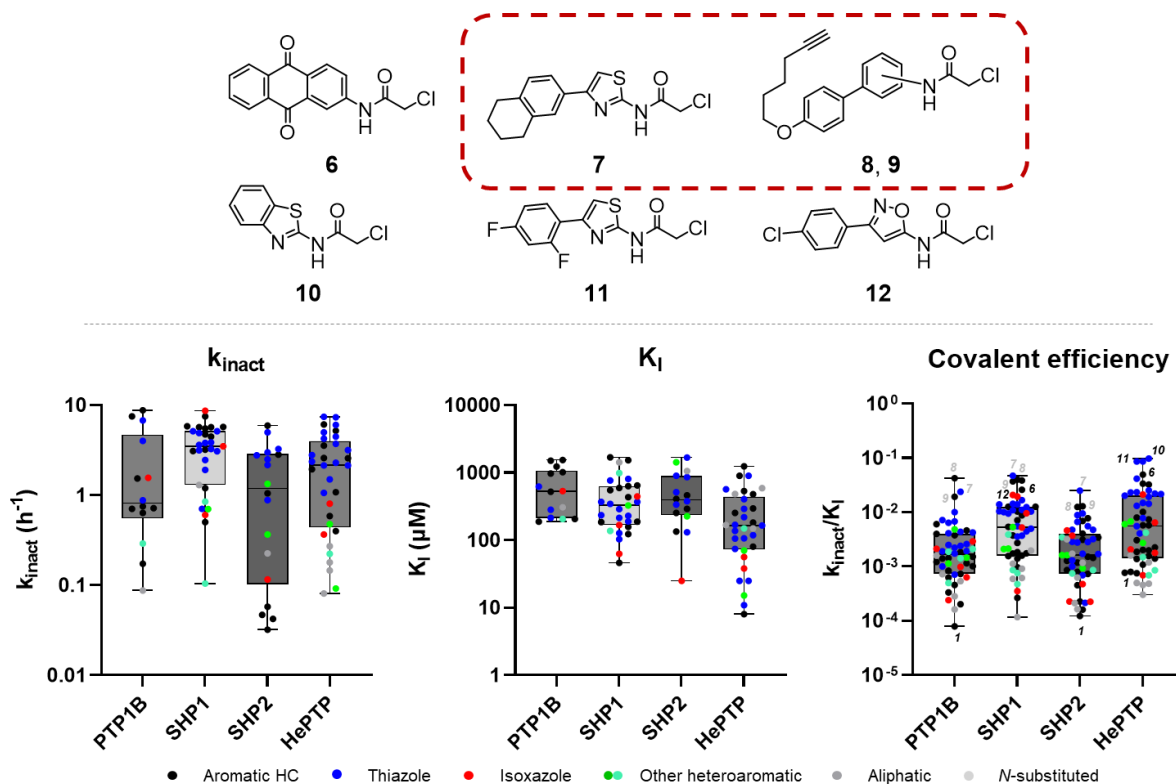
As with before, noisy data resulted in poorer  $k_{obs}$  fits, and there were cases where fixed fit did not produce a good fit (Figure 4.6). To only use more robust fits,  $k_{obs}$  fits reported as “unstable” by Prism were not used in the secondary plot. Secondary plots of compounds showed that most compounds had  $k_{obs}$  plateauing before 1 mM for HePTP, in contrast to PTP1B, where most compounds had  $k_{obs}$  linear up to  $[I]$  of 1 mM. While we can still obtain  $k_{inact}/K_I$  value from those linear plots, to analyze individual rate constants further, we only considered compounds with  $K_I$  calculated to be less than 2 mM, which is twice the highest concentration tested. After that, PTP1B had 13 compounds with determined  $K_I$  values, whereas HePTP had 35. Determined  $K_I$  values were in mid to high micromolar range except for a few, and HePTP had the lowest value on average (Table 4.1).



**Figure 4.6: Fitting library screen data to obtain  $k_{obs}$ .** Examples of compounds that do (right) and do not (left) closely follow a single exponential decay from 100 % are shown.

We identified a few interesting compounds from individual  $k_{\text{inact}}$  and  $K_{\text{I}}$  data (Figure 4.7). 2-(9,10-dioxo-9,10-dihydroanthracenyl) chloroacetamide (**6**), which stood out in analysis using  $\text{IC}_{50}$  and intrinsic reactivity, showed high affinity for SHP1 and HePTP. Aggregation-inducing compounds such as 2-(4-(5,6,7,8-tetrahydro-2-naphthyl)thiazolyl) chloroacetamide (**7**) and the hexynyl ether-tagged variant of 4- and 3-biphenyl chloroacetamide (**8, 9**) had misleading data here as well, displaying a high  $k_{\text{inact}}$  and a low  $K_{\text{I}}$ . A number of thiazole scaffolds performed well in this analysis as well, such as 2-benzothiazole chloroacetamide (**10**) or 2-(4-(2,4-difluorophenyl)thiazolyl) chloroacetamide (**11**), and a few isoxazoles such as 5-(3-(4-chlorophenyl)isoxazolyl) chloroacetamide (**12**). Phenyl chloroacetamide (**1**), which was analyzed in the previous section through inactivation assay with more timepoints and concentrations, had a comparable order of magnitude for the obtained  $k_{\text{inact}}$ , but had  $K_{\text{I}}$  that is about an order of magnitude higher. Notably, it placed at the bottom of the scale for covalent efficiency alongside its carboxy derivatives, suggesting that more complex structures may improve covalent efficiency even when  $\text{IC}_{50}$  might not show it directly. However, it must be noted that individual  $k_{\text{inact}}$  and  $K_{\text{I}}$  values are not determined for a considerable number of compounds tested, so the constants may range beyond what we obtained.

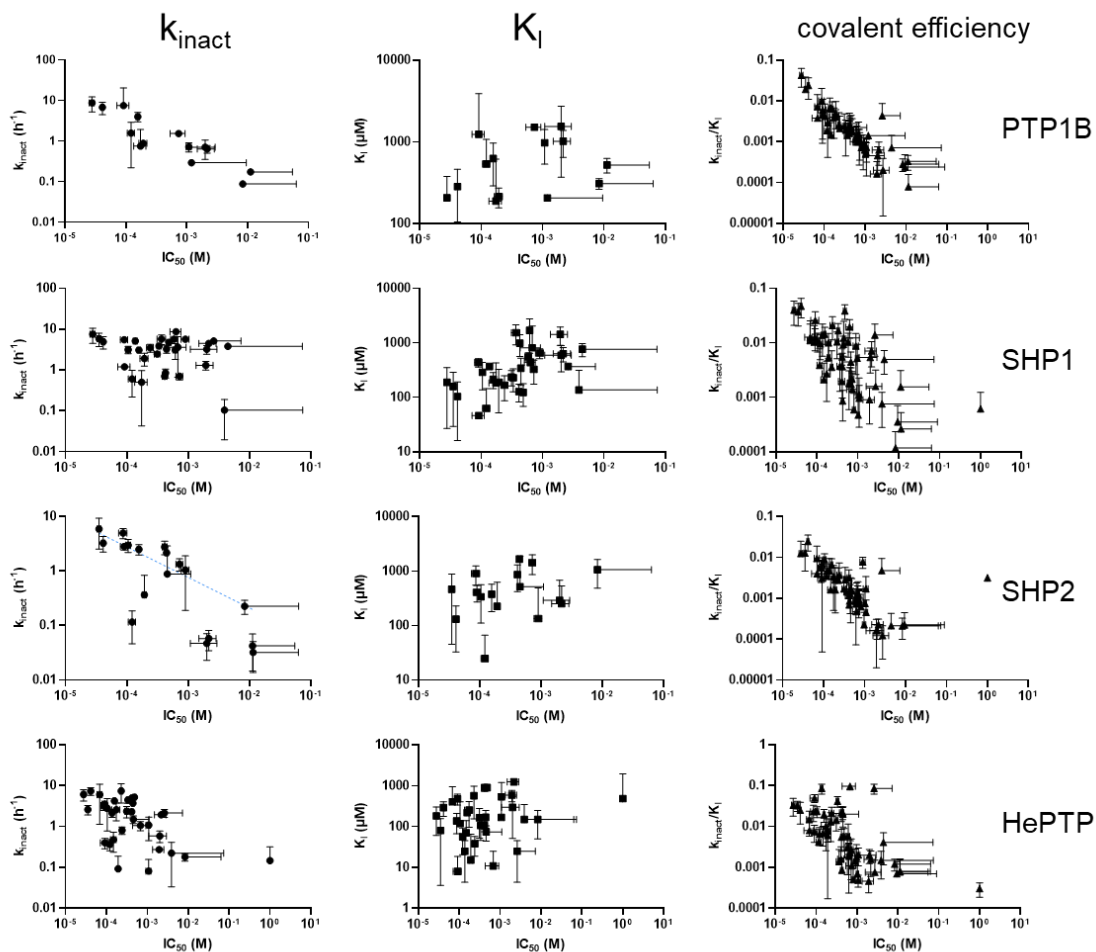




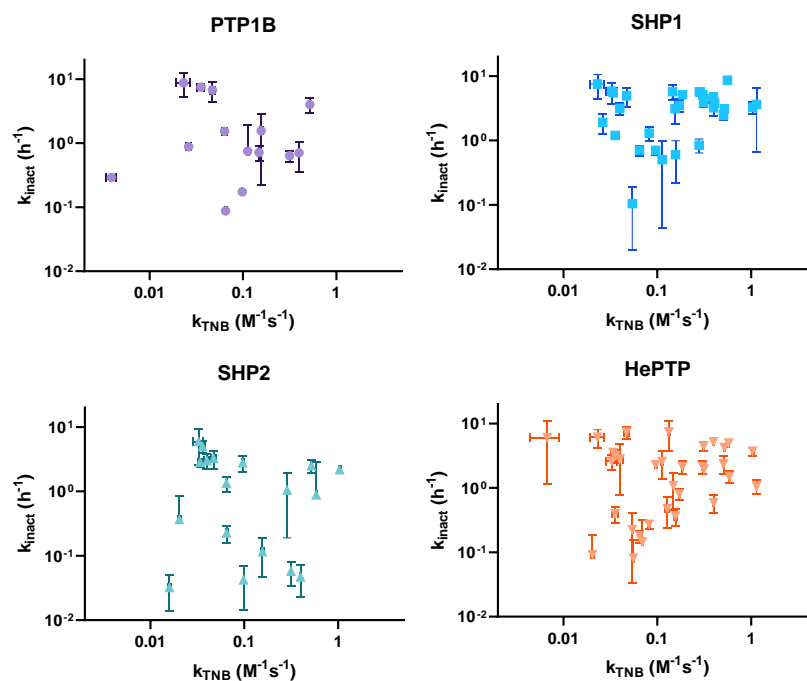
**Figure 4.7:**  $k_{\text{inact}}$ ,  $K_{\text{I}}$  and covalent efficiency for individual PTPs. For covalent efficiency, notable compounds are denoted with their numbers. Aggregating compounds are in the red rectangle.

Finally, we compared the covalent efficiency obtained from the fits, with  $\text{IC}_{50}$  values for the same compounds. Plotting both values in log scale confirmed that two values are largely correlated (Figure 4.8). This is expected, and researchers have reported using  $\text{IC}_{50}$  values to extract kinetic rate or equilibrium constants<sup>112–114</sup>. Interestingly, comparing  $\text{IC}_{50}$  with  $k_{\text{inact}}$  and  $K_{\text{I}}$  individually revealed differences between PTPs. For PTP1B, SHP2 and HePTP,  $K_{\text{I}}$  was less correlated to  $\text{IC}_{50}$  than  $k_{\text{inact}}$  was; but for SHP1,  $k_{\text{inact}}$  was rather flat, and  $K_{\text{I}}$  was the major factor driving covalent efficiency correlation. Then, we also compared  $k_{\text{inact}}$  values for each enzyme with the  $k_{\text{TNB}}$  obtained from intrinsic reactivity assay (Figure 4.8). While the two constants did not show a high correlation, many hexynyl ether-tagged probes, including those that cause aggregation, were

placed in the upper left corner, having a high  $k_{\text{inact}}$  for the same  $k_{\text{TNB}}$ . The range of  $k_{\text{inact}}$ , spanning about two orders of magnitude, was smaller than that of  $k_{\text{TNB}}$ . This may be explained by the fact that the covalent bond formation in the active site is different from the same reaction in general, as the active site contains structures and general acid/base residues to stabilize the transition state in a mechanism-based way.



**Figure 4.8: Obtained constants plotted against  $IC_{50}$  for each PTP. For SHP1, there was a statistically significant correlation between  $k_{\text{inact}}$  and  $IC_{50}$  (blue line).**



**Figure 4.9:**  $k_{\text{inact}}$  for each PTP compared to small molecule thiol reaction constant.

**Table 4.2:** All values of  $k_{\text{inact}}$ ,  $K_{\text{I}}$  and covalent efficiency for individual PTPs. Compounds are numbered in the same scheme as Chapter 3 (all are chloroacetamides, or 5). As explained, values that were filtered out are not shown.

	PTP1B			SHP1			SHP2			HePTP		
	$k_{\text{inact}}$ ( $\text{h}^{-1}$ )	$K_{\text{I}}$ ( $\mu\text{M}$ )	$k_{\text{inact}}/K_{\text{I}}$ ( $\text{M}^{-1}\text{h}^{-1}$ )	$k_{\text{inact}}$	$K_{\text{I}}$	$k_{\text{inact}}/K_{\text{I}}$	$k_{\text{inact}}$	$K_{\text{I}}$	$k_{\text{inact}}/K_{\text{I}}$	$k_{\text{inact}}$	$K_{\text{I}}$	$k_{\text{inact}}/K_{\text{I}}$
<b>aa</b>			202			1590			123			755
<b>ab</b>	1.53	1510	1010			1400			800			1090
<b>ac</b>			79.3			264						800
<b>ad</b>	0.174	522	333			1560						765
<b>ae</b>			834	5.7	650	8770	1.05	135	7790			1410
<b>af</b>	0.637	1020	627	4.48	632	7090	0.0578	254	228	1.95	1250	1560
<b>ag</b>	0.706	1550	456	3.21	594	5410	0.047	292	161	0.589	297	1980
<b>ah</b>			1150	5.69	552	10300			859			3070
<b>ai</b>	0.717	975	735			967			455	1.08	535	2020
<b>aj</b>			4640	5.51	436	12600	2.83	409	6910	3.56	480	7420
<b>ak</b>			1830	4.75	123	38600			1520	5.21	901	5780
<b>al</b>			1520			1740			1270			2250
<b>am</b>	7.51	1240	6070	1.2	46.3	25900			3120	0.399	8.05	49500

<b>an</b>			1410	3.11	1690	1840			734			5750
<b>ao</b>			1430	5.72	1530	3730			1450			1370
<b>ap</b>			1220			1760			1180			1880
<b>aq</b>	0.751	189	3980	0.503	185	2720			3450	2.58	254	10200
<b>ar</b>			3790			12700			4030			7990
<b>as</b>	8.76	207	42400	7.53	189	39900			12500	6.12	182	33600
<b>at</b>			19300	5.85	158	37000	5.95	464	12800	2.6	80.2	32400
<b>ba</b>			713	3.81	766	4970			215			4080
<b>bb</b>			4350	5.14	369	13900			4770	2.14	24.9	85900
<b>bc</b>			2590			3580			2790	7.43	568	13100
<b>bd</b>			1930	3.87	230	16800			2860	4.47	107	41800
<b>be</b>	4.01	626	6400	3.09	215	14400	2.5	378	6620	4.24	219	19400
<b>bf</b>	6.76	284	23800	4.91	104	47300	3.27	131	24900	7.36	291	25300
<b>bg</b>			6760	5.14	369	13900			4770	2.14	24.9	85900
<b>bh</b>			4230	3.15	289	10900	3	338	8860	2.79	119	23500
<b>bj</b>			1000	3.62	812	4460			1000	1.06	10.9	97400
<b>bk</b>	0.88	214	4110	1.91	190	10000			2840			7810
<b>bl</b>			2510	0.705	128	5520	2.77	864	3210	2.3	106	21700
<b>bm</b>			2400	2.46	235	10400			1710	2.36	166	14300
<b>bn</b>			1920			3710			692	4.99	903	5530
<b>bo</b>			7190			11000			9490	6.03	406	14900
<b>bp</b>			2230			12400	0.884	516	1710	1.5	74.2	20200
<b>bq</b>			10100			10300	4.99	902	5530	3.18	138	23000
<b>ca</b>			633			5240			226			1790
<b>cb</b>			242			354			224			692
<b>cc</b>			2120	3.5	168	20800			3700	0.808	38.2	21100
<b>cd</b>	1.56	536	2910	0.605	63.1	9590	0.116	25	4640	0.365	56.5	6460
<b>ce</b>			982	8.66	444	19500			477			2070
<b>da</b>			1140	0.703	327	2150	1.34	1430	934			2730
<b>db</b>			1460			2110			1620	0.478	70.6	6760
<b>dc</b>			4850			5350	0.365	226	1620	0.092	15.2	6030
<b>e</b>			2110	0.848	988	858			858			855
<b>f</b>			499			475			747			690
<b>g</b>	0.292	206	1420									
<b>h</b>			1870			3890			3480			4200
<b>i</b>				0.104	137	761				0.223	150	1490
<b>ja</b>						615			3180	0.146	484	301
<b>jb</b>			727			594			630			498
<b>jc</b>			1600			1960			1170			1590
<b>jd</b>			163	1.3	1430	912			164	0.272	592	459
<b>jg</b>			540			1130			1720	0.0812	168	484
<b>jh</b>	0.0873	309	283			118	0.226	1060	213	0.179	150	1200
<b>k</b>			1740	3.27	345	9470	2.17	1670	1300	3.73	170	22000

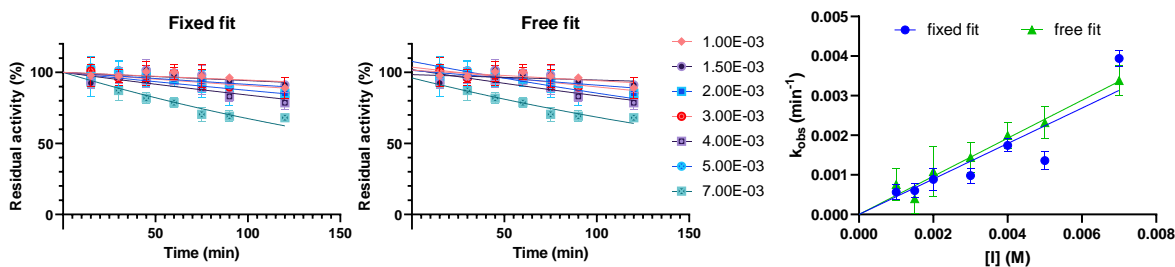
## 4.6. Discussion

First, we conducted inactivation assay with a few enzyme-compound pair, and fit the obtained data multiple ways. For the conventional method of fitting secondary plots with steady-state approximations, when fitting activity over time for  $k_{\text{obs}}$ , in many cases setting  $y_0$  fixed and free resulted in large differences in the obtained  $k_{\text{inact}}$  and  $K_I$  values, although  $k_{\text{inact}}/K_I$  was more conserved. Still, for compounds with lower reactivity, fixed and free fits produced very similar constants. With the global fit method, we saw that fits generated by fitting to  $[\varepsilon]$  is roughly similar to single decay fits with fixed  $y_0$ , and those generated by fitting to  $[E]$  similar to free fits to exponential decay. This suggested that 10-fold dilution from preincubation mixture to substrate reaction mixture may not be enough to dissociate bound but unreacted species. Lastly, we used our screening data and extracted  $k_{\text{inact}}$  and  $K_I$ , by using fixed  $y_0$  with just 3 timepoints. As expected,  $k_{\text{inact}}/K_I$  mostly correlated well with  $IC_{50}$  data. The result showed promise that  $k_{\text{inact}}$  and  $K_I$  may be modeled to get more information about affinity and specificity of the compounds, as they seemed to reveal similar information that we were only able to infer from having a second assay such as small molecule thiol reactivity besides  $IC_{50}$ . Overall, this analysis for kinetic rate constants can guide us further in identifying lead candidates that will be suitable for proteomics upon refinement.

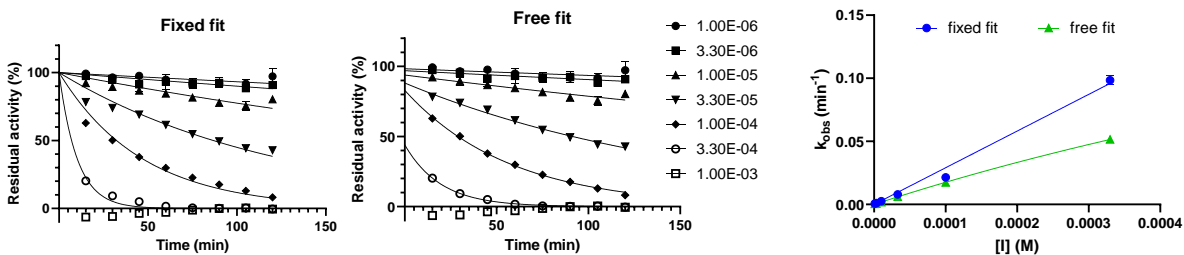
## 4.7. Methods

*Inactivation assay.* The procedure for pNPP assay in Chapter 2 was followed. Aliquots of preincubation mixture was taken every 15 min for 2 h. For high compound concentration range of between 1 mM and 7 mM, more 100 mM stock solution was used, resulting in a higher DMSO concentration of 7 % (instead of 1 %, v/v) in the preincubation mixture.

*Nonlinear regression and statistical analysis.* For secondary plot fitting, inhibition over time data was fit to a single exponential decay with plateau as 0 and top as 100 % or free between 0 % and 200 % (up to section 4.4) or 0 % and 100 % (section 4.5). Then the obtained  $k_{obs}$  values were plotted over time and was fit to a Michaelis-Menten kinetics equation. For global fitting, inhibition over time data was fit to a double exponential decay according to the equations 2 and 3.  $[I]$  was set as input for each data set, and all parameters were set to be shared across data sets. R-squared values were used as provided. Extra sum-of-squares F test was carried out with the  $y_0$ -constrained model as the simpler model. Correlation between fit  $k_{inact}$ ,  $K_I$ , covalent efficiency and  $IC_{50}$  data was analyzed using Pearson's  $r$ .



**Figure 4.10: Example of a noisy data fitting.** Data is of YopH and 2. Note that free fit translates into a better secondary fit.



**Figure 4.11: Example of a less noisy data fitting well without  $y_0$  constraint.** Data is of PTP1B and 3. At all concentrations, free fit had higher R-squared value.

## **Chapter 5: Utilizing active site-directed covalent probes in proteomics**

### **5.1. Background**

Genome sequencing has provided researchers with vast amount of data of scales never seen before. With the information, the next challenge was to understand the roles or functions of each protein encoded in the genome.<sup>115</sup> While genomic data itself can be used to infer protein function, the existence of post-translational modification that further diversities genome into the proteome means that there are clear limitations to using just the genomic data to obtain physiological contexts of the cell. Advent of proteomics through LC-MS/MS platforms, especially what is known as “bottom-up” approach where mixtures of proteins are subjected to enzymatic cleavage before analysis, gifted researchers with a reliable method to generate quantitative data with high throughput. However, one thing that is still missing in data produced by proteomic methods is the catalytic activity of the proteins, which may differ from the expression levels. As the activity is, in many cases, what ultimately dictates the role of each protein in the cell, additional methods designed to interrogate the activity of a subset of the proteins in their native state have been developed. One example is activity-based protein profiling, which has been fruitful and made into a standard toolset for researchers looking to characterize protein families for their function. Designing suitable covalent probes tends to be the main challenge that researchers need to overcome to reach success at profiling a particular family of proteins, which was the premise for this body of work.

Activity-based probes can broadly be categorized into directed and nondirected probes; directed probes are designed to target a specific family of enzymes. Ideally, this is done with probes of which the labeling reaction mechanism has a high degree of similarity with the catalytic

mechanism of the target. It can also be done by incorporated high-affinity binding groups that are specific to the target family. Either way, it requires prior knowledge about the target family and their structure, mechanism, or substrate specificity. Nondirected probes are mild electrophiles or photoreactive groups with some structural features that will react with a variety of proteins without a specific target. This strategy is useful when less-known enzyme families are being targeted.

As our goal was to design activity-based probes for a particular family, protein tyrosine phosphatases, we tested our probes from the fragment-like library with the activity-based protein profiling methods, including the LC-MS/MS based method. This required that probes are augmented with tags that can incorporate reporter groups. We found that even probes made from a fragment-like structure displayed selectivity towards a few target proteins.

## **5.2. Design and synthesis of tagged probes**

While it is possible to directly attach reporter groups to activity-based probes, having a bulky attachment may affect the probe functionally, affecting its ability to label targets, or cell permeability. Thus, a common approach has been to attach a small tag that can be reacted with a bio-orthogonal click chemistry. Attaching an alkyne or azide tag allows copper-catalyzed (or strain-promoted) cycloaddition with the reporter group that has the other tag. As for which tag to be attached to the probes, we chose alkyne groups for all our probes, as the reporter tag is added in excess during click chemistry, and due to higher reactivity of the activated alkyne, having a alkyne-reporter can increase background signal level.<sup>116</sup> It should be noted that terminal alkyne groups are reactive under certain circumstances. They can be oxidized to a Michael acceptor, which has been exploited in activity-based probing of several oxidases.<sup>117,118</sup> In fact, thiol-yne coupling methods, where a photoinitiator is used to activate a triple bond and induce reaction with

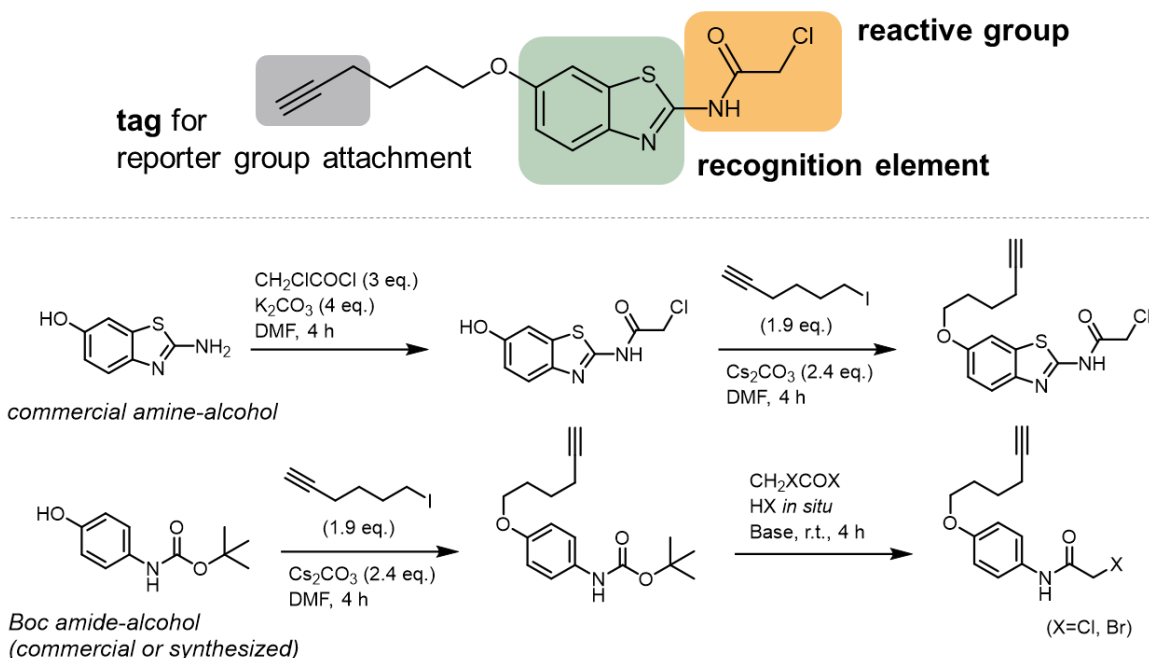


thiol groups via free radical mechanism, have been used for decades.<sup>119</sup> They can also function as a Michael acceptor towards select cysteine residues in pockets of enzymes such as ubiquitinases.<sup>120–122</sup> However, outside of these special contexts terminal alkyne groups are mostly biorthogonal. Even more electron-deficient alkynes, that have been used as a reactive group for thiols,<sup>123</sup> are relatively mild in reactivity and thus is featured in covalent drugs in clinical practice, such as acalabrutinib.<sup>124</sup>

Probes with general structures that could serve as a representation of a chemotype were chosen to be tagged; this included phenyl, biphenyl, naphthyl and a few thiazole moieties (Scheme 5.1). In most cases, an amine-alcohol precursor was used, to which alkyne tag was attached through Williamson ether formation with the hydroxy group, and the reactive group was formed from amine through the same reaction as untagged molecules. There were cases where an alkyne containing precursor was used (4-ethynylphenyl) or a different method of amide coupling with propargylamine was used from a carboxylic acid precursor to attach an alkyne group. In these cases, later a hexynyl equivalent was synthesized to compare all structures with the same hexynyl ether tag; however different tags did not lead to much difference (Figure in methods section).

To prevent the two steps of warhead formation and tag attachment from interfering with each other, in some cases Boc (*t*-butyl carbamate)-protection was used, either through commercially available precursors or through protection with Boc anhydride. Notably, in some cases we used a previously reported one-pot conversion method<sup>125</sup> from Boc to haloacetamides, using *in situ* generation of acid to cleave off Boc group.

## Activity-based probe



Scheme 5.1 ABP structure and representative synthesis pathways.

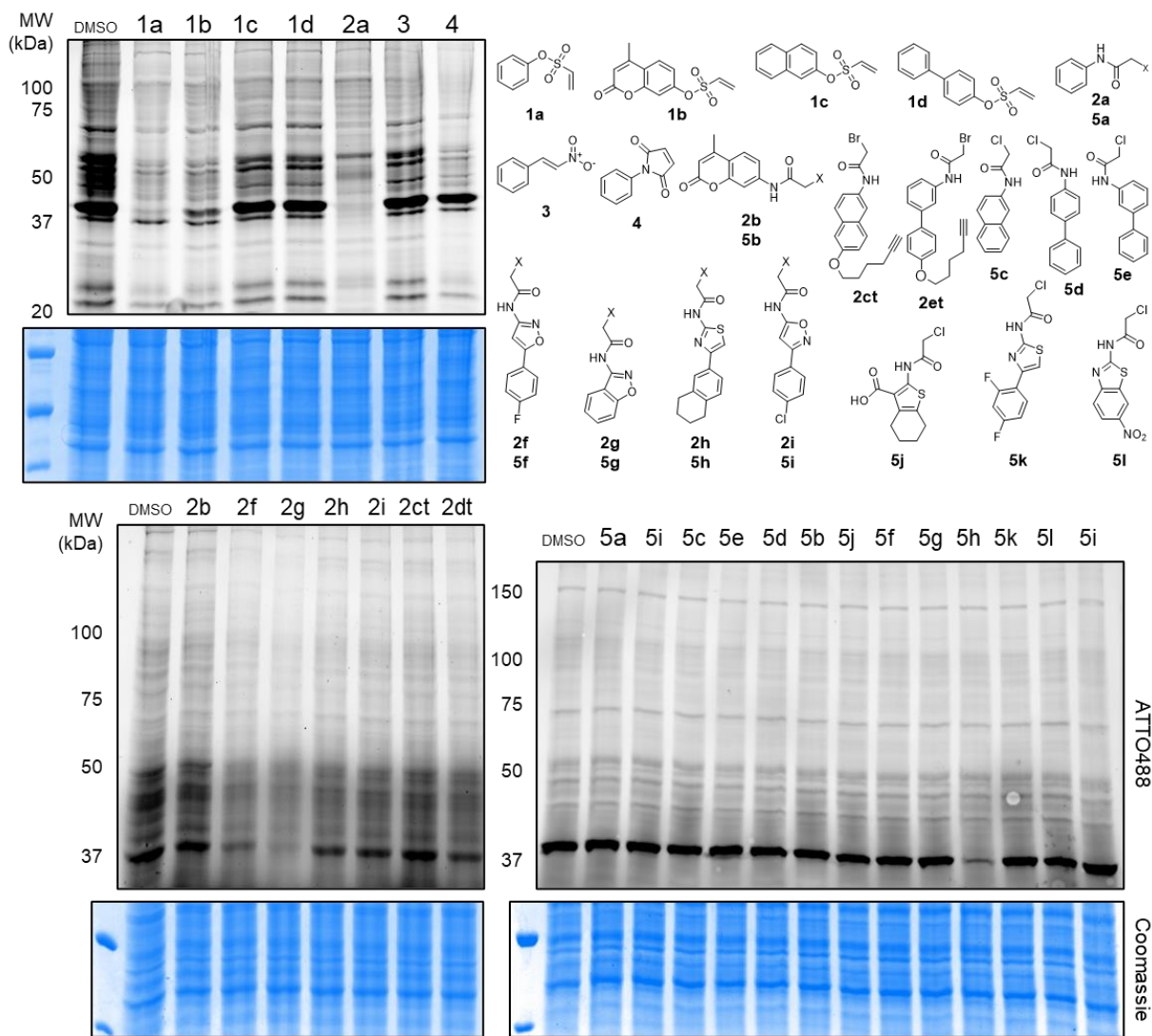
### 5.3. Gel-based ABPP methods

In gel-based ABPP approaches, proteome treated with probes are resolved through SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis), either one- or two-dimensional, and read out either through in-gel fluorescence or by blotting methods. It is a reliable technique, being based on a mature, standard methodology; while it does not reveal anything about each protein in the result besides molecular weight, downstream techniques such as in-gel digestion and MS/MS analysis can be used to determine the identity of protein.

Requirement of a tagged variant for any probe can be a barrier to assessing as many compounds against the proteome as possible; synthesis can consume both time and resource, and for some scaffold structures a path to a tagged probe may not be as straightforward either. While

a variety of biochemical assays we laid out in the previous chapters can be good substitutes, testing against the whole proteome is valuable in assessing whether a compound can be a good activity-based probe when tagged. One way to do so is by using a fluorophore broadly reactive towards thiol, to probe proteins that were not already labeled by the preceding probe treatment. In a way, this is similar to competitive ABPP<sup>126</sup> but without the narrowed scope for only the target sites. Therefore, while it may not be suitable for finding a probe with great affinity for the target active sites, it can still give information about the reactivity for off-target cysteines across the proteome. We used this method, with an iodoacetamide-ATTO488 dye conjugate, by treating HEK293 lysate samples with different candidate compounds and following up with the dye conjugate, then resolving with SDS-PAGE and measuring in-gel fluorescence of the dye.

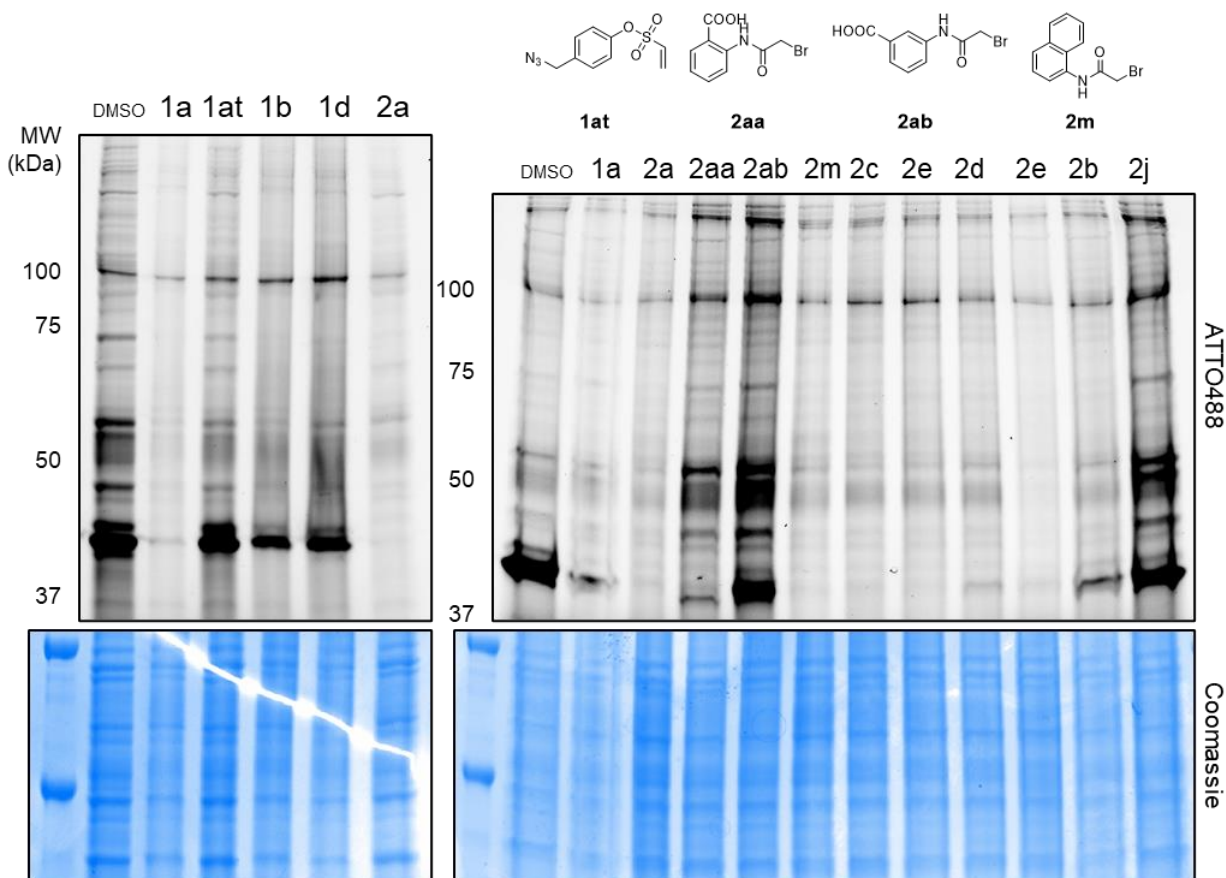
First, we assessed some thiol-reactive groups, which showed some difference in band patterns as well as intensity across the lane (Figure 5.1). Phenyl maleimide (**4**) was the most reactive except towards one band around 40 kDa. PVSN (**1a**), MUVSN (**1b**) and phenyl bromoacetamide (**2**) showed high reactivity across the proteome, while naphthyl and biphenyl vinylsulfonates (**1c-d**) seemed to have a lower reactivity for cysteines in the proteome. We also assessed bromoacetamides with different scaffold structures, which showed that bromoacetamide on methylcoumarin scaffold (**2b**) was not as reactive as on naphthyl and biphenyl (**2c-d**), both of which are tagged. It also showed that several heterocyclic compounds had more reactivity, a finding that is in line with the biochemical assays. For chloroacetamides (**5**), same experiment did not generate much information as all groups showed nearly identical results, probably due to the low reactivities of the warhead; however, for the bright band at 40 kDa, 3-(1,2-benzisoxazolyl) chloroacetamide (**5h**) showed a significant reduction in signal, suggesting that the compound had some specificity for that protein.



**Figure 5.1: Chasing with a broadly reactive dye conjugate.** Compounds were treated with 40  $\mu\text{M}$  for 1 h, then with 1  $\mu\text{M}$  IA-dye for 1 h. Lower signal in a band means that compounds reacted more.

Same experiments were also done by first treating HEK293 cells with probe-containing media, and then lysing the cells and treating with the same dye conjugate. Samples were resolved SDS-PAGE with and in-gel fluorescence was measured (Figure 5.2). This largely showed the same result as lysate treatment, showing that the probes tested were capable of permeating cell

membranes. Intensity differences between lanes revealed that azide- (**1at**) and carboxylate-containing compounds (**2aa-ab**, **2j**) are less cell-permeable than others.

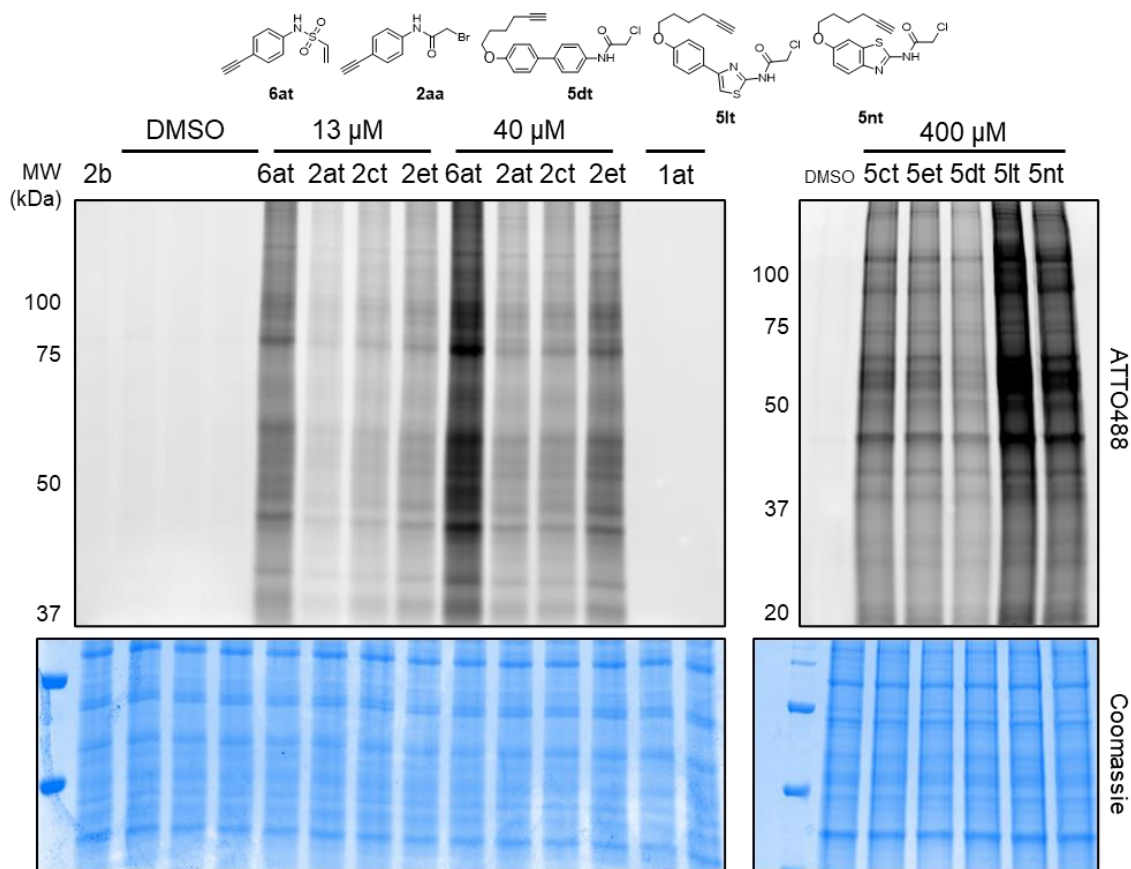


**Figure 5.2: Dye chasing after direct cell treatment.** Cells were incubated in media containing 100  $\mu$ M compounds for 1 h, then lysate was treated with 1  $\mu$ M IA-dye for 1 h.

Next, we used the tagged probes to check for labeling on the proteome. We used both azide-ATTO633 conjugate for fluorescence and azide-PEG-biotin for streptavidin blotting or enrichment. First, HEK293 cell lysates were treated with tagged compounds **2ct**, **2et**, **1at**, **6at**, **5ct-5lt** and subsequently CuAAC click chemistry was performed, either with azide-dye conjugate or azide-PEG-biotin. This showed that all tested compounds exhibit broad reactivity across the

proteome, with very modest differences in band patterns (Figure 5.3). Click chemistry steps were verified by removing each component upon which signal was extinguished (Figure in methods section).

Then, we also treated HEK293 cells directly with the compounds, and then cells were lysed to be subjected to the same gel-based ABP procedure. The result again showed that all compounds tested were able to permeate the cell membrane and label the proteins inside, indicating an improvement over previously reported probes such as  $\alpha$ -bromobenzylphosphonate. Many of the gel-based experiments were run in parallel with MS-based proteomics experiments, taking samples from each group for blotting with streptavidin-dye conjugate (see Figure 5.10).



**Figure 5.3: In-gel ABPP using synthesized probes.** Compounds were incubated with lysates at indicated concentrations for 30 min. **1at** did not display any signal as it is an azide probe.

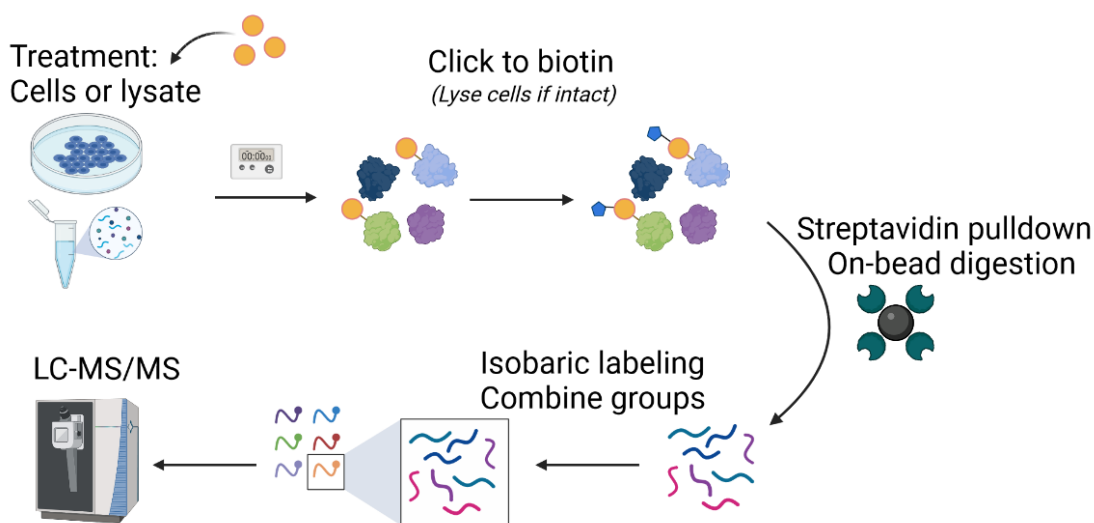
#### 5.4. Mass spectrometry-based ABPP methods

To assess which proteins in the proteome are being targeted with our probes further, we used LC-MS/MS based proteomics, which is more sensitive and qualitative than a gel-based method (Scheme 5.2). This part of work was done in collaboration with Jovanovic Lab (Biological Sciences, Columbia University), with Sarah Xi. Several quantification methods exist for MS proteomics; absolute quantification can only be done with an internal standard spiked into the sample, which can be costly. Therefore it is desirable to have a reliable relative quantification method in between the groups, to accurately compare the proteomic data for different states, in this case treatment with different compounds. MS1-based label-free quantification is an intuitive way that integrates the MS1 signal for a given peptide. While this can be used to compare signals from an unlimited number of different runs without any additional steps and cost, there are limitations, namely low precision, variability in signal and detection, and requirement for replicates.<sup>127</sup> Isotopic methods allow for relative MS1 quantification within the same run by combining samples labeled isotopically, which gives much higher reproducibility. Isotopes can be introduced multiple ways: a metabolic way of adding heavy amino acids to culture media, known as SILAC (stable isotope labeling with amino acids in cell cultures), has seen widespread use; *in vitro* methods such as the use of light and heavy cleavable linkers are commonly used for ABPP.<sup>128</sup> However, as this increases the complexity of the MS1 spectrum, usually only 2 samples are compared in each experiment (used for each group against a control group), and at most 3.

Isobaric mass tags have provided solutions for further multiplexing, where each tag has the same mass and only differs by the distribution of isotopes in the tag. This allows MS1 spectra for identical peptides to be a single peak, and the groups are quantified according to their

fragmentation in MS2. Commonly used, commercially available isobaric tags are mostly amine-reactive, and include iTRAQ (isobaric tags for relative and absolute quantitation) and TMT (tandem mass tags).<sup>127</sup>

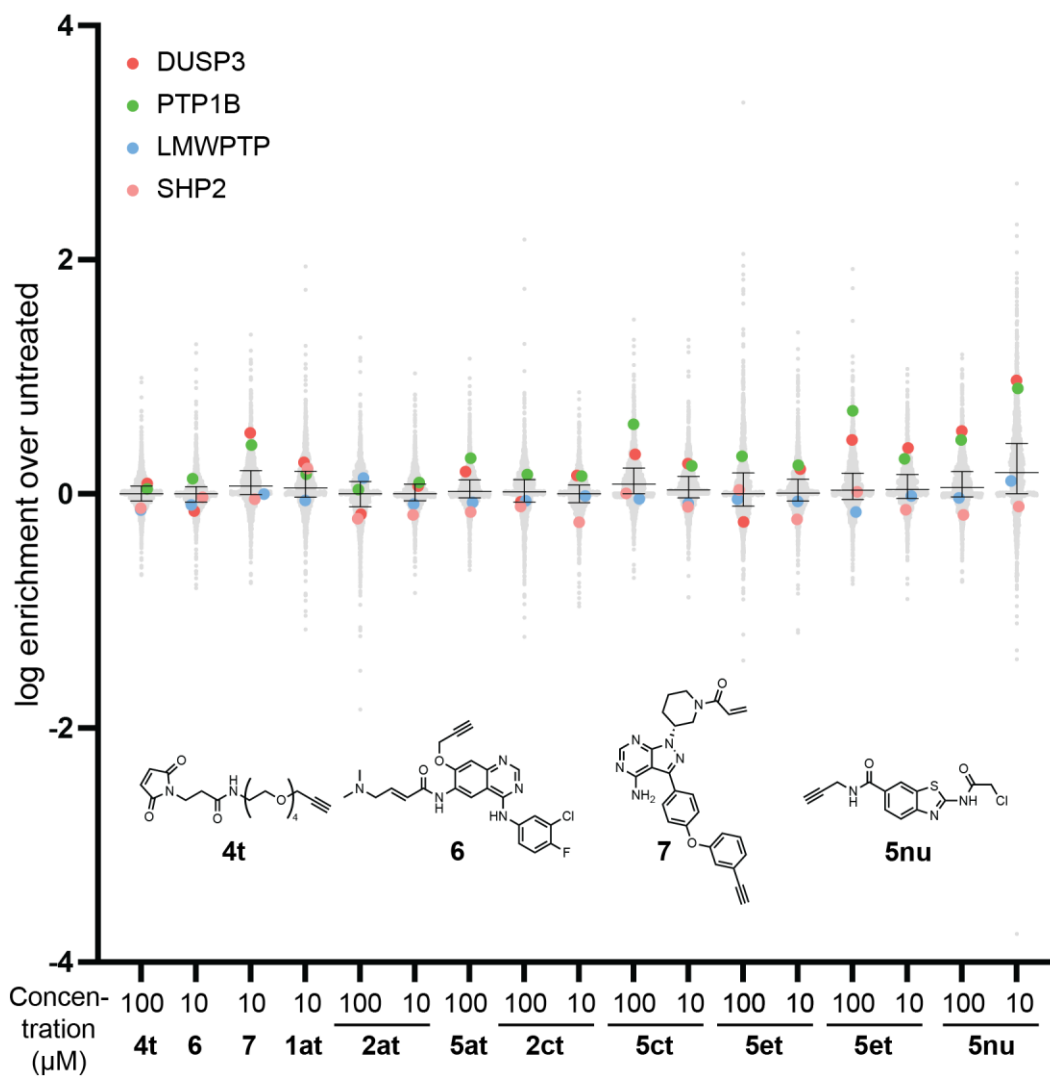
For the experiments described here, groups were multiplexed using TMT-18 for isobaric labeling and MS2 quantification between groups, and label-free MS1 quantification of iBAQ (intensity based absolute quantification) was used to normalize for each protein group. Data-dependent acquisition (DDA) was used. As the set of compounds to test, we chose our tagged chloroacetamide (**5at**, **5ct**, **5et**, **5nu**) and bromoacetamide (**2at**, **2ct**, **2et**) probes on top of azPVSN (**1at**, used with alkyne-biotin) to compare the performance to. Two drug molecules, alkyne derivatives of ibrutinib and afatinib – kinases inhibitors that targets Cys residues of BTK (Bruton's tyrosine kinase) and EGFR (epidermal growth factor receptor) with acrylamide (**7**) and 4-dimethylaminoacrylamide (**6**) groups, respectively – were added to assess the enrichment patterns of target-specific drugs. A commercially available kinase ABP, ActivX, as well as two controls, one untreated DMSO group and a broad thiol-reacting maleimide group (**4t**), were added.



Scheme 5.2: Workflow for MS proteomics-ABPP.



Two human cell lines were chosen for this part: adherent HEK293 cells and suspension Jurkat T cells. As with before, this works both by treating our probes on lysates, or directly treating the cells and then lysing. Then lysates were treated with alkyne-tagged probes, which were then clicked to azide-PEG-biotin. This allowed labeled proteins to be enriched for with streptavidin magnetic beads, followed by on-bead digestion with trypsin. The enrichment protocol and its parameters, such as incubation conditions and washing, had been optimized for efficient washing (Figure in methods section). Finally, different groups, including a DMSO-treated control group, were labeled with isobaric labeling reagents and combined to be measured through LC-MS/MS (Scheme 5.2). Each group was compared to the untreated DMSO group for the difference in protein signals. The result showed two PTPs, PTP1B and DUSP3, as enriched in many groups; both were enriched within the top 10 % of all proteins in the groups treated with **5et**, **5nu** and **7**. DUSP3 was enriched by **5at** and **5ct** within top 10 % as well. Plus, PTP1B was among the top 5 % of **5nu** and **7**, and DUSP3 was in the top 2 % for **5et** and top 5 % for **5ct**. Importantly, the corresponding bromoacetamides **2ct** and **2et** did not show as high enrichment, suggesting that the reduced reactivity was important in bringing out the ability of scaffolds to function as a recognition element. More reactive **1at**, **2at** and **4t** did not show significant enrichment of any PTP. While the two drug-alkyne molecules did not show enrichment for their targets as they were not detected in the lysates, **7** enriching PTP1B and DUSP3 may show that a less reactive warhead is more likely to enrich specific targets.



**Figure 5.4: Enrichment of proteins from intact cells treated with probes.** Intact HEK2993 cells were incubated in media containing compounds of indicated concentrations for 1 h. After quantification of each channel with MS2 labels, signal intensity for each protein group was compared to that of the DMSO control group. All 4 detected PTPs are colored.

## 5.5. Discussion

Here, we showed several methods of assessing the reactivity of candidate compounds for the proteome. Untagged molecules were tested through a following reaction with a broadly reactive dye conjugate. Then, probes with alkyne groups were synthesized, and used in both gel-based mass spectrometry-based ABPP. Probes did not have any issue with cell permeability.

As our goal is to develop directed probes for a particular family of phosphatases, the selectivity of the probes needs to be improved further. With electrophilic probes that do not have reaction mechanism identical to that of the natural substrate, side reactions can only be avoided by increasing the affinity, thereby reducing the dose of probes treated. We showed that our probes with simple structures are capable of enriching for PTPs, and that lower reactivity is essential to do so. With a more complex structure, probes may be tuned to have a higher selectivity for PTPs.

## 5.6. Methods

*Boc protection of amino naphthol.* 10 mmol of 6-amino-1-naphthol was dissolved in 50 mL DMF. 5.56 mL TEA was added. To this mixture, 20 mmol Boc anhydride (4.59 mL after liquifying at 37 °C) was added, and the mixture was stirred at 45 °C while being monitored with TLC. Upon complete protection to *N-bis*-Boc amide naphthol, DMF was removed on rotary evaporator. 100 mL MeOH and 30 mmol potassium carbonate was added. This was refluxed while being stirred and monitored with TLC. Upon conversion to *N-mono*-Boc amide naphthol, 50 mL 0.5 M HCl and 50 mL EtOAc was added, and aqueous phase was extracted twice more with 25 mL EtOAc. Organic layers were combined and washed with 25 mL brine, dried with sodium sulfate and concentrated. Product was purified with flash column chromatography.

*Synthesis of chloroacetamides with a hydroxy group.* **5nt** and **5lt** followed this method, after which ether formation gave final probes. To a stirred solution of 2-amino-6-hydroxybenzothiazole 4 mmol and potassium carbonate 16 mmol in DMF (10 mL) on ice, 12 mmol chloroacetyl chloride was added. The reaction was brought to r.t. as it was run for 4 h. Work up was done by adding 20 mL 1 M sodium bicarbonate solution and 20 mL EtOAc, and washing organic layer with 20 mL 1

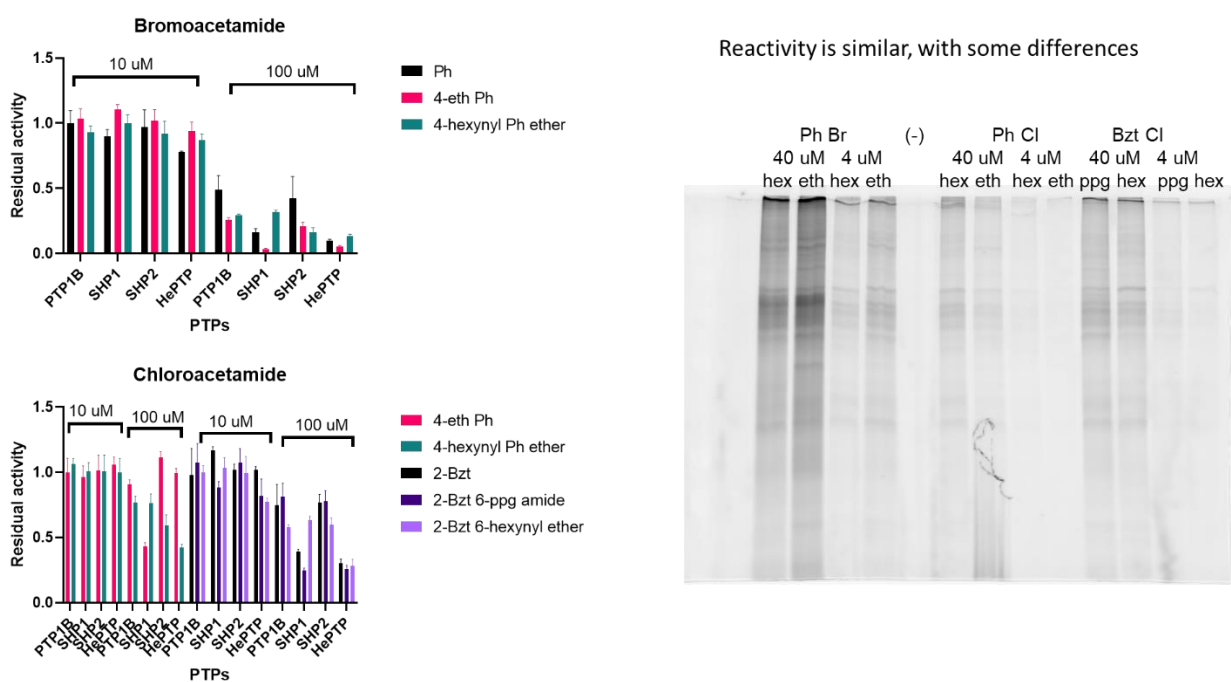
M HCl and then 20 mL brine. Flash column chromatography in 0 % to 50 % MeOH in DCM afforded 2-chloroacetamido-6-hydroxybenzothiazole. 4-(2-aminothiazol-4-yl)phenol was dissolved in DCM (10 mL) and 8.2 mmol potassium carbonate was added on ice. 4.2 mmol chloroacetyl chloride was added and the reaction was brought to r.t. as it was run for 1 h. After concentration, the mixture was redissolved in 10 mL ACN. Adding 10 mL of water gradually, 4-(2-chloroacetamidothiazol-4-yl)phenol was precipitated out; it was filtered and used.

*General procedure for hexynyl ether formation.* In 14 mL DMF, 2 mmol of alcohol molecule, either *N*-Boc amides (Ambeed) or chloroacetamides, and 4.8 mmol cesium carbonate was added. 3.8 mmol of 6-iodohexyne was added, and the mixture was stirred at r.t. while being monitored with TLC. Typical reaction lasted for 1 h. Product was extracted by adding 20 mL EtOAc and washing with 20 mL dI water 3 times, then drying organic layer with sodium sulfate and concentrating. Product was purified on flash column chromatography using 0 % to 25 % EtOAc in hexane.

*General procedure for conversion of Boc amides to haloacetamides.* **2ct-et**, **5ct-dt** and **5nu** followed this method. 1 mmol of *t*-butyl carbamates was dissolved in 10 mL solvent (DCM for **2**, ACN for **5**). For **2**, 4 mmol bromoacetyl bromide was added, and for **5**, 4 mmol chloroacetyl chloride, then 2 mmol of NaI was added. For both, then 80  $\mu$ L of MeOH was added and the mixture stirred at r.t. for 20 min. For **2**, 3 mmol potassium carbonate was added at r.t. and stirred for 1 h. Solution was filtered on a pad of Celite and concentrated to give pure bromoacetamide product. For **5**, 4 mmol DIPEA was added on ice then mixture was moved to r.t. and stirred for 1 h. Adding

15 mL 1 M HCl and 30 mL diethyl ether, and washing with 1 M sodium bicarbonate and drying with sodium sulfate and concentrating gave chloroacetamide product.

*Synthesis of N-Boc amide benzothiazole propargylamide.* 1 mmol of *N*-Boc benzothiazole 6-acetic acid (Combi-Block) was dissolved in 10 mL ACN. 5 mmol of DIPEA, 0.1 mmol of HOBt, 1 mmol of DMAP, and 1.2 mmol of propargylamine was added. Finally 1 mmol EDC was added, and the mixture was stirred at r.t. and monitored with TLC. Typical reaction lasted 24 h. Then it was concentrated and extracted by adding 20 mL EtOAc and 20 mL 1 M sodium bicarbonate and extracting out twice more with 20 mL EtOAc. Combined organic layer was dried and concentrated, and flash column chromatography with 5 % to 30 % MeOH in DCM gave pure amide product.



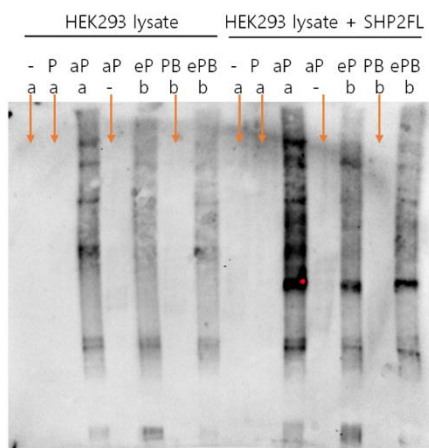
**Figure 5.5: Comparison of hexynyl-tagged probes with ethynyl- and propargylamide-tagged probes.** Only minor differences in reactivity were observed, both with pNPP (left) and by in-gel ABPP (right).

*Cell culture.* HEK293 cells were obtained from ATCC and grown in DMEM with L-alanine-L-glutamine (Gibco) supplemented with 10 % FBS (Gibco) and 1 % antibiotics/antimycotics mix (Gibco). Jurkat T cells were obtained from ATCC and grown in RPMI with L-alanine-L-glutamine (Gibco) supplemented with 10 % FBS and 1 % antibiotics/antimycotics mix.

*Cell lysis.* Whether before treatment or after, cell lysis was performed with the lysis buffer containing 50 mM HEPES, 1 % Triton X-100, 0.1 % sodium deoxycholate, NaCl so that ionic strength is 150 mM, and protease inhibitor cocktail added at each time of use. Cells were washed with PBS first, then scraped off mechanically and pelleted. Dry pellet was resuspended in 3-5 times the volume of lysis buffer and incubated at 0-4 °C for 25 min. Insoluble part was removed by centrifuging at 17000 rpm for 15 min in a refrigerated tabletop centrifuge. Lysate was measured for total protein content using BCA assay, then diluted with lysis buffer to 1 g/L.

*Visualizing probe reaction on proteome with broadly-reactive dye.* When done on lysate, 24  $\mu$ L lysate was mixed with 1  $\mu$ L of compound stock solution and incubated for 30 min. When done on cells, HEK293 cells were grown on 6-well plates to about 90 % confluence. Medium was exchanged to 1980  $\mu$ L medium (DMEM) + 20  $\mu$ L compound stocks. After incubating for 1 h, cells were immediately lysed according to lysis protocol. The resulting lysate (25  $\mu$ L) from either was incubated with 1  $\mu$ L 25  $\mu$ M IA-ATTO488 stock solution and reacted for another 1 h. Reaction was quenched by adding 4x sample buffer, and SDS-PAGE was run enough to ensure free dye removal. In-gel fluorescence was imaged using either ChemiDoc MP (Bio-Rad) or Typhoon 5 (GE) imager.

*Gel-based ABPP.* When done on lysate, 24  $\mu\text{L}$  lysate was mixed with 1  $\mu\text{L}$  of compound stock solution and incubated for 30 min. When done on cells, HEK293 cells were grown on 6-well plates to about 90 % confluence. Medium was exchanged to 1980  $\mu\text{L}$  medium (DMEM) + 20  $\mu\text{L}$  compound stocks. After incubating for 1 h, cells were immediately lysed according to lysis protocol. The resulting lysate (25  $\mu\text{L}$ ) was mixed with 1.25  $\mu\text{L}$  of azide-ATTO633 (For fluorescence) or azide-PEG<sub>3</sub>-biotin stock (for blotting, 5 mM in DMSO, alkyne-biotin in case of using **1at** probe), 1.25  $\mu\text{L}$  of 50 mM sodium ascorbate aqueous solution, 3.75  $\mu\text{L}$  of TBTA stock (1.7 mM in DMSO) and 1.25  $\mu\text{L}$  of copper sulfate stock (50 mM in water). After 1 h of incubation, 10.83  $\mu\text{L}$  4x sample loading buffer was added and standard SDS-PAGE protocol was followed. In-gel fluorescence was imaged on Typhoon 5 (GE) imager. In case of blotting, transfer to membrane was done using Turbo-Blot (Bio-Rad), blocking was done for 1 h in 5 % BSA (w/v) in TBS-T, then the membrane was incubated in 10 mL 1:1000 streptavidin-IR680 (Li-Cor) in blocking solution for 1 h. After washing 3 times in TBS-T. Blots were imaged on Typhoon 5 (GE) imager. Another gel was loaded identically and stained with Coomassie Blue for loading control.



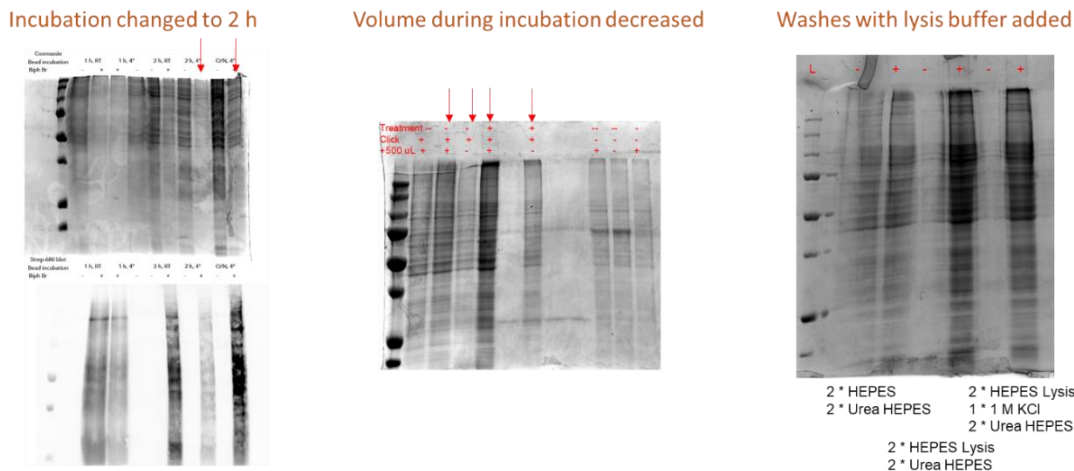
**Figure 5.6: Checking for background signals of click chemistry.** Groups that were not treated with compounds at all, those that were treated with untagged compounds, and those that were

not added with a linker during click chemistry step (all marked with arrows) did not show signal after blotting with streptavidin-HRP.

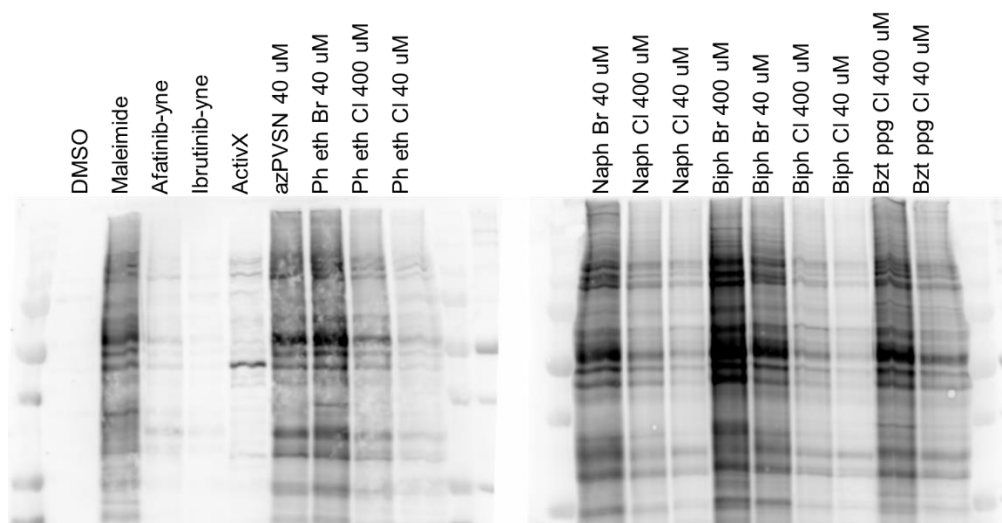
*Sample preparation for MS-based ABPP.* When done on lysate, 96  $\mu\text{L}$  lysate was mixed with 4  $\mu\text{L}$  of compound stock solution and incubated for 30 min. When done on cells, cells were grown in 6-well plates to about 90 % confluence. Medium was exchanged to 1.98 mL medium (DMEM) + 20  $\mu\text{L}$  compound stocks. After incubating for 1 h, cells were immediately lysed according to lysis protocol. The resulting lysate (100  $\mu\text{L}$ ) was mixed with 70  $\mu\text{L}$  HEPES buffer (pH 7.5, 50 mM HEPES, ionic strength 150 mM), then 5  $\mu\text{L}$  azide-PEG<sub>3</sub>-biotin stock (for blotting, 5 mM in DMSO, alkyne-biotin in case of using **1at** probe), 5  $\mu\text{L}$  of 50 mM sodium ascorbate aqueous solution, 15  $\mu\text{L}$  of TBTA stock (1.7 mM in DMSO) and 5  $\mu\text{L}$  of copper sulfate stock (50 mM in water) were added. After 1 h of incubation, click reaction was quenched by adding EDTA solution for a final 2.5 mM, and samples were desalted using desalting columns (Pierce, MWCO 7k). 10  $\mu\text{L}$  of each sample was saved for blot analysis. Streptavidin magnetic beads (Pierce) were washed and each sample was mixed with 17  $\mu\text{L}$  of the beads solution, and was rotated in 4 °C for 2 h. Beads were washed with 200  $\mu\text{L}$  lysis buffer twice, then with 2 M urea in HEPES buffer. Beads were reconstituted in 80  $\mu\text{L}$  2 M urea in HEPES buffer that contains 1 mM DTT and 0.4  $\mu\text{g}$  trypsin (sequencing grade, Promega). This was shaken (1000 rpm) for 1 h at r.t.. Supernatant was combined with two washes of 60  $\mu\text{L}$  2 M urea in HEPES. DTT was added for a final 4 mM and mixture shaken (1000 rpm) for 30 min at r.t.. IAA was added for a final 10 mM and mixture shaken (1000 rpm) for 45 min in the dark. 0.5  $\mu\text{g}$  trypsin was added per sample, and sample was shaken (500 rpm) overnight. Samples were acidified to a final concentration of 1 % formic acid. Peptides were desalted using C18 StageTips. Samples were dried, then reconstituted in 40  $\mu\text{L}$  of pH 8.5 HEPES. Each TMTpro 18-plex (Thermo Scientific) reagent was added to each group (0.2 mg in



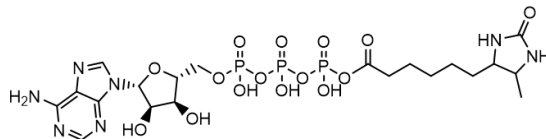
16.4  $\mu\text{L}$  ACN). After 1h incubation, the reaction was stopped with 4  $\mu\text{L}$  5 % hydroxylamine for 20 min at r.t.. Differently labeled peptides were mixed, desalted using C18 cartridges (Waters), evaporated to dryness in a vacuum concentrator, and reconstituted in 3 % ACN, 0.2 % formic acid solution.



**Figure 5.7: Optimization of bead washing protocol to minimize nonspecific binding.**



**Figure 5.8: An example of a sample blot during proteomics sample preparation.** The lane intensities of each group reflect the reactivity of the compounds well.



Name	ActivX enrichment score
DCN1-like protein;DCN1-like protein 1	6.603667
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	3.574839
UDP-glucose:glycoprotein glucosyltransferase 1	2.949668
Endoplasmic	2.758632
Ribonuclease H2 subunit C	2.746093
NF-kappa-B essential modulator	2.69782
La-related protein 4B	2.549297
Cyclin-dependent-like kinase 5	2.478134
Inosine-5-monophosphate dehydrogenase 2	2.464429
Ribosomal protein S6 kinase;Ribosomal protein S6 kinase alpha-2;Ribosomal protein S6 kinase alpha-3	2.449229
NEDD8-activating enzyme E1 catalytic subunit	2.403686
Serine/threonine-protein kinase OSR1	2.166272
Activated RNA polymerase II transcriptional coactivator p15	2.115437
JMP-CMP kinase	2.114771
Acetyl-CoA carboxylase 1;Biotin carboxylase	2.105273
Arfaptin-2	2.101284
Ketosamine-3-kinase	2.090871
Myosin regulatory light chain 12A;Myosin regulatory light chain 12B;Myosin regulatory light polypeptide 9	2.086995
Thymidylate kinase	2.042776

**Figure 5.9: Verification of MS proteomics protocol.** A commercial kinase ABP kit (ActivX, structure shown) was included in lysate-based proteomics work. Out of top 19 protein groups enriched, 6 were kinases.

*LC-MS/MS analysis.* About 1  $\mu\text{g}$  of total peptides were analyzed on a Waters M-Class UPLC using a 25 cm Ionopticks Aurora column coupled to a benchtop Orbitrap Q Exactive HF (Thermo Fisher) mass spectrometer. Peptides were separated at a flow rate of 400 nL/min with a 100 min gradient, including sample loading and column equilibration times. Data was acquired in data dependent mode using Xcalibur 4.1 software. MS1 Spectra were measured with a resolution of 120000, an AGC target of 3e6 and a mass range from 300 to 1800 m/z. Up to 12 MS2 spectra per duty cycle were triggered at a resolution of 60,000, an AGC target of 1e5, an isolation window of 0.8 m/z, a normalized collision energy of 28, a scan range of 200 to 2000 m/z, and a fixed first mass of 110 m/z.

*Analysis of MS proteomics result.* All raw data were analyzed with MaxQuant software version 2.1.3.0 using a UniProt human database (UP000005640), and MS/MS searches were performed with the following parameters: TMTpro 18-plex labeling on the MS2 level, M oxidation and N-terminal acetylation as variable modifications; C carbamidomethylation as fixed modification; Trypsin/P as the digestion enzyme; precursor ion mass tolerances of 20 ppm for the first search (used for nonlinear mass recalibration) and 4.5 ppm for the main search, and a fragment ion mass tolerance of 20 ppm. For identification, we applied a maximum FDR of 1% separately on protein and peptide level. We required 1 or more unique/razor peptides for protein identification and at least two MS/MS spectra ratio counts for quantification for each TMT channel. This gave us a total of 2062 quantified protein groups. Next, we normalized the corrected TMT MS2 intensity such that at each condition (sample) these intensity values added up to exactly 1000000, therefore each protein group value can be regarded as a normalized microshare – separately for each TMT channel for all proteins that made our filter cutoff in all channels – of the label-free quantification from MS1 of each protein group.

## Chapter 6: Conclusion and future outlooks

### 6.1. Overview

In this thesis, I described what started from an analysis of existing vinylsulfonate probes and turned into an extensive analysis of covalent ligands for the catalytic cysteines of protein tyrosine phosphatases. We took a systematic approach of assessing different reactive groups, then looking into how various scaffold structure may serve as recognition elements for target PTPs. For this process, a series of assays for the reactivity of thiol-reactive compounds, both towards the active site catalytic residue and all other off-target residues, was developed. While on a simple phenyl scaffold highly reactive warheads inhibit PTPs in the micromolar range, less reactive warheads show little to no inhibition. Upon examining a series of structures of compounds with several warheads, we were able to find out the relative reactivity of individual PTPs with the compounds, and also compare within different groups of compounds: for example, thiazoles showed faster inhibition rate than aromatic hydrocarbon structures did. We then also examined the intrinsic reactivity of these compounds, which showed a larger range than the reactivity towards the active site cysteine. Moreover, there was not a noticeable correlation between the two values, and thus some groups performed better on average as inhibitors while having a similar reactivity towards thiols in general. Structure-activity relationships could also be extracted from comparing intrinsic reactivity with the reactivity with PTPs. We also discussed what considerations researchers should take into account when analyzing kinetic measurements of covalent inhibitors, and showed that it is possible to extract more information from a small time series of a concentration series inhibition data. Compounds were then tested against the proteome, first by a competitive method that does not require a tag moiety on the probes. Finally, we showed that the

compounds can be used to probe the proteome, with the alkyne-tagged derivatives of several compounds that were synthesized showing enrichment of several PTPs.

## **6.2. Towards an ideal ABP**

The key to any successful ABPP approach is having the right probe that balances reactivity and selectivity. If a probe is too reactive, it will label the proteome indiscriminately and it will be hard to get meaningful information. If it is too unreactive, researchers will have a hard time getting a consistent readout of anything that the probe labels. Then, if the probe is to be a directed activity-based probe, the specificity needs to be targeted towards a specific class of enzymes. A reliable way to ensure that the probe is specific to a target family is if the probe is highly mechanism-based, in which case a microenvironment specific to the target family is necessary for the labeling reaction to take place. In this case, though, we opted for simpler electrophiles for cell permeability and stability.

For directed ABPs, the reactivity for target active sites is essential, and having that ensures that probes are able to light up the target proteins in MS-based ABPP experiments. Moreover, as long as the probes are reactive towards target active sites they can be employed in a competitive ABPP screens for potential inhibitors. This may be quantified easily through the use of  $IC_{50}$  values, or similarly, covalent efficiency. We showed that many of our probes are capable of this, having a low micromolar to high nanomolar  $IC_{50}$ ; application of our probes to competitive inhibitor screening is certainly an interesting possibility. Then for the specificity aspect, the matter is to drive down the off-target reactivity sufficiently without sacrificing on-target reactivity too much, for the differences in response for target active sites in various biological contexts to be read out without noise drowning them out. This is not inherently connected to the kinetic parameters as

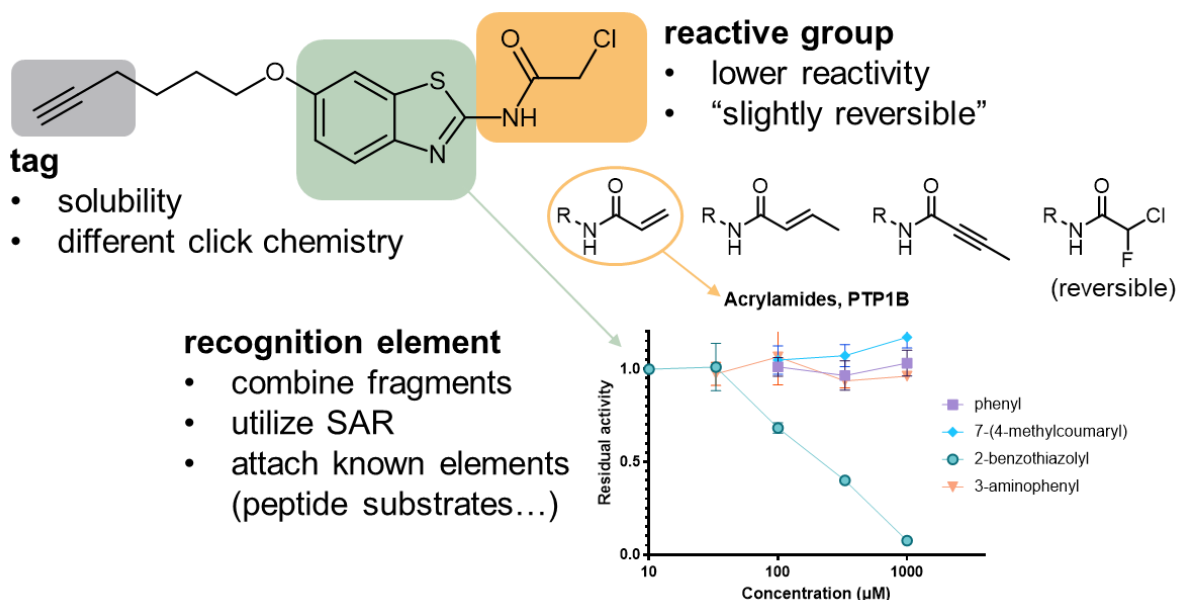
those are PTP-specific constants – for example, a sufficiently mechanism-based probe will not have off-target reactivity regardless of those constants – but  $k_{\text{inact}}$  may serve as a predictor for off-target reactivity, as was shown in Chapter 4.

Lastly, an ideal ABP will have preference for the active conformation of PTPs over the inactive one, thus reporting on the activity states. A built-in function with any thiol-reactive probe is that it will not react with post-translationally modified Cys, such as oxidation or nitrosylation, meaning that our probes can report on one regulation aspect of PTPs.

### 6.3. Future directions

To improve on the proteomics application of this research, several things can be done (Figure 6.1). Seeing that reducing the reactivity of the compound can increase specificity, several less reactive warheads, such as acrylamide or butynamide, can be substituted in. Indeed, upon quick synthesis and testing 2-benzothiazolyl acrylamide showed promise, displaying PTP inhibition on par with that of many chloroacetamides, despite the fact that other acrylamide molecules did not display any inhibitory activities. As our probes are mostly modular in synthesis, it will be possible to find an ideal combination of scaffold and warhead for their performance as ABPs. Scaffold structure can be improved further for ABPP directed to PTPs with high specificity. Fragments identified from our screen may be brought together for a more potent design. Another promising way of doing that is attachment of known recognition elements, such as target-specific peptides or nanobodies. A tag-free approach for proteomics may also be attempted by utilizing a competitive cysteine-reactive tag in similar fashion to the competitive in-gel assay, seeing that most compounds are quite specific to cysteine. Last but not least, proteomic approaches should be

utilized on probing a biological system, such as how the activity levels of PTPs change before and after T cell receptor activation, or insulin signaling.



**Figure 6.1: Potential improvements to functional parts of probes.** Some examples of less reactive groups are shown. Sample inhibition data for acrylamides are shown.

Another significance of this work is that we have laid out a rather unusual path of target-centric approach in the development of an active site-directed probe. We showed an iterative process of assessing various reactive groups, expanding the scaffold selections by sampling from promising chemical space, and analyzing the two-dimensional matrix by a unique combination of several biochemical assays. Similar courses can be utilized in future developments of activity-based probes both for PTPs and other families of enzymes with a defined target residue but without a suitable probe with a high resemblance to catalytic mechanism. Targets may even be identified first from a broad fragment screening data, upon which one can build using a warhead and scaffold varying approach.

One byproduct of this approach is that we identified a series of broadly reactive groups. There are various useful implications to cysteine reactivity, such as alkylation of cysteine groups for digestion and analysis – while researchers routinely use IAA, issues such as overalkylation exist,<sup>129–131</sup> and chloroacetamides and acrylamides are also options<sup>131,132</sup> – and deeper dives into the “cysteinome” in the context of disulfide bridges<sup>133</sup> or cysteine oxidation.<sup>40,134</sup> We showed that probes that are broadly reactive can still have distinct reactivity profiles. A more concrete characterization of the reactivity profile for several compounds with tuned scaffold structures for local environment sensing, *a la* “ligandability” obtained from fragment screens,<sup>135</sup> can yield a combination of probes that will provide valuable new insights, especially with the rapid advancement of data handling techniques. Finally, application to conformational mapping<sup>136</sup> or crosslinking MS analysis<sup>137</sup> using various broadly-reactive probes, including lysine-reactive or multi-reactive ones can be an interesting way to probe conformations and interactions across the proteome at different states.



## References

1. Lothrop, A. P., Torres, M. P. & Fuchs, S. M. Deciphering post-translational modification codes. *FEBS Lett.* **587**, 1247–1257 (2013).
2. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed* **44**, 7342–7372 (2005).
3. Uversky, V. N. Posttranslational Modification. in *Brenner's encyclopedia of genetics* 425–430 (Elsevier, 2013). doi:10.1016/B978-0-12-374984-0.01203-1.
4. Prabakaran, S., Lippens, G., Steen, H. & Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **4**, 565–583 (2012).
5. Modi, V. & Dunbrack, R. L. Defining a new nomenclature for the structures of active and inactive kinases. *Proc Natl Acad Sci USA* **116**, 6818–6827 (2019).
6. Zhang, M. *et al.* Construction and Deciphering of Human Phosphorylation-Mediated Signaling Transduction Networks. *J. Proteome Res.* **14**, 2745–2757 (2015).
7. Balla, T. Phosphoinositides: tiny lipids with giant impact on cell regulation. *Physiol. Rev.* **93**, 1019–1137 (2013).
8. Camden, A. J., Walsh, S. M., Suk, S. H. & Silverman, S. K. DNA Oligonucleotide 3'-Phosphorylation by a DNA Enzyme. *Biochemistry* **55**, 2671–2676 (2016).
9. Wilson, J. E. Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *J. Exp. Biol.* **206**, 2049–2057 (2003).

10. Lim, S. *et al.* Regulation of mitochondrial functions by protein phosphorylation and dephosphorylation. *Cell Biosci.* **6**, 25 (2016).
11. Tonks, N. K. Protein tyrosine phosphatases--from housekeeping enzymes to master regulators of signal transduction. *FEBS J.* **280**, 346–378 (2013).
12. Stanford, S. M. & Bottini, N. Targeting tyrosine phosphatases: time to end the stigma. *Trends Pharmacol. Sci.* **38**, 524–540 (2017).
13. Köhn, M. Turn and face the strange: A new view on phosphatases. *ACS Cent. Sci.* **6**, 467–477 (2020).
14. Shi, Y. Serine/threonine phosphatases: mechanism through structure. *Cell* **139**, 468–484 (2009).
15. Tautz, L., Critton, D. A. & Grotegut, S. Protein tyrosine phosphatases: structure, function, and implication in human disease. *Methods Mol. Biol.* **1053**, 179–221 (2013).
16. Chen, M. J., Dixon, J. E. & Manning, G. Genomics and evolution of protein phosphatases. *Sci. Signal.* **10**, (2017).
17. Lin, G., Aranda, V., Muthuswamy, S. K. & Tonks, N. K. Identification of PTPN23 as a novel regulator of cell invasion in mammary epithelial cells from a loss-of-function screen of the “PTP-ome”. *Genes Dev.* **25**, 1412–1425 (2011).
18. Krishnamurthy, D. & Barrios, A. M. Profiling protein tyrosine phosphatase activity with mechanistic probes. *Curr. Opin. Chem. Biol.* **13**, 375–381 (2009).
19. Alonso, A. *et al.* Protein tyrosine phosphatases in the human genome. *Cell* **117**, 699–711 (2004).
20. Barr, A. J. *et al.* Large-scale structural analysis of the classical human protein tyrosine phosphatome. *Cell* **136**, 352–363 (2009).

21. Li, H. *et al.* Crystal structure and substrate specificity of PTPN12. *Cell Rep.* **15**, 1345–1358 (2016).
22. Peters, G. H. *et al.* Residue 259 is a key determinant of substrate specificity of protein-tyrosine phosphatases 1B and alpha. *J. Biol. Chem.* **275**, 18201–18209 (2000).
23. Hippen, K. L. *et al.* Acidic residues are involved in substrate recognition by two soluble protein tyrosine phosphatases, PTP-5 and rrbPTP-1. *Biochemistry* **32**, 12405–12412 (1993).
24. Chen, K.-E. *et al.* Substrate specificity and plasticity of FERM-containing protein tyrosine phosphatases. *Structure* **23**, 653–664 (2015).
25. Choy, M. S. *et al.* Conformational rigidity and protein dynamics at distinct timescales regulate PTP1B activity and allostery. *Mol. Cell* **65**, 644–658.e5 (2017).
26. Zhang, Z.-Y. Drugging the undruggable: therapeutic potential of targeting protein tyrosine phosphatases. *Acc. Chem. Res.* **50**, 122–129 (2017).
27. Lee, P. Y., Yeoh, Y. & Low, T. Y. A recent update on small-molecule kinase inhibitors for targeted cancer therapy and their therapeutic insights from mass spectrometry-based proteomic analysis. *FEBS J.* (2022) doi:10.1111/febs.16442.
28. Attwood, M. M., Fabbro, D., Sokolov, A. V., Knapp, S. & Schiöth, H. B. Trends in kinase drug discovery: targets, indications and inhibitor design. *Nat. Rev. Drug Discov.* **20**, 839–861 (2021).
29. Qiu, W. *et al.* Structural insights into Noonan/LEOPARD syndrome-related mutants of protein-tyrosine phosphatase SHP2 (PTPN11). *BMC Struct. Biol.* **14**, 10 (2014).
30. Tiganis, T. PTP1B and TCPTP--nonredundant phosphatases in insulin signaling and glucose homeostasis. *FEBS J.* **280**, 445–458 (2013).

31. Courtney, A. H. *et al.* CD45 functions as a signaling gatekeeper in T cells. *Sci. Signal.* **12**, (2019).
32. Thura, M. *et al.* PRL3-zumab as an immunotherapy to inhibit tumors expressing PRL3 oncoprotein. *Nat. Commun.* **10**, 2484 (2019).
33. Boutros, R., Lobjois, V. & Ducommun, B. CDC25 phosphatases in cancer cells: key players? Good targets? *Nat. Rev. Cancer* **7**, 495–507 (2007).
34. Zhang, Z. Y., Wang, Y. & Dixon, J. E. Dissecting the catalytic mechanism of protein-tyrosine phosphatases. *Proc Natl Acad Sci USA* **91**, 1624–1627 (1994).
35. Zhang, Z. Y. Kinetic and mechanistic characterization of a mammalian protein-tyrosine phosphatase, PTP1. *J. Biol. Chem.* **270**, 11199–11204 (1995).
36. Jia, Z., Barford, D., Flint, A. J. & Tonks, N. K. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science* **268**, 1754–1758 (1995).
37. Karisch, R. & Neel, B. G. Methods to monitor classical protein-tyrosine phosphatase oxidation. *FEBS J.* **280**, 459–475 (2013).
38. Yu, Z.-H. & Zhang, Z.-Y. Regulatory mechanisms and novel therapeutic targeting strategies for protein tyrosine phosphatases. *Chem. Rev.* **118**, 1069–1091 (2018).
39. Salmeen, A. *et al.* Redox regulation of protein tyrosine phosphatase 1B involves a sulphenyl-amide intermediate. *Nature* **423**, 769–773 (2003).
40. Chen, Y.-Y. *et al.* Cysteine S-nitrosylation protects protein-tyrosine phosphatase 1B against oxidation-induced permanent inactivation. *J. Biol. Chem.* **283**, 35265–35272 (2008).
41. Chen, Y.-N. P. *et al.* Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* **535**, 148–152 (2016).

42. Yuan, X., Bu, H., Zhou, J., Yang, C.-Y. & Zhang, H. Recent advances of SHP2 inhibitors in cancer therapy: current development and clinical application. *J. Med. Chem.* **63**, 11368–11396 (2020).
43. Guo, Y., Xu, Y., Dong, X. & Zhang, J. Cross the undruggable barrier, the development of SHP2 inhibitors: from catalytic site inhibitors to allosteric inhibitors. *ChemistrySelect* **6**, 5504–5523 (2021).
44. Cui, D. S., Lipchock, J. M., Brookner, D. & Loria, J. P. Uncovering the molecular interactions in the catalytic loop that modulate the conformational dynamics in protein tyrosine phosphatase 1B. *J. Am. Chem. Soc.* **141**, 12634–12647 (2019).
45. Haque, A., Andersen, J. N., Salmeen, A., Barford, D. & Tonks, N. K. Conformation-sensing antibodies stabilize the oxidized form of PTP1B and inhibit its phosphatase activity. *Cell* **147**, 185–198 (2011).
46. Li, S. *et al.* The mechanism of allosteric inhibition of protein tyrosine phosphatase 1B. *PLoS ONE* **9**, e97668 (2014).
47. Zhang, J., Zhou, B., Zheng, C. & Zhang, Z. A bipartite mechanism for ERK2 recognition by its cognate regulators and substrates. *J. Biol. Chem.* **278**, 29901–29912 (2003).
48. Bilwes, A. M., den Hertog, J., Hunter, T. & Noel, J. P. Structural basis for inhibition of receptor protein-tyrosine phosphatase- $\alpha$  by dimerization. *Nature* **382**, 555–559 (1996).
49. Heinrich, F. *et al.* The PTEN tumor suppressor forms homodimers in solution. *Structure* **23**, 1952–1957 (2015).
50. Papa, A. *et al.* Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. *Cell* **157**, 595–610 (2014).

51. Fueller, J. *et al.* Subcellular partitioning of protein tyrosine phosphatase 1B to the endoplasmic reticulum and mitochondria depends sensitively on the composition of its tail anchor. *PLoS ONE* **10**, e0139429 (2015).
52. den Hertog, J., Ostman, A. & Böhmer, F.-D. Protein tyrosine phosphatases: regulatory mechanisms. *FEBS J.* **275**, 831–847 (2008).
53. Willems, L. I., Overkleeft, H. S. & van Kasteren, S. I. Current developments in activity-based protein profiling. *Bioconjug. Chem.* **25**, 1181–1191 (2014).
54. Verhelst, S. H. L., Bongers, K. M. & Willems, L. I. Bioorthogonal Reactions in Activity-Based Protein Profiling. *Molecules* **25**, (2020).
55. Killinger, B. J., Brandvold, K. R., Ramos-Hunter, S. J. & Wright, A. T. Chemoproteomic analyses by activity-based protein profiling. in *Mass spectrometry-based chemical proteomics* (eds. Tao, W. A. & Zhang, Y.) 67–99 (Wiley, 2019). doi:10.1002/9781118970195.ch3.
56. Niphakis, M. J. & Cravatt, B. F. Enzyme inhibitor discovery by activity-based protein profiling. *Annu. Rev. Biochem.* **83**, 341–377 (2014).
57. Simon, G. M. & Cravatt, B. F. Activity-based proteomics of enzyme superfamilies: serine hydrolases as a case study. *J. Biol. Chem.* **285**, 11051–11055 (2010).
58. Wu, L. *et al.* An overview of activity-based probes for glycosidases. *Curr. Opin. Chem. Biol.* **53**, 25–36 (2019).
59. Joyce, J. A. *et al.* Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. *Cancer Cell* **5**, 443–453 (2004).
60. Cravatt, B. F., Wright, A. T. & Kozarich, J. W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* **77**, 383–414 (2008).

61. Saghatelian, A., Jessani, N., Joseph, A., Humphrey, M. & Cravatt, B. F. Activity-based probes for the proteomic profiling of metalloproteases. *Proc Natl Acad Sci USA* **101**, 10000–10005 (2004).
62. Chen, X. *et al.* Target identification with quantitative activity based protein profiling (ABPP). *Proteomics* **17**, (2017).
63. Nomura, D. K., Dix, M. M. & Cravatt, B. F. Activity-based protein profiling for biochemical pathway discovery in cancer. *Nat. Rev. Cancer* **10**, 630–638 (2010).
64. Ruddraraju, K. V. & Zhang, Z.-Y. Covalent inhibition of protein tyrosine phosphatases. *Mol. Biosyst.* **13**, 1257–1279 (2017).
65. Wang, Q., Dechert, U., Jirik, F. & Withers, S. G. Suicide inactivation of human prostatic acid phosphatase and a phosphotyrosine phosphatase. *Biochem. Biophys. Res. Commun.* **200**, 577–583 (1994).
66. Lo, L.-C. *et al.* Design and synthesis of class-selective activity probes for protein tyrosine phosphatases. *J. Proteome Res.* **1**, 35–40 (2002).
67. Zhu, Q., Huang, X., Chen, G. Y. J. & Yao, S. Q. Activity-based fluorescent probes that target phosphatases. *Tetrahedron Lett.* **44**, 2669–2672 (2003).
68. Kumar, S. *et al.* Global analysis of protein tyrosine phosphatase activity with ultra-sensitive fluorescent probes. *J. Proteome Res.* **5**, 1898–1905 (2006).
69. Zhang, Z. Y. & Dixon, J. E. Active site labeling of the Yersinia protein tyrosine phosphatase: the determination of the pK<sub>a</sub> of the active site cysteine and the function of the conserved histidine 402. *Biochemistry* **32**, 9340–9345 (1993).
70. Hsu, M.-F. *et al.* S-nitrosylation of endogenous protein tyrosine phosphatases in endothelial insulin signaling. *Free Radic. Biol. Med.* **99**, 199–213 (2016).

71. Liu, S. *et al.* Aryl vinyl sulfonates and sulfones as active site-directed and mechanism-based probes for protein tyrosine phosphatases. *J. Am. Chem. Soc.* **130**, 8251–8260 (2008).
72. Awoonor-Williams, E. & Rowley, C. N. Evaluation of methods for the calculation of the pKa of cysteine residues in proteins. *J. Chem. Theory Comput.* **12**, 4662–4673 (2016).
73. McCain, D. F., Catrina, I. E., Hengge, A. C. & Zhang, Z.-Y. The catalytic mechanism of Cdc25A phosphatase. *J. Biol. Chem.* **277**, 11190–11200 (2002).
74. Wang, W.-Q. *et al.* Kinetic and mechanistic studies of a cell cycle protein phosphatase Cdc14. *J. Biol. Chem.* **279**, 30459–30468 (2004).
75. Krishnan, N., Bencze, G., Cohen, P. & Tonks, N. K. The anti-inflammatory compound BAY-11-7082 is a potent inhibitor of protein tyrosine phosphatases. *FEBS J.* **280**, 2830–2841 (2013).
76. Lee, J., Rhee, M. H., Kim, E. & Cho, J. Y. BAY 11-7082 is a broad-spectrum inhibitor with anti-inflammatory activity against multiple targets. *Mediators Inflamm.* **2012**, 416036 (2012).
77. Ritorto, M. S. *et al.* Screening of DUB activity and specificity by MALDI-TOF mass spectrometry. *Nat. Commun.* **5**, 4763 (2014).
78. Strickson, S. *et al.* The anti-inflammatory drug BAY 11-7082 suppresses the MyD88-dependent signalling network by targeting the ubiquitin system. *Biochem. J.* **451**, 427–437 (2013).
79. Scapin, G., Patel, S., Patel, V., Kennedy, B. & Asante-Appiah, E. The structure of apo protein-tyrosine phosphatase 1B C215S mutant: more than just an S → O change. *Protein Sci.* **10**, 1596–1605 (2001).



80. Pádua, R. A. P. *et al.* Mechanism of activating mutations and allosteric drug inhibition of the phosphatase SHP2. *Nat. Commun.* **9**, 4507 (2018).
81. Shannon, D. A. & Weerapana, E. Covalent protein modification: the current landscape of residue-specific electrophiles. *Curr. Opin. Chem. Biol.* **24**, 18–26 (2015).
82. Petri, L., Ábrányi-Balogh, P., Varga, P. R., Imre, T. & Keserű, G. M. Comparative reactivity analysis of small-molecule thiol surrogates. *Bioorg. Med. Chem.* **28**, 115357 (2020).
83. Ábrányi-Balogh, P. *et al.* A road map for prioritizing warheads for cysteine targeting covalent inhibitors. *Eur. J. Med. Chem.* **160**, 94–107 (2018).
84. Böhme, A., Thaens, D., Paschke, A. & Schüürmann, G. Kinetic glutathione chemoassay to quantify thiol reactivity of organic electrophiles--application to alpha,beta-unsaturated ketones, acrylates, and propiolates. *Chem. Res. Toxicol.* **22**, 742–750 (2009).
85. Lonsdale, R. *et al.* Expanding the armory: predicting and tuning covalent warhead reactivity. *J. Chem. Inf. Model.* **57**, 3124–3137 (2017).
86. Resnick, E. *et al.* Rapid Covalent-Probe Discovery by Electrophile-Fragment Screening. *J. Am. Chem. Soc.* **141**, 8951–8968 (2019).
87. Martin, J. S., MacKenzie, C. J., Fletcher, D. & Gilbert, I. H. Characterising covalent warhead reactivity. *Bioorg. Med. Chem.* **27**, 2066–2074 (2019).
88. Flanagan, M. E. *et al.* Chemical and computational methods for the characterization of covalent reactive groups for the prospective design of irreversible inhibitors. *J. Med. Chem.* **57**, 10072–10079 (2014).
89. Kirsch, P., Hartman, A. M., Hirsch, A. K. H. & Empting, M. Concepts and Core Principles of Fragment-Based Drug Design. *Molecules* **24**, (2019).

90. Sun, J.-P., Wu, L., Fedorov, A. A., Almo, S. C. & Zhang, Z.-Y. Crystal structure of the Yersinia protein-tyrosine phosphatase YopH complexed with a specific small molecule inhibitor. *J. Biol. Chem.* **278**, 33392–33399 (2003).
91. Moise, G., Gallup, N. M., Alexandrova, A. N., Hengge, A. C. & Johnson, S. J. Conservative tryptophan mutants of the protein tyrosine phosphatase YopH exhibit impaired WPD-loop function and crystallize with divanadate esters in their active sites. *Biochemistry* **54**, 6490–6500 (2015).
92. Pola, S. Significance of Thiazole-based Heterocycles for Bioactive Systems. in *Scope of Selective Heterocycles from Organic and Pharmaceutical Perspective* (ed. Varala, R.) (InTech, 2016). doi:10.5772/62077.
93. Agrawal, N. & Mishra, P. The synthetic and therapeutic expedition of isoxazole and its analogs. *Med. Chem. Res.* **27**, 1309–1344 (2018).
94. Mathew, B., Hobrath, J. V., Connelly, M. C., Guy, R. K. & Reynolds, R. C. Oxazole and thiazole analogs of sulindac for cancer prevention. *Future Med. Chem.* **10**, 743–753 (2018).
95. Petrou, A., Fesatidou, M. & Geronikaki, A. Thiazole Ring-A Biologically Active Scaffold. *Molecules* **26**, (2021).
96. Perrin, C. L. Linear or Nonlinear Least-Squares Analysis of Kinetic Data? *J. Chem. Educ.* **94**, 669–672 (2017).
97. Kathman, S. G., Xu, Z. & Statsyuk, A. V. A fragment-based method to discover irreversible covalent inhibitors of cysteine proteases. *J. Med. Chem.* **57**, 4969–4974 (2014).
98. Andersen, H. S. *et al.* 2-(oxalylamino)-benzoic acid is a general, competitive inhibitor of protein-tyrosine phosphatases. *J. Biol. Chem.* **275**, 7101–7108 (2000).

99. Strelow, J. M. A perspective on the kinetics of covalent and irreversible inhibition. *SLAS Discov.* **22**, 3–20 (2017).
100. Rodriguez, J.-M. G., Hux, N. P., Philips, S. J. & Towns, M. H. Michaelis–menten graphs, lineweaver–burk plots, and reaction schemes: investigating introductory biochemistry students’ conceptions of representations in enzyme kinetics. *J. Chem. Educ.* (2019) doi:10.1021/acs.jchemed.9b00396.
101. Copeland, R. A. Irreversible Enzyme Inactivators. in *Evaluation of enzyme inhibitors in drug discovery: A guide for medicinal chemists and pharmacologists* 345–382 (John Wiley & Sons, Inc., 2013). doi:10.1002/9781118540398.ch9.
102. Kitz, R. & Wilson, I. B. Esters of methanesulfonic acid as irreversible inhibitors of acetylcholinesterase. *J. Biol. Chem.* **237**, 3245–3249 (1962).
103. Malcolm, A. D. & Radda, G. K. The reaction of glutamate dehydrogenase with 4-iodoacetamido salicylic acid. *Eur. J. Biochem.* **15**, 555–561 (1970).
104. Kuzmič, P. A steady-state algebraic model for the time course of covalent enzyme inhibition. *BioRxiv* (2020) doi:10.1101/2020.06.10.144220.
105. Fersht, A. *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding.* (WORLD SCIENTIFIC, 2017). doi:10.1142/10574.
106. Childs, R. E. & Bardsley, W. G. Time-dependent inhibition of enzymes by active-site-directed reagents. A theoretical treatment of the kinetics of affinity labelling. *J. Theor. Biol.* **53**, 381–394 (1975).
107. Lente, G. Solving Rate Equations. in *Deterministic kinetics in chemistry and systems biology* 21–59 (Springer International Publishing, 2015). doi:10.1007/978-3-319-15482-4\_2.

108. Constales, D., Yablonsky, G. S. & Marin, G. B. Intersections and coincidences in chemical kinetics: Linear two-step reversible–irreversible reaction mechanism. *Computers & Mathematics with Applications* **65**, 1614–1624 (2013).
109. Cornish-Bowden, A. Validity of a “steady-state” treatment of inactivation kinetics. *Eur. J. Biochem.* **93**, 383–385 (1979).
110. McWhirter, C. Kinetic mechanisms of covalent inhibition. in (Elsevier, 2021). doi:10.1016/bs.armc.2020.11.001.
111. Kim, H., Hwang, Y. S., Kim, M. & Park, S. B. Recent advances in the development of covalent inhibitors. *RSC Med. Chem.* **12**, 1037–1045 (2021).
112. Cheng, Y. & Prusoff, W. H. Relationship between the inhibition constant (K<sub>1</sub>) and the concentration of inhibitor which causes 50 per cent inhibition (I<sub>50</sub>) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).
113. Krippendorff, B.-F., Neuhaus, R., Lienau, P., Reichel, A. & Huisinga, W. Mechanism-based inhibition: deriving K(I) and k(inact) directly from time-dependent IC(50) values. *J. Biomol. Screen.* **14**, 913–923 (2009).
114. Kuzmič, P. A two-point IC<sub>50</sub> method for evaluating the biochemical potency of irreversible enzyme inhibitors. *BioRxiv* (2020) doi:10.1101/2020.06.25.171207.
115. Saghatelian, A. & Cravatt, B. F. Assignment of protein function in the postgenomic era. *Nat. Chem. Biol.* **1**, 130–142 (2005).
116. Speers, A. E. & Cravatt, B. F. Profiling enzyme activities in vivo using click chemistry methods. *Chem. Biol.* **11**, 535–546 (2004).
117. Wright, A. T., Song, J. D. & Cravatt, B. F. A suite of activity-based probes for human cytochrome P450 enzymes. *J. Am. Chem. Soc.* **131**, 10692–10700 (2009).

118. Krysiak, J. M. *et al.* Activity-based probes for studying the activity of flavin-dependent oxidases and for the protein target profiling of monoamine oxidase inhibitors. *Angew. Chem. Int. Ed* **51**, 7035–7040 (2012).
119. Li, Y., Pan, M., Li, Y., Huang, Y. & Guo, Q. Thiol-yne radical reaction mediated site-specific protein labeling via genetic incorporation of an alkynyl-L-lysine analogue. *Org. Biomol. Chem.* **11**, 2624–2629 (2013).
120. Ekkebus, R. *et al.* On terminal alkynes that can react with active-site cysteine nucleophiles in proteases. *J. Am. Chem. Soc.* **135**, 2867–2870 (2013).
121. Qiao, Y. *et al.* Expressed Protein Ligation without Intein. *J. Am. Chem. Soc.* **142**, 7047–7054 (2020).
122. Mons, E. *et al.* The alkyne moiety as a latent electrophile in irreversible covalent small molecule inhibitors of cathepsin K. *J. Am. Chem. Soc.* **141**, 3507–3514 (2019).
123. Shiu, H.-Y. *et al.* Electron-deficient alkynes as cleavable reagents for the modification of cysteine-containing peptides in aqueous medium. *Chem. Eur. J* **15**, 3839–3850 (2009).
124. Gehringer, M. & Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **62**, 5673–5724 (2019).
125. Nazih, A. & Heissler, D. One-pot Conversion of t-Butyl Carbamates to Amides with Acyl Halide-Methanol Mixtures. *Synthesis (Mass)* **2002**, (2004).
126. Motiwala, H. F., Kuo, Y.-H., Stinger, B. L., Palfey, B. A. & Martin, B. R. Tunable Heteroaromatic Sulfones Enhance in-Cell Cysteine Profiling. *J. Am. Chem. Soc.* **142**, 1801–1810 (2020).

127. Pappireddi, N., Martin, L. & Wühr, M. A review on quantitative multiplexed proteomics. *Chembiochem* **20**, 1210–1224 (2019).
128. Wang, S. *et al.* Advanced Activity-Based Protein Profiling Application Strategies for Drug Development. *Front. Pharmacol.* **9**, 353 (2018).
129. Boja, E. S. & Fales, H. M. Overalkylation of a protein digest with iodoacetamide. *Anal. Chem.* **73**, 3576–3582 (2001).
130. Nielsen, M. L. *et al.* Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat. Methods* **5**, 459–460 (2008).
131. Müller, T. & Winter, D. Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents. *Mol. Cell. Proteomics* **16**, 1173–1187 (2017).
132. Suttapitugsakul, S., Xiao, H., Smekens, J. & Wu, R. Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics. *Mol. Biosyst.* **13**, 2574–2582 (2017).
133. Xu, L., Silva, M. J. S. A., Gois, P. M. P., Kuan, S. L. & Weil, T. Chemoselective cysteine or disulfide modification via single atom substitution in chloromethyl acryl reagents. *Chem. Sci.* **12**, 13321–13330 (2021).
134. Shakir, S., Vinh, J. & Chiappetta, G. Quantitative analysis of the cysteine redoxome by iodoacetyl tandem mass tags. *Anal. Bioanal. Chem.* **409**, 3821–3830 (2017).
135. Vinogradova, E. V. & Cravatt, B. F. Multiplexed proteomic profiling of cysteine reactivity and ligandability in human T cells. *STAR Protocols* **2**, 100458 (2021).
136. Potter, Z. E. *et al.* Parallel chemoselective profiling for mapping protein structure. *Cell Chem. Biol.* **27**, 1084-1096.e4 (2020).

137. Tang, X., Wippel, H. H., Chavez, J. D. & Bruce, J. E. Crosslinking mass spectrometry: A link between structural biology and systems biology. *Protein Sci.* **30**, 773–784 (2021).