Mapping the chemical space of active-site targeted covalent ligands for protein tyrosine phosphatases

Dr. Suk ho Hong\textsuperscript{1,+}, Sarah Y. Xi\textsuperscript{1,+}, Andrew C. Johns\textsuperscript{1}, Lauren C. Tang\textsuperscript{2}, Allyson Li\textsuperscript{1}, Madeleine N. Hum\textsuperscript{1}, Cassandra A. Chartier\textsuperscript{1}, Prof. Marko Jovanovic\textsuperscript{2}, Prof. Neel H. Shah\textsuperscript{1,*}

\textsuperscript{1} Department of Chemistry, Columbia University, New York, NY 10027
\textsuperscript{2} Department of Biological Sciences, Columbia University, New York, NY 10027

* These authors contributed equally to this work.
* corresponding author, email: neel.shah@columbia.edu

ORCIDs
Suk ho Hong: 0000-0002-6024-0685
Sarah Y. Xi: 0000-0001-7590-3202
Andrew Johns: 0000-0002-3197-1390
Lauren C. Tang: 0000-0001-8786-5418
Allyson Li: 0000-0003-2359-7703
Madeleine N. Hum: 0009-0004-3350-1677
Cassandra A. Chartier: 0009-0004-2474-9937
Marko Jovanovic: 0000-0001-9707-3377
Neel H. Shah: 0000-0002-1186-0626

Table of Contents Graphic

Table of Contents Text
Covalent chemical probes and inhibitors for tyrosine phosphatases have historically been challenging to develop. We report the characterization of diverse cysteine-reactive electrophiles and fragment-like scaffolds as covalent ligands for tyrosine phosphatases. Our investigation reveals chemical constraints for tyrosine phosphatase covalent inhibitor design and sheds light on the structural features of phosphatases that govern their susceptibility to covalent inhibition.
Abstract

Protein tyrosine phosphatases (PTPs) are an important class of enzymes that modulate essential cellular processes through protein dephosphorylation and are dysregulated in various disease states. There is demand for new compounds that target the active sites of these enzymes, for use as chemical tools to dissect their biological roles or as leads for the development of new therapeutics. In this study, we explore an array of electrophiles and fragment scaffolds to investigate the required chemical parameters for covalent inhibition of tyrosine phosphatases. Our analysis juxtaposes the intrinsic electrophilicity of these compounds with their potency against several classical PTPs, revealing chemotypes that inhibit tyrosine phosphatases while minimizing excessive, potentially non-specific reactivity. We also assess sequence divergence at key residues in PTPs to explain their differential susceptibility to covalent inhibition. We anticipate that our study will inspire new strategies to develop covalent probes and inhibitors for tyrosine phosphatases.

Keywords

tyrosine phosphatase, covalent inhibition, cysteine labeling, enzymes, inhibitors, structure-activity relationships, protein structures
Introduction

Protein tyrosine phosphorylation is a prevalent post-translational modification that can modulate enzyme activity, protein localization, protein stability, and protein-protein interactions. Protein tyrosine phosphatases (PTPs) make up a large family of over 100 enzymes in humans that catalyze the dephosphorylation of tyrosine residues on proteins.\[^1\,^2\] These enzymes counteract protein tyrosine kinases, which phosphorylate proteins on tyrosine residues. Many diseases, including cancers, arise from the dysregulation of tyrosine phosphorylation, and tyrosine phosphatases have been proposed as candidate drug targets for the treatment of these diseases.\[^3\] For example, the phosphatase PTP1B has been linked to diabetes and obesity, as it acts as a negative regulator of the insulin receptor signaling pathway by dephosphorylating the insulin receptor and its substrates.\[^4\,^5\] The phosphatase SHP2 has been shown to promote oncogenic signaling downstream of receptor tyrosine kinases.\[^6\,^7\] Both of these enzymes have been the targets of extensive preclinical drug discovery campaigns, and allosteric inhibitors of SHP2 are currently in clinical trials.\[^8\,^9\] Despite their physiological importance, the regulatory mechanisms and biological roles of many PTPs remain poorly defined, in part due to the lack of robust chemical tools to selectively inhibit PTPs or monitor their activity in cells.\[^10\]

Most tyrosine phosphatases have a catalytic cysteine that is directly responsible for substrate dephosphorylation. They also have a catalytic arginine on the same loop as the cysteine residue (P loop), as well as conserved aspartic acid and glutamine residues on the WPD and Q loops, respectively. These residues make up the conserved catalytic core of classical PTPs (Figure 1A, B).\[^11\,^12\] The WPD loop can adopt an open or closed conformation that is dependent on substrate or inhibitor binding (Figure 1A, left panel), and it has been suggested that this loop plays a regulatory role.\[^11\,^14\] In some phosphatases, such as SHP2, auxiliary non-catalytic domains physically occlude the active site to regulate activity.\[^15\,^18\] Furthermore, the catalytic cysteines of most PTPs are susceptible to reversible oxidation by reactive-oxygen species, and oxidation has been proposed as a dynamic regulatory mechanism.\[^19\,^21\] Other uncharacterized regulatory mechanisms likely exist for PTPs, making these enzymes an opportune target class for activity-based profiling.\[^22\] In this context, it is noteworthy that the catalytic cysteine in PTPs has a characteristically low \(pK_a\) due to the active site environment,\[^23\,^24\] which should make it a good target for covalent probes.

There has been an exciting revolution in chemical biology to examine the proteome-wide reactivity and ligandability of cysteine residues using mass spectrometry proteomics. These studies have served as a starting point for identification of new ligandable protein targets and the development of covalent probes and drugs.\[^25\,^26\] Cysteine reactivity across the proteome is generally assessed by measuring labeling efficiency with a highly reactive iodoacetamide probe.\[^27\,^29\] Cysteine ligandability, on the other hand, is examined by treating the proteome with small-molecule covalent fragments bearing weaker electrophiles, like chloroacetamide or acrylamide, followed by treatment with an iodoacetamide probe. The loss of iodoacetamide labeling at specific cysteines reveals the selectivity profile of fragment electrophiles across the proteome.\[^29\] Cysteine ligandability screens have the potential to simultaneously identify target proteins and lead compounds for covalent probe development.

In general, these chemoproteomic efforts are not targeted – instead, they yield large datasets on the ligandability of thousands of cysteine residues. These datasets can be leveraged for the development of covalent inhibitors, probes, and degraders for a variety of protein targets, particularly for non-active site cysteine residues.\[^25\] However, published proteome-wide ligandability screens have not yielded any substantial leads for PTP active site cysteine residues.\[^50\] Rather, these screens suggest that PTP catalytic cysteines are not very reactive relative to the entire proteome, despite their low \(pK_a\) values,\[^23\,^24\] and they are apparently unliganded by most fragments used in existing screens. Consequently, it is difficult to extract information from these global ligandability screens on how different structures impact covalent inhibition of PTPs.

In this study, we have taken a target-centric approach to identify structural leads for the development of covalent PTP inhibitors and probes. PTPs have historically been difficult to inhibit potently and selectively, due to their highly conserved and charged active site architectures.\[^31\] One solution has been to pursue allosteric inhibitors, which has proven successful for SHP2.\[^8\] Covalent inhibition provides a different opportunity, as the long lifetimes of protein-ligand complexes often yield potent responses.\[^25\] A key challenge in covalent inhibitor design, however, is to optimize potency toward the desired target while mitigating off-
target reactivity. Thus, we set out to understand the chemical parameters that govern covalent targeting of the conserved active-site cysteine across the classical PTP subfamily. Along these lines, there have been several noteworthy efforts towards the development of activity-based covalent probes for PTPs over the past two decades, but these efforts have focused primarily on phosphotyrosine isosteres (Figure 1C). Early candidates included mechanism-based quinone methide generating probes and α-bromobenzylphosphonates. These molecules generally have poor membrane permeability, display nonspecific labeling, or are susceptible to solvolysis at physiological pH. Aryl vinyl sulfones and sulfonates were presented as a promising class of molecules that alleviated previous shortcomings and posed potential for further development.

Here, we report that aryl vinyl sulfonates have exceedingly fast non-specific reactivity with cysteines at physiological pH, making them unsuitable for selective phosphatase labeling or inhibition in a cellular context. We present a broader exploration of chemical space beyond reactive phosphotyrosine isosteres for the covalent inhibition of tyrosine phosphatases. We assessed a variety of thiol-reactive groups and fragment scaffolds using biochemical assays on an array of purified PTPs. We combined this enzyme-targeted approach with compound reactivity assays and protein sequence/structure analyses to better understand how both ligand and protein structural features govern covalent PTP inhibition. Motivated by the lack of suitable covalent chemical tools to investigate PTPs, our studies identify lead compounds that will guide the future development of inhibitors and chemical probes for PTPs.

Results and Discussion

Aryl vinylsulfonates inhibit tyrosine phosphatases but are highly non-specific at physiological pH.

Phenyl vinyl sulfone (PVS, 1) and phenyl vinylsulfonate (PVSN, 2) were previously reported as promising active site-directed covalent ligands for PTPs. They were shown to be cell-permeable, owing to their relatively simple structures and lack of charged functional groups. An azide-tagged variant of PVSN was also reported as the starting point for activity-based profiling. We envisioned that a more elaborate scaffold structure, larger than a phenyl ring, might improve the potency of PVSN while retaining the potentially...
favorable features of the vinyl sulfonate moiety. Thus, we synthesized a series of aryl vinylsulfonates and assessed their ability to covalently label and inhibit tyrosine phosphatases using two complementary assays (Figure 2A,B). In the first assay, the residual activities of several tyrosine phosphatase catalytic domains were measured using the colorogenic substrate p-nitrophenyl phosphate (pNPP) after treatment with each covalent inhibitor. In the second assay, we assessed the extent of labeling of the same PTPs by each vinyl sulfonate using intact protein mass spectrometry.

As expected, aryl vinylsulfonates were able to inhibit several PTPs after treatment with 100 μM compound for one hour (Figure 2C). Phenyl vinyl sulfone (1) showed very little activity against every
phosphatase except HePTP. In general, vinylsulfonates on bulkier scaffolds (compounds 3-6) showed increased potency. However, intact protein mass spectrometry revealed that these molecules were labeling other residues in addition to the catalytic cysteine, as demonstrated by the existence of protein species with more than one adduct (Figure 2D, top panel, Figure 2E, and Figure S1). Although the bulkier scaffolds were more potent inhibitors, they did not show enhanced selectivity for the active site, suggesting that the intrinsic reactivity of the vinylsulfonate was obscuring any effects of scaffold structure on binding (Figure 2C,E). We note that off-target covalent labeling is more prevalent in our assays than in the original report of PVSN.\[37\]

This can be attributed to the fact our assays were done at pH 7.5, whereas PVSN and PVS were previously characterized at pH 6. Labeling of PTP1B by compound 2 was greatly attenuated at pH 6 in our assay (Figure 2D, bottom panel). Lower pH allows for differentiation of the active site Cys residue from off-target Cys residues, due to its lower pK\(_a\).[23] but these conditions could limit labeling applications on intact cells. A survey of published reports showing electrophilic labeling of PTPs in cell lysates revealed that most experiments were conducted between pH 5.5 and 6.\[36,39,40\] These observations suggest that selective active site labeling of PTPs at physiological pH will require less reactive molecules with higher affinity for the active site.

**Thiol-reactive groups display a wide range of PTP inhibition.**

We assessed a broad range of thiol-reactive electrophilic functional groups (warheads) for their ability to target the active site cysteine of PTPs (Figure 3A and Table S1). To isolate the effect of the reactive group,
we focused on compounds with a simple phenyl scaffold and assessed phosphatase inhibition using the pNPP dephosphorylation assay described above (Figure 2A). These experiments revealed a general rank-order for the reactivity of different warheads towards PTPs (Figure 3B and Table S2). As expected, there was a wide range of potency across different warheads. Vinylsulfonate (compound 2) was one of the most reactive warheads (Figure 3B). Given the hyper-reactivity of vinylsulfonates observed at physiological pH, we conclude that electrophiles with similar reactivity may also be unsuitable for PTP inhibitor or probe design. Substitution of ester linkages with amides resulted in lower reactivity for all cases (compare 2 to 10, 9 to 12, and 13 to 15), as anticipated.[23] Chloroacetamide (compound 16), which is commonly used in fragment electrophile screening libraries,[29] showed a small but detectable amount of phosphatase inhibition under these conditions (Figure 3B and Table S2). We reasoned that the chloroacetamide warhead would be a good starting point for fragment scaffold screening, and that the electrophile survey presented here could be used to guide late-stage tuning of compound reactivity.

We were surprised to observe drastically different degrees of covalent inhibition across the eight classical PTP catalytic domains included in our experiments, despite the use of a simple phenyl scaffold (Figure 3B). Whereas HePTP, SHP1, and TCPTP were generally inhibited the most by compounds in the phenyl-electrophile series, other PTPs such as CD45, CD148, and YopH were not readily inhibited. These data suggest that the catalytic cysteine residues of individual phosphatases either have different reactivity or accessibility. These phosphatases also showed significantly different intrinsic catalytic activities in the absence of any inhibitor, consistent with previous observations (Figure 3C).[13,41] The susceptibility to covalent inhibition showed a roughly inverse correlation with intrinsic catalytic activity (Figures 3B,C). The underlying molecular basis for this inverse correlation between covalent inhibition and catalytic activity is currently unclear. It could reflect the possibility that the rate-limiting step in the tyrosine phosphatase catalytic cycle is hydrolysis of the phospho-enzyme intermediate, as opposed to the initial nucleophilic attack by the catalytic cysteine residue, as suggested previously (Figure 1B).[42,43] Alternatively, the ligand on-rates may be too slow to efficiently access the active sites of the highly active phosphatases, which have very fast WPD-loop closure rates.[44]

Scaffold exploration reveals global trends in tyrosine phosphatase covalent inhibition.

We next sought to explore the scope of scaffold chemical space that is suitable for covalent inhibition of classical tyrosine phosphatases. We selected chloroacetamide as our electrophilic warhead given its precedent in fragment electrophile screens and because the phenyl species showed sufficiently detectable phosphatase inhibition to build upon without the significant multi-site labeling seen for aryl vinyl sulfonates at physiological pH (Figure S2). We constructed a small library containing 65 chloroacetamide-functionalized fragments bearing diverse aryl and alkyl substituents (Figure 4A and Table S1). These molecules were tested in the same pNPP dephosphorylation assay described above against seven classical human tyrosine phosphatases. Inhibition was assessed over a range of concentrations and time points to obtain time-dependent IC\textsubscript{50} values (Figure 4B, Figure S3, and Table S3,4). For select compounds, the presence of active site labeling in the phosphatase HePTP was corroborated by LC/MS/MS (Figure S4).

The chemically diverse scaffolds in our library are likely to impact both active site binding affinity, as well as intrinsic reactivity of the chloroacetamide warhead. To account for changes in intrinsic reactivity, we measured the rate at which each chloroacetamide reacted with TNB\textsuperscript{2−}, a reduced form of Ellman’s reagent (5,5-dithio-bis-2-nitrobenzoic acid, DTNB) using a previously reported colorimetric assay.[45] The compounds in our library showed a wide range of reactivities, from compounds with no discernable reaction to rates around 1 M\textsuperscript{−1}s\textsuperscript{−1} (Figure 4C and Table S5). We also selected a few compounds spanning the reactivity spectrum and measured their rates of reaction with reduced glutathione (GSH), as GSH is commonly used as a biologically-relevant model thiol for reactivity studies (Figure S5).[46] The relative reactivities of these compounds showed good agreement between GSH and TNB\textsuperscript{2−} (Figure S5D,E). Juxtaposition of enzyme inhibition data from the pNPP dephosphorylation assay with thiol reactivity from the DTNB assay demonstrated that there was no significant correlation between reactivity towards PTPs and intrinsic thiol reactivity (Figure 4D and Figure S6). Thus, we reasoned that we could identify covalent fragments that inhibit PTPs without having excessive intrinsic reactivity that could undermine the downstream development of a drug molecule or chemoproteomic probe.
Figure 4. Exploring diverse chloroacetamide-based fragments for tyrosine phosphatase inhibition. (A) Structures of the core chemotypes found in our chloroacetamide library. For the thiazole category, there is one 4,5-dihydrothiazole. (B) Schematic diagram of the inhibition assay (top) and distributions of HePTP inhibition values for molecules bearing different core chemotypes (bottom). (C) Schematic diagram of the DTNB reactivity assay (top) and distributions of reaction rates with the reduced form of DTNB for molecules bearing different core chemotypes (bottom). (D) Scatterplot juxtaposing representative IC$_{50}$ values for HePTP at one inhibition time point with intrinsic reactivity data obtained using the DTNB assay. (E) Distribution of IC$_{50}$ values across the library for each tyrosine phosphatase after one hour of compound treatment. Phosphatases are ranked as shown in Figure 3C. Compounds denoted “poor fit” had data that could not be fit well to a dose-response curve, due to low degrees of inhibition. (F) Pearson’s correlation coefficients for inhibition datasets of each pair of phosphatases tested against the chloroacetamide library. Correlation coefficients were calculated using IC$_{50}$ values after one hour of compound treatment.
The chloroacetamide fragment survey revealed a series of informative general trends, both with respect to the enzymes and the scaffold structures. As seen in our warhead survey, we observed substantially different degrees of inhibition of each phosphatase that was roughly inversely correlated with intrinsic catalytic activity (Figure 4E). On average across the library, HePTP and SHP1 were inactivated at a faster rate than other phosphatases, and both CD45 and CD148 were hard to inhibit. Notably, we observed unexpected correlations between the inhibition profiles of different phosphatase pairs (Figure 4F). PTP1B and SHP2 displayed the highest correlation, which was surprising, given the close sequence and structural homology between PTP1B and TCPTP and between SHP2 and SHP1. More nuanced enzyme-differentiating trends are discussed in the subsequent sections.

In terms of overall fragment structure, our data revealed that aliphatic chloroacetamides were generally poor inhibitors of all PTPs tested, irrespective of their intrinsic reactivity (Figure 4A-C). Of the aryl chloroacetamides, thiazole-containing compounds, including those with relatively average intrinsic reactivity, showed significant phosphatase inhibition when compared to most other chemotypes (Figure 4A-D and Figure S6). This is particularly noteworthy, given recent reports of thiazole-derived sulfophenyl acetic amides as non-covalent inhibitors of the low molecular weight protein tyrosine phosphatase (LMW-PTP).[47,48] Although the architecture of the LMW-PTP active site is distinct from that of classical PTPs, there are some structurally conserved features, including a catalytic loop and pTyr loop. We compared a crystal structure of LMW-PTP bound to a benzothiazole-containing non-covalent inhibitor[49] with a covalent docking model of PTP1B bound to a benzothiazole-derived chloroacetamide (compound 74). Both proteins engage the benzothiazole in a similar orientation, suggesting that it may be a preferred scaffold for diverse tyrosine phosphatases (Figure S7).

To further confirm that the observed scaffold effects were not exclusively due to changes in reactivity, we selected a few compounds for determination of the rate of inactivation $k_{\text{inact}}$, binding constant $K_i$ (Figure S8), and covalent efficiency ($k_{\text{inact}}/K_i$) with the phosphatase HePTP.[49] Phenyl chloroacetamide (compound 16), which features a simple scaffold and has low reactivity against GSH and TNB,[2], shows low $k_{\text{inact}}$ and high $K_i$ (weak binding), contributing to the low covalent efficiency. On the other hand, the benzothiazole-derived chloroacetamide (compound 74), a potent inhibitor of HePTP with high intrinsic reactivity, shows the expected high $k_{\text{inact}}$ and high covalent efficiency. Notably, while compound 37 shows a low $k_{\text{inact}}$ similar to compound 16, it also has high covalent efficiency arising from a lower $K_i$. These data show how changes in scaffold modulate both $k_{\text{inact}}$ and $K_i$ and illustrate the potential to optimize scaffolds for tighter binding to achieve high covalent efficiency without excessive reactivity.

**Scaffold exploration reveals constraints on fragment structure for tyrosine phosphatase inhibition.**

Most efforts to covalently inhibit tyrosine phosphatases have focused on phenyl-based scaffolds, building on the structure of the endogenous substrate phosphotyrosine.[32] As noted above, our assessment of chloroacetamides suggests that some heterocyclic scaffolds may be preferable over a phenyl scaffold (Figure 4A-C). A comparison of phenyl, benzyl, phenylethyl, and cyclohexyl chloroacetamides showed that an aromatic ring is beneficial near the warhead, mostly likely both for active site binding as well as chloroacetamide activation through inductive effects (Figure 5A). Our data also revealed that the classical PTP active site cannot tolerate steric bulk close to the chloroacetamide warhead. Substitution of the amide nitrogen with large substituents diminished activity in a variety of contexts (Figure 5B and Figure S9A), as did introduction of large functional groups at the ortho-position relative to the chloroacetamide (Figure S9B). For example, we examined a few tertiary alkyl amides that have been used in protein-wide cysteine ligandability studies (compounds 31 and 32) and found that they were not as potent as structurally similar secondary aryl amides, despite their high intrinsic reactivity (Figure 5B).[28,29,50] Recent work in this realm has focused on the development of fragment libraries with even more sterically crowded electrophiles to yield lead fragments with proteome-level selectivity.[51] Our data suggest that the compounds in these libraries may not be optimal for targeting the active sites of tyrosine phosphatases.

We were able to extract other useful structure-activity relationships in relation to previously reported non-covalent inhibitors. For example, in the context of a phenyl scaffold, we found that a carboxy substitution at the 2-position reduced intrinsic reactivity, which could be useful for probe design. However, this substitution did not improve potency (Figure 5C). Previous reports have shown that 2-carboxy substitutions are favorable in unnatural phosphotyrosine-like substrates[52] and non-covalent inhibitors,[53,54] due to an intimate electro-
Figure 5. Structure-activity relationships for covalent tyrosine phosphatase inhibition. (A) Inhibition of PTPs by phenyl, cyclohexyl, benzyl, and phenylethyl chloroacetamide. (B) The effect of chloroacetamide-proximal steric bulk on phosphatase inhibition. (C) The effects of 2-carboxy and 2-hydroxy substituents on covalent inhibition. Known phosphatase inhibitors with the same functional groups are shown. (D) Potency and reactivity effects of substituents on a thiazole scaffold. (E) Biased inhibition of SHP1 by isoxazole chloroacetamides.
static interaction with a conserved lysine residue in PTPs. In the context of covalent inhibition, the significant decrease in chloroacetamide reactivity caused by the 2-carboxy substituent likely negates beneficial binding effects of this appendage. By contrast, a 2-hydroxy substitution both modestly reduced reactivity and enhanced potency (Figure 5C). Non-covalent inhibitors of TCPTP and PTP1B have recently entered the clinical pipeline, and these molecules have a 2-hydroxy substitution on an aryl ring, adjacent to the phosphomimetic “head group” that lies in the same position as the electrophile in our molecules (Figure 5C).\[55,56\]

Our fragment library contains several thiazole-derived molecules that also reveal informative structure-activity relationships (Figure 5D). Compound 65, comprised of a simple thiazole scaffold, shows moderate potency towards PTPs relative to our entire library. However, amongst thiazoles, it was one of the weakest inhibitors of PTPs and showed high intrinsic thiol reactivity. Modifications to this base scaffold generally imparted increased potency against PTPs and in most cases reduced intrinsic thiol reactivity. Benzothiazole (compound 74), was both more reactive and more potent than the unsubstituted thiazole (compound 65). However, further substitution off the benzothiazole decreased intrinsic reactivity, in some cases with mild or even favorable impact on potency against PTPs (Figure S10). As with phenyl-based compounds, aromatic substitutions on the thiazole resulted in increased potency (Figure 5D). Altering the substituent on the coupled aromatic ring further modulated PTP inhibition and thiol reactivity, as seen with compounds 71 and 73. These distal substituents also had an impact on phosphatase selectivity, as exemplified by the biased inhibition of HePTP, and to a lesser extent SHP1, by compound 71, the 4-hydroxy derivative (Figure 5D).

As with thiazole-based compounds, isoxazole-based compounds also offered improvements over phenyl-based compounds. In general, modifications to this base scaffold similarly improved potency against PTPs, particularly with aromatic substitutions from the isoxazole ring. Notably, SHP1 showed a distinct preference for isoxazole-based compounds over analogous thiazole-based compounds when compared to other PTPs (Figure 5E and Figure S11). For example, benzothiazole 74 is preferred over benzisoxazole 39 for every PTP except for SHP1. Moreover, while compounds 37 and 73 perform essentially identically for most other PTPs, compound 37, the isoxazole, is a much more potent inhibitor of SHP1, and compound 36 is the most potent inhibitor of SHP1 (Figure 5E).

During these assays, compound 66, a tetralin-substituted benzothiazole, showed visible aggregation at high concentrations, raising concern that this compound may be partially inhibiting the enzymes by inducing aggregation. To investigate the risk of aggregation, we assessed some compounds with and without the addition of Triton X-100, a nonionic surfactant which has been previously used in enzymatic assays to prevent colloidal aggregation.\[57,58\] When assayed with Triton X-100, compound 66 showed a noticeable reduction in inhibition, though it still showed potent inhibition of PTPs, suggesting that aggregation is partially contributing to the inhibition with this compound. By comparison, compound 74 showed no visible aggregation and likewise retained inhibitory activity towards PTPs in the presence of Triton X-100 (Figure S12). These results indicate that future screening and optimization of this class of inhibitors should include a critical evaluation of misleading aggregate-based inhibition.

Differences in phosphatase covalent inhibition are dictated by residues beyond the active site.

Our experiments have revealed significant differences in susceptibility to covalent inhibition between tyrosine phosphatases, as well as more granular effects of ligand structure on potency and selectivity. We sought to identify divergent structural features of these enzymes that might explain their differences in covalent inhibition. We first used covalent docking to predict plausible binding poses for a few select compounds against four phosphatases: PTP1B, HePTP, SHP1, and SHP2 (Figure S13). Based on these models, we identified residues that made direct contact with the molecules and examined conservation across the panel of phosphatases used in this study. Most first-shell residues were conserved, ruling out these residues as differentiating factors in our analysis (Figure 6A).

As noted above, differences in overall covalent inhibition across the tested phosphatases showed an approximately inverse correlation with the catalytic activity of those phosphatases. Thus, we hypothesize that divergent structural features which affect catalytic activity may also affect covalent inhibition. The WPD loop of phosphatases is the active site feature that has been most directly connected to differences in catalytic activity between phosphatases.\[43,59–63\] Sequence divergence in this region can lead to differences loop dynamics, which in turn impacts catalysis. Two proline residues found in PTP1B (Pro 185 and Pro 188), but not YopH, are reported to constrain the dynamics of the PTP1B WPD loop, potentially leading to its lower
While Pro 185 is invariant in the human PTPs, Pro 188 is replaced by threonine in CD148, and alanine in HePTP, the least and most inhibitable human phosphatases examined in our study, respectively (Figure 6A). Previous studies have shown that the P188A mutation in PTP1B alters WPD loop structure and dynamics and also impacts the rates of multiple steps in the phosphatase catalytic cycle. Based on these observations, we expect that the identity of the P188-analogous residue in other phosphatases is likely to also impact their rates of covalent inhibition.

Our sequence and structure analysis revealed several additional residues that are not conserved across the phosphatases and may impact covalent inhibition. Arg 47 and Asp 48 in PTP1B lie on the pTyr loop, alongside Tyr 46, which forms the conserved hydrophobic floor of the active site (Figure 6C). These two residues diverge across the phosphatases tested and are likely to impact selectivity for the multi-ring molecules in our library (Figure 6A). For example, we observed a distinctive improvement in potency against HePTP for compound 71 (Figure 5D). This molecule has a hydroxy group that is well-positioned to engage the unique Thr 127 on the HePTP pTyr loop (Asp 48 in PTP1B). A previous investigation exploited an electrostatic interaction with Asp 48 to design selective non-covalent inhibitors of PTP1B, as this charged residue is not present in all classical tyrosine phosphatases. We also identified two residues with modest conservation (Leu 110 and Asn 111) that buttress the catalytic cysteine (Cys 215 in PTP1B) (Figure 6A,D). A recent investigation showed that the dynamics of Leu 110 in PTP1B play a role in enzyme catalysis. SHP1 and SHP2 both have a unique Thr residue in place of Leu 110, and this Thr residue hydrogen bonds to the backbone carbonyl preceding the catalytic cysteine in those enzymes. Thus, although Leu 110 and Asn 111 do not make direct contact with substrates or covalent ligands, we hypothesize that changes at these positions will impact the orientation, dynamics, and reactivity of the catalytic cysteine.

---

**Figure 6. Structural features of tyrosine phosphatases that may govern covalent inhibition.** (A) Sequence alignment of the eight tyrosine phosphatases used in this study, highlighting conservation in the loops that line the active site. Residues that directly contact active site ligands are indicated with a black asterisk below the alignment. Six positions with noteworthy sequence divergence, as discussed in the main text, are indicated with a red arrow below the alignment. (B) Potential reactivity- and specificity-defining residues on the WPD loop, mapped onto a model of PTP1B bound to compound 74. (C) Residues on the pTyr loop that can engage the side of covalent fragments that is distal to the electrophile, highlighted on a model of PTP1B bound to compound 73. (D) A modestly conserved buttress for the catalytic cysteine residue on the E loop of PTPs, highlighted on a model of PTP1B bound to compound 74.
Perhaps the most surprising difference between phosphatases observed in our screens was the divergence in ligand engagement by SHP1 and SHP2. The catalytic domains of these enzymes have 59% sequence identity, which increases to 88% identity within 10 Å of the active site, and both proteins share a common domain architecture and regulatory mechanism.® The high degree of structural homology, SHP1 is more readily inhibited by aryl chloroacetamides than SHP2, and it has a unique preference for isoxazole-containing molecules (Figure 5E). Based on docking models and sequence alignments, there are no ligand-proximal residues that differentiate SHP1 and SHP2. The only potentially significant active site difference we identified between the two enzymes is the first residue of the WPD loop (Figure 6A,B, Thr 177 in PTP1B, Leu 415 in SHP1, and Arg 421 in SHP2). A previous study showed that the T177A mutation in PTP1B stabilizes a closed state of the WPD loop but only marginally impacts catalytic activity.® Mutations to the bulkier Leu or Arg found in SHP1 and SHP2 have not been examined, but we anticipate that they will have a substantial impact on loop dynamics and covalent inhibition.

**Conclusions**

A handful of protein tyrosine phosphatases are well-characterized and have been identified as potential therapeutic targets for the treatment of cancer, metabolic diseases, and immune disorders. Nevertheless, the biological roles and regulatory mechanisms of most members of this enzyme family remain poorly characterized. Given their promise as therapeutic targets and the need for more phosphatase-focused discovery biology, there have been extensive efforts to design new chemical tools for tyrosine phosphatases. These can be broadly classified into: (1) non-covalent orthosteric and allosteric inhibitors that serve as both chemical probes and lead compounds for drug discovery,® (2) fluorescent substrates that report on the activity of tyrosine phosphatases in cells,® and (3) covalent probes to profile tyrosine phosphatase activity and regulation in the proteome.® On the one hand, the design of drug molecules and enzyme-specific reporters requires selectivity for individual phosphatases. On the other hand, probes for activity-based profiling of tyrosine phosphatases would ideally be family-wide, and the identification of labeled phosphatases would be enabled by mass spectrometry proteomics. Tackling either of these goals requires a comprehensive view of the small molecule chemotypes that are compatible with tyrosine phosphatases, as well as a deep understanding of the features of tyrosine phosphatases that govern their selective ligandability.

Inspired by a recent resurgence in covalent inhibitors for challenging therapeutic targets, we set out to explore the chemical parameters required for active site covalent inhibition of phosphatases. While covalent ligands for tyrosine phosphatases have been previously identified, most studies have focused on phosphotyrosine isosteres and a handful of reactive warheads.® These past efforts have identified promising molecules but have not yielded significant drug leads or probes that are compatible with mass spectrometry proteomics. As an alternative approach, we explored a broad range of electrophiles and compound structures that deviate from previously characterized chemotypes. Our focused screening yielded a series of guiding principles for future tyrosine phosphatase inhibitor and probe design: (1) we have determined which of the common electrophilic warhead molecules are compatible with tyrosine phosphatase inhibition and identified chloroacetamide as useful starting point for ligand design, (2) our study points to thiazoles and isoxazoles as unique core scaffolds for tyrosine phosphatase inhibitors, and (3) our data show that aliphatic scaffolds and bulky tertiary amide-derived electrophiles have uniquely low potency against all phosphatases that we tested.

A unique feature of our study is the parallel characterization of covalent inhibition for several members of the classical protein tyrosine phosphatase family. This comparative approach revealed unexpected trends in reactivity and selectivity across different phosphatases, even for a relatively small compound library comprised entirely of fragment-like molecules. Most notably, we observed that tyrosine phosphatases with high catalytic activity are difficult to inhibit covalently, and we found that the closely related enzymes SHP1 and SHP2 have distinct fragment preferences. Our study provides new insights into sequence-structure-activity relationships in PTPs by analyzing these enzymes through the lens of covalent inhibition. Altogether, we anticipate that our findings will guide future ligand design efforts and will fuel new biophysical and biochemical investigations into tyrosine phosphatase activity and regulation.
Experimental Section

Expression and purification of tyrosine phosphatases

Constructs bearing the catalytic domains of PTP1B, SHP1, SHP2, HePTP, TCPTP, and CD148 were prepared by over-expression in E. coli (see Supporting Information for protein sequences). All of these constructs had an N-terminal His6-tag followed by a TEV protease cleavage site. For CD45, our construct contained the tandem phosphatase and pseudophosphatase domains and an N-terminal His6-SUMO tag, and for YopH, a tagless version of the full-length protein was expressed. BL21(DE3) cells transformed with a phosphatase-encoding plasmid were grown at 37 °C in 2 L of terrific broth (TB), supplemented with the appropriate antibiotic. After cells reached an optical density at 600 nm of 0.5, isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added for a final concentration of 0.5 mM to induce the expression of proteins, and the cultures were incubated at 18 °C overnight. Cells were harvested by centrifugation (7800 rcf at 4 °C for 30 min), resuspended in lysis buffer (pH 8, 50 mM Tris, 300 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 (33,000 rcf at 4 °C for 45 minutes), the supernatant was applied to a 5 mL HiTrap Ni-NTA column (Cytiva). The resin was washed with lysis buffer and wash buffer (50 mM Tris, pH 8.5, 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) and passed through a 5 mL HiTrap Q anion exchange column. The His6-tag of the collected fractions was cleaved by incubating with TEV protease or Ulp1 (SUMO protease) overnight. The reaction mixture was subsequently flowed through 2 mL of Ni-NTA resin (ThermoFisher). The cleaved protein was collected in the flow-through and washed. Proteins were further purified using a Superdex 75 or 200 16/600 gel filtration column (Cytiva) in 10 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM TCEP, 10 % glycerol). Pure fractions were pooled then concentrated with centrifugal filters (Millipore). The resulting solution was aliquoted, and flash frozen in liquid N2 for long-term storage at -80 °C. For YopH, which was tagless, the protein was purified by cation exchange of the soluble fraction of the cell lysate over a 5 mL HiTrap SP column (Cytiva), followed by size-exclusion chromatography as described for the other phosphatases.

Chemical synthesis of electrophilic compounds

Detailed experimental procedures for the synthesis of compounds is provided in the Supporting Information, along with NMR and MS characterization data.

Intact protein mass spectrometry

Each enzyme was diluted in pH 7.5 Tris buffer to achieve a final concentration 500 nM after addition of compound. Enzymes were pretreated by adding 2 µL of compound stock to 198 µL of enzyme solution. Unless otherwise specified, the final concentration of each compound was 100 µM. After one hour, the mixture was injected onto a BEH C8 column (Waters) on a UPLC-MS system (Xevo QToF, Waters). Reverse-phase liquid chromatography was carried out with gradient of 5% to 95% of acetonitril (with 0.02% formic acid) for 8.5 min. The protein typically eluted around 4 min; this peak on the chromatogram was integrated and deconvoluted using the 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 ± 1 Da. The abundance of each detectable enzyme species was normalized to the total intensity of all enzyme species.

Analysis of phosphatase activity using a colorimetric assay

Phosphatase catalytic activity was measured by monitoring dephosphorylation of p-Nitrophenyl phosphate (pNPP). All assays were performed in Tris buffer containing 50 mM Tris adjusted to an ionic strength of 150 mM with NaCl at pH 7.5. All measurements were taken at 30 ºC. Each enzyme was diluted to 250 nM in Tris buffer. Enzymes were pretreated with compound: for each enzyme, an aliquot (2.2 µL) of each compound in DMSO stock was dispensed into a well of a 96-well plate. A DMSO group was included to serve as an uninhibited control. Subsequently, enzyme solution (217.8 µL, final DMSO concentration: 1% v/v) was dispensed and mixed in each well. Similarly, a well was prepared with DMSO and Tris buffer without enzyme to serve as a substrate-only control. At 1, 2, and 4 hour time points, the pretreated enzyme solutions were aliquot (20 µL) in triplicate to a 96-well plate, and pNPP solution (180 µL) 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 20M during measurement. For CD148 only, this solution was diluted 10x for a final pNPP Hong, Xi, et al., 2023 – Main Text & Figures - page 14 of 20
concentration of 2 mM. Absorbance measurements at 405 nm were acquired every 30 seconds for 10 min in a microplate reader. The data were fit to a linear regression to obtain an initial rate from the slope, and the slope of the substrate-only control was subtracted from each group. The residual activity of each enzyme after incubation with compound was calculated from the slope of the compound-treated group as a percentage of the slope of the uninhibited group. pNPP dephosphorylation assays were conducted against each enzyme for each compound across 6 concentrations (3.3 µM, 10 µM, 33 µM, 100 µM, 330 µM, 1 mM). The IC50 value for the inhibition of phosphatase activity was determined by fitting the plot of residual enzyme activity against the inhibitor concentrations using a four-parameter dose-response nonlinear regression in GraphPad Prism.[70]

Analysis of K_I and k_inact for phosphatase inhibition using a colorimetric assay

Phosphatase catalytic activity was measured using p-nitrophenyl phosphate (pNPP) as described above. Enzymes were treated with compound in DMSO stock (2.4 µL stock in 237.6 µL enzyme, final DMSO concentration: 1% v/v). Concentration ranges were selected per compound to cover the range of measurable inhibition in a 1 hour period. The percent inhibition was determined as described above at each time point and the inhibition over time was fit to a first-order rate equation at each inhibitor concentration to obtain an observed first-order rate constant, k_{obs}. Six time points were collected over the course of 1 hour. The double reciprocal plot of k_{obs} versus inhibitor concentration was then fit to the following equation to obtain K_I and k_{inact}:

$$\frac{1}{k_{obs}} = \frac{K_I + [I]}{k_{inact}[I]} = \frac{K_I}{k_{inact}[I]} + \frac{1}{k_{inact}}$$

Analysis of compound intrinsic reactivity against TNB^2- in a colorimetric assay

Thiol reactivity assays were carried out based on a previously reported assay by Resnick et al.[45] All assays were performed in buffer containing 20 mM sodium phosphates and 150 mM NaCl at pH 7.4. All measurements were taken at 37 °C. 50 µM DTNB and 200 µM TCEP were incubated in sodium phosphate buffer for 5 min at 37 °C to obtain TNB^2-. Separately, 200 µM TCEP was incubated without DTNB in sodium phosphate buffer for 5 min at 37 °C. For each compound, TNB^2- solution (73.5 µL) was treated with 200 µM compound (1.5 µL of 10 mM stock in DMSO) in triplicate in a 384-well plate. Additionally, TCEP solution without TNB^2- was treated with 200 µM compound in triplicate to control for background absorbance. A DMSO group was included to serve as a negative control. UV absorbance measurements were acquired every 7.5 min for 7 hours at 412 nm on a microplate reader. For each compound, the background absorbance without TNB^2- was subtracted from the absorbance with TNB^2-. The initial inhibitor concentration, [I]_0, and TNB^2- concentration, [TNB^2-]_0, were 200 µM and 100 µM, respectively. The remaining concentrations [I] and [TNB^2-] were determined as a function of time from absorbance data using absorbance of the DMSO group as a reference for 100 µM TNB^2-. Subtracted reaction progress curves were directly fit to the second-order rate equation below, using GraphPad Prism, to extract a second-order rate constant:[70,71]

$$[\text{TNB}^2-]_0 = \frac{[\text{TNB}^2-]_0([I]_0 - [\text{TNB}^2-]_0)}{[I]_0 e^{([I]_0 - [\text{TNB}^2-]_0)kt} - [\text{TNB}^2-]_0}$$

In the original report for this assay, data were fit by linear least squares to obtain the reaction rate constant, where k([I]_0-[TNB^2-]) is the slope of ln([I][TNB^2-][I]_0) plot versus time.[45] We also analyzed our data using this method, and the rate constants from both analyses were in good agreement (Table S5).

Analysis of compound intrinsic reactivity against glutathione in an RP-HPLC assay

All assays were performed in Dulbecco’s phosphate buffered saline. GSH solution was prepared with 1 mM TCEP for a final GSH concentration of 2 mM after compound treatment. For each compound, GSH solution (396 µL) was treated with 100 µM compound (4 µL of 10 mM stock in DMSO). Reaction progress was monitored by reverse-phase high performance liquid chromatography (RP-HPLC) using an Agilent 1200 series, monitoring absorbance at 280 nm or 260 nm. At each time point, 50 µL of the reaction mixture was injected onto a C18 column (ZORBAX 300SB-C18, 5 µm, 4.6 x 150 mm). The samples were eluted using water with 0.1% trifluoroacetic acid (solvent A) and acetonitrile with 0.1% trifluoroacetic acid (solvent B) over the following linear gradients: 0-2 minutes: 5% B, 2-12 minutes: 5-95% B, 12-13 minutes: 95% B, 13-14 minutes: 95-5% B, and 14-18 minutes: 5% B. For each reaction, six time points were analyzed over the course of 6-12 hours. Control reactions were run with 100 µM compound in 1 mM TCEP without GSH to identify

Hong, Xi, et al., 2023 – Main Text & Figures - page 15 of 20
peaks corresponding to degradation products (Figure S5A). The peak areas corresponding to the compound, the compound + GSH adduct, and the degradation product were calculated using the Agilent OpenLAB ChemStation software. The fractional peak areas were used to determine the concentrations of [I+GSH] and [I]. The initial inhibitor concentration, [I]_0, and GSH concentration, [GSH]_0, were 100 µM and 2 mM, respectively. The remaining concentrations of [GSH] and [I] were assumed to be equal to [GSH]_0 – [I+GSH], [I]_0 – [I+GSH], respectively, disregarding the degradation product. The data were fit by linear least squares to obtain the reaction rate constant, where k([GSH]_0-[I]_0) is the slope of ln([GSH]_0/[GSH]_0[I]) plot versus time.

\[
\ln \left( \frac{[GSH]_0}{[GSH]_0[I]} \right) = k([GSH]_0 - [I]_0) * t
\]

Analysis of active site Cys labeling in HePTP by LC-MS/MS

HePTP was diluted in a buffer containing 50 mM Tris, pH 7.5, and 150 mM NaCl. For each sample, HePTP was treated with compound in DMSO stock for a final concentration of 1 µM HePTP and 100 µM compound (6 µL stock in 594 µL enzyme, final DMSO concentration: 1% v/v) and incubated for 1 hour at room temperature. Then, the reactions were quenched with 10 mM DTT, and the samples were precipitated with methanol:chloroform:water (4:1:4). The resulting pellets were solubilized in a buffer containing 50 mM Tris, pH 7.5, 9 M urea, and 10 mM DTT, and incubated for 30 minutes while shaking at room temperature. Free cysteines were subsequently alkylated with 20 mM iodoacetamide in the dark for 45 minutes while shaking at room temperature. The samples were diluted to 1 M urea in 50 mM Tris pH 7.5 and digested with Promega sequencing grade modified trypsin (2 µg). After overnight digestion, samples were acidified with formic acid to a final concentration of 1% (v/v), desalted on C18 StageTips according to Rappsilber et al.,[72] evaporated to dryness in a vacuum concentrator, and reconstituted in 15 µL of 3% acetonitrile / 0.2% formic acid.

For each sample, about 1 µg of total tryptic peptides were analyzed on a Waters M-Class UPLC using a 15cm Thermo EASY-spray column (75µm inner diameter; 3µm particle size; p/n ES900) coupled to a benchtop Thermo Fisher Scientific Orbitrap Q Exactive HF mass spectrometer. Peptides were separated at a flow rate of 400 nL/min over a 65 min gradient, including sample loading and column equilibration times. Data was acquired in data independent mode using Xcalibur 4.5 software. All raw data were analyzed with Spectronaut software version 17.2.[73] using directDIA analysis methodology and searched against Uniprot P35236 (HePTP). Oxidation of methionine, protein N-terminal acetylation, carbamidomethylation on cysteines, and modification by the treated compound on cysteines were set as variable modifications. Trypsin/P was set as the digestion enzyme.

Covalent docking of compounds to tyrosine phosphatase

All docking was performed in Maestro by Schrödinger,[74] according to the “Covalent Docking for Virtual Screening and Pose Prediction” and “Introduction to Structure Preparation and Visualization” tutorials published by Schrödinger.[75] Briefly, each compound was prepared using LigPrep to generate possible tautomers, conformations, and ionization states between pH 5-9.[76] Structures for PTP1B (1KAK), SHP1 (4HJQ), SHP2 (6CMQ), and HePTP (3D42) were downloaded from the Protein Data Bank.[77] These structures were prepared using the Protein Preparation Wizard[78] to optimize charge, add hydrogens, and remove water molecules. All structures were aligned to PTP1B (1KAK), and the receptor grid box for docking calculations was centered on the co-crystallized ligand in 1KAK. Covalently docked complexes were generated using the CovDock module in the Pose Prediction docking mode, and the top scoring complex for each compound was selected as a representative pose for further structural analysis.[76]

Acknowledgements

We thank members of the Shah and Jovanovic labs for their guidance throughout this project, especially Anne van Vlimmeren for her constructive feedback on experimental design. We also thank Fereshteh Zandkarimi and Brandon Fowler from the Columbia Chemistry mass spectrometry facility for their assistance with mass spectrometry and Daniel Keedy for critically reading the manuscript. Bacterial expression vectors for HePTP and SHP2 were gifts from Nicola Burgess-Brown (Addgene plasmid #s 38945 and 38965). Plasmids encoding PTP1B, TCPTP, and SHP1 were gifts from Pau Creixell. Plasmids encoding CD45 and CD148 were gifts from Arthur Weiss. This work was funded NIH grant R35 GM128014 awarded to Hong, Xi, et al., 2023 – Main Text & Figures - page 16 of 20
NHS. M.J. is funded by the NIH (R35GM128802; R01AG071869 and R01HG012216) and NSF (Award 2224211). Experiments involving cryoprobe NMR data were supported by funding from the NIH Office of the Director (award # S10OD026749). SYX is supported by a Barry Goldwater Scholarship. LCT is supported by an NSF Graduate Research Fellowship (award # 2036197). CAC is supported by an NSF Graduate Research Fellowship.

Conflict of Interest
The authors declare no conflict of interest.

Supporting Information
Supporting information for this manuscript includes the following: (1) Amino acid sequences of all phosphatase constructs used in this study, (2) a table containing the structures, SMILES strings, and vendor/catalog number (when applicable) for all compounds used in this study, (3) complete synthetic methods, and (4) supplementary data figures and tables.

References

Hong, Xi, et al., 2023 – Main Text & Figures - page 17 of 20


Hong, Xi et al., 2023 – Main Text & Figures - page 18 of 20


[70] D. Radushev, *GraphPad Prism*, GraphPad Software, LLC., **2022**.


