Published in final edited form as:

J Am Chem Soc. 2018 January 10; 140(1): 200–210. doi:10.1021/jacs.7b08366.

"Inverse Drug Discovery" strategy to identify proteins that are targeted by latent electrophiles – exemplified by arylfluorosulfates

David E. Mortenson[1,2], **Gabriel J. Brighty**[1,2], **Lars Plate**[1,2], **Grant Bare**[2], **Wentao Chen**[1,2], **Suhua Li**[2], **Hua Wang**[2], **Benjamin F. Cravatt**[1,2,4], **Stefano Forli**[3], **Evan T. Powers**[2], **K. Barry Sharpless**[2,4,*], **Ian A. Wilson**[3,4,*], and **Jeffery W. Kelly**[1,2,4,*]

[1]Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA

[2]Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037, USA

[3]Department of Integrative, Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[4]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

Abstract

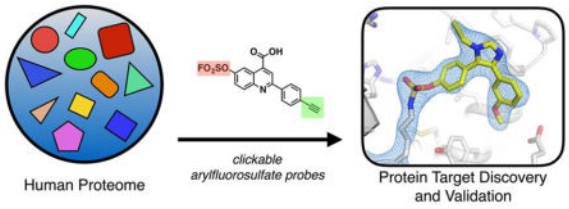
Drug candidates are generally discovered using biochemical screens employing an isolated target protein or by utilizing cell-based phenotypic assays. Both non-covalent and covalent hits emerge from such endeavors. Herein, we exemplify an "Inverse Drug Discovery" strategy in which organic compounds of intermediate complexity harboring weak, but activatable, electrophiles are matched with the protein(s) they react with in cells or cell lysate. An alkyne substructure in each candidate small molecule enables affinity chromatography–mass spectrometry, which produces a list of proteins that each distinct compound reacts with. A notable feature of this approach is that it is agnostic with respect to the cellular proteins targeted. To illustrate this strategy we employed arylfluorosulfates, an underexplored class of sulfur(VI) halides, that are generally unreactive unless activated by protein binding. Reversible arylfluorosulfate binding, correct juxtaposition of protein side chain functional groups, and transition state stabilization of the S(VI) exchange reaction all seem to be critical for conjugate formation. The arylfluorosulfates studied thus far exhibit chemoselective reactivity toward Lys and, particularly, Tyr side chains, and can be used to target non-enzymes (e.g., a hormone carrier or a small-molecule carrier protein) as well as enzymes. The "Inverse Drug Discovery" strategy should be particularly attractive as a means to explore latent electrophiles not typically used in medicinal chemistry efforts, until one reacts with a protein target of exceptional interest. Structure-activity data can then be used to enhance the selectivity of conjugate formation or the covalent probe can be used as a competitor to develop

*Corresponding authors: Professor Jeffery W. Kelly: jkelly@scripps.edu, 858-784-9880, Professor Ian A. Wilson: wilson@scripps.edu, 858-784-9706, Professor K. Barry Sharpless: sharples@scripps.edu, 858-784-8945.

The x-ray structures of the GSTO1-**1c** and NME1-**2** conjugates have been deposited in the PDB under accession codes 5UEH and 5UI4, respectively. Supporting Information containing experimental protocols, supplemental figures, and compound characterization data is available free of charge on the ACS publications website. Full tandem mass spectrometry data are available as a supporting table (XLSX format).

non-covalent drug candidates. Here we use the “Inverse Drug Discovery” platform to identify and validate covalent ligands for 11 different human proteins. In the case of 2 of these proteins, we have identified and validated ligands for the first time.

Graphical Abstract



INTRODUCTION

The majority of drugs modulate function by binding reversibly to proteins. However, a small percentage of regulatory agency-approved drugs form irreversible, covalent linkages with their target proteins, and only recently has this class of compounds been actively sought by drug companies.[1–2] The historic reluctance to seek covalent drug candidates stems from concerns that off-target reactivity could drive toxicity. In many cases, the covalent nature of approved drugs, such as clopidogrel (Plavix), was appreciated only after their medical usefulness had been demonstrated[3], although recently, fit-for-purpose covalent protein modifiers have been established (e.g., ibrutinib (Imbruvica) for B cell cancers).[4] Latent electrophiles, functional groups that become activated toward covalent bond formation only upon binding to a specific protein(s), are particularly desirable for drug development because they minimize off-target reactivity-linked toxicity.

A small number of latent electrophiles are known that form a covalent linkage with a specific protein, but are otherwise unreactive in the context of a cell.[5–9] Arylfluorosulfates are interesting candidates in this regard[10], as they are essentially unreactive toward the vast majority of the proteome. We have shown previously that structurally simple arylfluorosulfates react efficiently with a few intracellular lipid binding proteins (iLBPs), whose binding sites present a specific constellation of side-chain functional groups capable of catalyzing the sulfur-fluoride exchange (SuFEx[10]) reaction with a binding site Tyr residue.[11] iLBPs deliver hydrophobic small molecules to nuclear hormone receptors, activating their transcription factor functions.[12–13] Thus, formation of covalent arylfluorosulfate-iLBP conjugates can alter transcription by blocking ligand-mediated transcription factor activation. For example, a biphenyl arylfluorosulfate reacts with and inhibits the iLBP, cellular retinoic acid binding protein 2 (CRABP2)[11], attenuating the expression of genes downstream of the nuclear hormone receptor activated by retinoic acid.[14] The Tyr reactivity of the biphenyl arylfluorosulfate toward CRABP2 depends on two adjacent Arg side chains, which likely lower the pKa of the phenol and possibly stabilize the reaction transition state of the SuFEx reaction. Critically, unfolded CRABP2 does not react with arylfluorosulfates, indicating that the native protein structure is required to facilitate conjugate formation. Proteins must meet highly stringent criteria to undergo a SuFEx reaction with arylfluorosulfates, by facilitating abstraction of a fluoride ion leaving group

from the strong S(VI)-F bond. The low reactivity of this bond in nearly all contexts makes it an attractive platform for the selective covalent modification of proteins.[10]

Building on our earlier work with simpler, aromatic fragment-like structures attached to the fluorosulfate functional group that react principally with intracellular lipid binding proteins (iLBPs), here we describe the "Inverse Drug Discovery" approach and exemplify this strategy with three different structures of intermediate complexity (or lead-like aromatic structures) attached to the fluorosulfate functional group. In contrast to conventional drug discovery efforts, which involve screening up to ~1 million small molecules for their ability to modulate the function of an isolated protein or a cellular phenotype, the "Inverse Drug Discovery" approach involves treating a cell or a cell lysate with a small molecule of intermediate structural complexity harboring a latent electrophilic functional group to identify the protein(s) that it reacts with using an affinity purification mass spectrometry strategy. Importantly, this method is agnostic with respect to the protein(s) targeted by the small molecule or the cellular phenotypes that may be altered. The principal goal is to identify and validate a diverse set of proteins that efficiently form covalent conjugates with a given small molecule of intermediate structural complexity that harbors an activatable, latent electrophile. Identifying the protein(s) targeted by latent electrophiles not typically used in medicinal chemistry attached to lead-like structures should provide fundamental insights into the structural and chemical requirements to activate the functional group of interest. This "Inverse Drug Discovery" strategy requires medicinal chemistry to convert lead-like small molecules into drug candidates that selectively target proteins of exceptional interest, either covalently or non-covalently.

Previous studies that exhibit some of the features of the "Inverse Drug Discovery" approach have led to discovery of proteins that react covalently with sulfonate esters[15–16], sulfonyl fluorides[17–19], aryl halides[8, 20], isothiocyanates[21], and terminal alkynes[7], although not all of these efforts were agnostic with respect to the protein(s) being targeted. Moreover, in some of the cited examples, the functional groups being employed are more electrophilic and broadly reactive toward the proteome than we believe would be ideal for the "Inverse Drug Discovery" strategy. Notably, the "Inverse Drug Discovery" approach has features that distinguish it from activity-based protein profiling (ABPP) efforts, which typically employ more-reactive covalent probes that are directed toward a specific class of enzymes[22–25] or a large number of proteins containing a reactive amino acid side chain like that of Cys.[26]

In this paper, we match structurally unique arylfluorosulfates of intermediate complexity with the protein(s) that they react with from the human proteome in the context of cells or cell lysate. Reversible protein binding directs the latent S(VI) electrophile to react with cellular proteins that possess the appropriate constellation of functional groups capable of catalyzing the SuFEx reaction, thus it is important that the non-reactive organic substructure exhibit intermediate chemical complexity. The arylfluorosulfates employed always incorporate an alkyne functionality, which facilitates affinity chromatography–mass spectrometric analysis. This methodology enables identification of the proteins that react with each structurally distinct arylfluorosulfate.

We find that each arylfluorosulfate probe examined in this work reacts with a distinct group of human proteins. X-ray crystallographic and/or mass spectrometry data reveal that these proteins are conjugated to the probes through Tyr or Lys side chains in proximity to cationic side chains, providing new insights into the structural basis for the activatable reactivity of arylfluorosulfates. Many of the proteins identified by the "Inverse Drug Discovery" approach are relevant to human disease and based on the findings reported here, can now be targeted for covalent inhibition using arylfluorosulfate-based probes.

RESULTS

Synthesis of Arylfluorosulfate Probes of Intermediate Structural Complexity

Given that structurally simple or fragment-like arylfluorosulfates (Figure 1a) tend to react preferentially with iLBPs[11] in human cells, we prepared a new set of arylfluorosulfates of intermediate structural complexity (i.e., lead-like) featuring distinct shapes, positioning of H-bond donors and acceptors, distinct charge distributions, etc., (see Table S1) to discern whether different human proteins can be targeted. We envision that reversible protein binding will precede and facilitate the subsequent conjugation reaction, hence the importance of an organic substructure of intermediate structural complexity. The properties of these probes do adhere to Lipinksi's Rule of 5, however beyond that we did not assess metabolic stability, chemical stability, etc. We favored these probes because they were accessible using one-pot reactions that simultaneously introduce the alkyne and phenol functional groups into a moderately complex organic scaffold. The latter functional group is readily converted to an arylfluorosulfate via published conditions.[10] Arylfluorosulfate probe **1** comprises a quinolinic acid, whereas probes **2** and **3** are based on imidazoles featuring different aryl substituents (Figure 1a). All three molecules contain the latent activatable arylfluorosulfate electrophile, as well as an alkyne functional group to be used for the copper-catalyzed azide-alkyne cycloaddition (CuAAC) click reaction[27–29] with biotin, enabling affinity chromatography-mass spectrometric analysis. Alternatively, the presence of the alkyne in the conjugate allows fluorophore attachment for SDS-PAGE fluorescence visualization. We anticipated that the structural diversity amongst scaffolds **1** – **3** would allow for equilibrium binding to and subsequent reaction with distinct sets of human proteins.

To provide insight into which proteins react quantitatively with a given arylfluorosulfate, we prepared molecules **1c**, **2c** and **3c** that are structurally analogous to probes **1**, **2** and **3**, respectively, but lack the alkyne functional group (Figure 1a). When competitors **1c**, **2c** or **3c** are present in excess of the concentration of probes **1**, **2** or **3**, they should block reaction between target proteins and the alkyne probes **1**, **2** or **3**, leading to less fluorescence signal *via* SDS-PAGE or lowered protein abundance following affinity purification and tandem mass spectrometry (MS-MS) analysis. These competition experiments are important because they distinguish between a low-abundance protein that reacts quantitatively with **1**, **2** or **3** and the undesirable situation wherein a highly-abundant protein reacts only fractionally with **1, 2** or **3**.[27] In the former case, the alkyne-lacking arylfluorosulfates **1c**, **2c** or **3c** prevent conjugation with **1**, **2** or **3**, respectively. In the latter case, competitors **1c**, **2c** or **3c** do not

block or only modestly diminish conjugate formation between **1**, **2** or **3** and a highly abundant protein exhibiting fractional reactivity.

Identifying Proteins in the Human Proteome that are Reactive toward Arylfluorosulfates 1, 2 or 3

We treated HEK293T cell lysate with probes **1**, **2** or **3** (10 μM) in the absence or the presence of a 9-fold molar excess of their respective competitor probes **1c**, **2c** or **3c** (90 μM) for 24 h. A CuAAC reaction[28–30] was then performed to attach the azide-conjugated fluorophore, tetramethylrhodamine azide (TMR-N_3). SDS-PAGE visualized by fluorescence was used to discern whether any proteins in the proteome abundant enough to be detected by this method form conjugates with probes **1**, **2** or **3** (Figures 1b, S1a). In the presence of competitors **1c**, **2c** or **3c**, some of the bands in the cell lysate disappeared (Figure 1b, red arrows), indicating that the alkyne probes engaged certain proteins substantially. The same is true in living HEK293T cells treated with probes **2** or **3** and their respective competitors (Figures 1c, S1b). Generally, more proteins were labeled in cell lysate with probes **1** or **2** or **3** relative to live cell treatment (cf. Figure 1b and 1c). For the remainder of this manuscript, labeling experiments were carried out in cell lysate, to facilitate discovery of a maximum number of protein targets for further scrutiny. Although few gel bands were visibly diminished in intensity upon addition of the competitor **3c** to the cell lysate reaction with probe **3**, this result does not imply that probe **3** was not forming quantitative conjugates with members of the human proteome. Previous activity-based protein profiling studies reveal that SDS-PAGE-based proteome profiling is inherently biased toward abundant proteins, while low-abundance proteins are less likely to be visible above background.[31]

Quantitative Proteomics Identifies Human Proteins Reactive Toward Probes 1, 2 or 3

To identify the proteins in the human proteome that formed conjugates with probes **1**, **2** or **3** over a 24 h reaction period, a cleavable biotin azide (diazo biotin azide) was clicked onto the protein-probe conjugates using the CuAAC click reaction[28, 30] to enable biotin affinity chromatography of the conjugates, followed by MS-MS analysis.[29] The relative abundance of protein-probe conjugates enriched from each reaction condition was evaluated by mass spectrometry, using six-plex labeling with tandem mass tags (TMTs).[32] Biotin-modified and affinity-isolated protein conjugates were digested with trypsin and the resulting peptides were reacted with one of six unique isobaric TMT reagents (Figure 2a). Using a different TMT tag for each reaction condition (e.g., for **1**, **1+1c**, and **1c**) with human cell lysate, in duplicate reactions, enables quantitative comparisons of protein abundance between the three different treatment conditions (Figure 2b). Protein reactions with probes **1**, **2** or **3** that are strongly attenuated by the addition of competitors **1c**, **2c** or **3c**, respectively, were of highest interest. *Competition ratios* (Equation 1) were generated by comparing the abundance of a particular protein conjugate in an alkyne probe-only experiment (i.e., **1**, **2**, or **3**) to the abundance of that protein conjugate in the experiment where competitors **1c**, **2c** or **3c** are present at 9x the concentration of probe **1**, **2** or **3**. The *Enrichment Ratio* (Equation 2) afforded by comparing protein abundance in the alkyne-only (e.g., **1**) and competitor-only (e.g., **1c**) experiments was used to assess the efficiency of target protein enrichment. A high protein abundance when the competitor probes **1c**, **2c** or **3c** are used alone indicates non-

specific enrichment in our affinity purification protocol, since these competitor probes are not competent as affinity handles *via* the CuAAC reaction.

$$\text{Competition Ratio} = \frac{\text{Protein Abundance(alkyne probe)}}{\text{Protein Abundance(alkyne + competitor probe)}} \quad \text{Eq. 1}$$

$$\text{Enrichment Ratio} = \frac{\text{Protein Abundance(alkyne probe)}}{\text{Protein Abundance(competitor probe)}} \quad \text{Eq. 2}$$

We calculated the Competition Ratio (e.g., **1/(1+1c)**; Equation 1) and the Enrichment Ratio (e.g., **1/1c;** Equation 2) for each probe pair. The Competition Ratio was then employed to produce rank-ordered lists of the protein conjugates formed with each probe. We focused our validation efforts primarily on protein conjugates exhibiting the lowest p-values and highest ranks in the Competition Ratio experiment (Figure 2c; see Tables S2–S4 for representative data). The Enrichment Ratio was calculated to be sure that the affinity purification step was performing as expected, i.e., that the conjugate harboring an alkyne was enriched whereas the conjugate lacking the alkyne was not. Generally, those proteins exhibiting the highest ranks and low p-values in the Competition Ratio experiment (e.g., **1/(1+1c)**) also ranked prominently in Enrichment Ratio experiments (e.g., **1/1c**). However, comparison of the two ratios (e.g., **1/(1+1c)** *vs.* **1/1c**) for some protein targets indicated substantial differences. CRABP2, a validated arylfluorosulfate-reactive protein[11], displayed an enrichment ratio of 4.02 (**1/1c**; p-value: 0.048), however the competition ratio (i.e., **1/(1+1c)**) was only 1.04 (p-value: 0.68), suggesting that only a small fraction of this protein is conjugated by probe **1** in cell lysate. Maximal Enrichment Ratios of 4.02, 14.29, and 6.40 were determined for **1/1c**, **2/2c**, and **3/3c**, respectively, using data normalized with respect to total protein abundance. The relatively low values observed for **1/1c** and **3/3c** may arise due to a combination of the data normalization procedure, which assumes equal protein abundance across all TMT channels, and TMT ratio compression, which has been documented previously.[33] While TMT channel normalization allows for correction of systematic differences in protein abundance between replicate measurements, this procedure also leads to Enrichment Ratio compression. Notably, these effects do not interfere with the identification of protein targets for these sets of probes, which were validated with recombinant proteins in 11 cases.

Selection and Validation of Fluorosulfate-reactive Proteins Identified via Proteomics

To scrutinize the proteome-wide affinity purification MS-MS strategy for “Inverse Drug Discovery”, as well as to provide a structural basis for understanding arylfluorosulfate protein reactivity, we focused our efforts on validating twelve different proteins for which structures are already deposited in the Protein Data Bank[34] (Figure S2). Reactions between purified recombinant proteins and the respective reactive probes (i.e., **1**, **2** or **3**) were evaluated by LC-ESI-MS to establish whether labeling occurred at a single, specific site, and to estimate the extent of recombinant protein labeling. For eleven of the twelve proteins that we examined, partial or complete labeling at a specific site(s) was established. For one

protein target that we evaluated, reaction between the recombinant protein and the probe used to identify it in cell lysate was not observed in buffer. This finding may highlight the sensitivity of our method to a required protein post-translational modification or other factors that are unique to a cellular context to enable the conjugation reaction. In addition to LC-ESI-MS analysis, we used mutagenesis experiments, tandem mass spectrometry analysis (MS-MS), x-ray crystallography and reactive docking methods to identify specific arylfluorosulfate-reactive residues in validated protein targets.

Validation of Probe 1 Reactive Protein Targets

The probe **1** Competition Ratio experiment (i.e., **1/(1+1c)**) revealed that two classes of glutathione S-transferase (GST), omega and pi (GSTO1 and GSTP1, respectively), react substantially with probe **1** in HEK293T cell lysate (Figure 2c, left panel). The Competition Ratio experiment also revealed strong enrichment of biliverdin reductase A (*BLVRA*). Less enriched, but also showing significantly elevated protein abundance in the probe **1** Competition Ratio experiment are conjugates with hydroxysteroid dehydrogenase-like protein 2 (HSDL2), fructose 2,6-biphosphatase TIGAR, RNA demethylase ALKBH5, and FTO (ALKBH9), a homolog of ALKBH5[35] (Figure 2c, left panel).

Since GSTP1 also reacted with probe **2** (*vide infra*; Figure 2c, middle panel), we examined the reactivity of recombinant GSTP1 toward both arylfluorosulfate probes **1** and **2**. Initial labeling experiments monitored by LC-ESI-MS indicated formation of singly-labeled GSTP1 by either probe (Figures S3, S4). A previous study identified Tyr108 as the primary site of covalent conjugate formation in mouse GSTP1 using probes bearing a sulfonyl fluoride[18], a functional group that is related to the arylfluorosulfate[11], but substantially more reactive and more prone to hydrolysis.[36] Adjacent to Tyr108 in the GSTP1 binding pocket is Tyr7, which serves as a catalytic base in glutathione conjugation to electrophilic xenobiotics.[37] In the labeling reactions with probes **1** or **2**, probe **2** exhibited more efficient reaction with GSTP1 than probe **1** and was used for the SDS-PAGE-based Tyr→Phe mutagenesis labeling experiments. We confirmed that both Tyr7 and Tyr108 are capable of reaction with arylfluorosulfate **2**, as mutation of either Tyr individually to Phe maintained some degree of labeling (Figure 3a). These gel-based experiments indicate that covalent conjugate formation likely occurs primarily *via* Tyr7, as the Y7F mutation led to a greater attenuation in GSTP1 labeling than did the Y108F mutation. Mutation of both Tyr residues (Tyr7 and Tyr108) to Phe eliminates GSTP1 labeling. We also examined the role of the cationic side chain adjacent to Tyr7 and Tyr108: while GSTP1 harboring the R13K mutation maintained native-like labeling efficiency, R13Q strongly attenuated labeling (Figure 3b). This finding indicates that a nearby cation is likely to be essential for fluorosulfate activation and/or Tyr pKa perturbation. Critically, GSTP1 labeling in all cases was strongly diminished by pre-boiling of the enzyme in buffer containing 0.1% SDS, indicating that the native protein structure is required to facilitate arylfluorosulfate conjugate formation. Consistent with the results of our mutagenesis studies, we found strong evidence from the MS-MS experiment for labeling of GSTP1 by probes **1** and **2** at Tyr7, indicating that this residue is the primary site of labeling (Figures 4a, S5, S6). We also found that treatment of GSTP1 with either probe **1** or **2** attenuates its activity (Figure S7).

We next examined the labeling of the six other human proteins identified in the probe **1** Competition Ratio experiment (GSTO1, BLVRA, TIGAR, HSDL2, ALKBH5 and FTO) by LC-ESI-MS and MS-MS analysis. For five of these targets (all but FTO), labeling by probe **1** was established by LC-ESI-MS (Figures S8–S12). Labeling sites were then established *via* MS-MS analysis (Figures 4b–4e, S13–S16). Consistent with our assignment of probe **1** covalent labeling in the active site of BLVRA, we found that probe **1** treatment of this protein completely abolishes its enzymatic activity (Figure S17). While FTO (ALKBH9) appears to be structurally primed for reaction with a fluorosulfate (Figure S18), we did not observe appreciable labeling when recombinant FTO was incubated with probe **1** (Figure S19). We suspect that differences between the recombinant and cellular FTO, for example, a post-translational modification of cellular FTO that renders it conjugation competent, may explain our results.

The covalent conjugate between GSTO1 and **1c** was formed quantitatively following a 48 h reaction period in pH 8.5 buffer (Figure S20), and the purified conjugate was crystallized for x-ray structure determination (see Table S5 for data processing and model refinement statistics). The structure of the GSTO1-**1c** conjugate, refined at 2.00 Å resolution, unambiguously shows a diaryl sulfate linkage attached to Tyr229, with the remainder of the probe being located in the small-molecule binding pocket of GSTO1 (Figures 5a, S21). The phenolic oxygen atom of Tyr229 lies in close proximity to the side-chain amine (N_ε) of Lys57, suggesting that this ammonium functionality may play a role in activating Tyr229, likely through pKa perturbation. Intriguingly, the chlorophenyl moiety of probe **1c** resides in a pocket formed by Arg183, Trp222 and Leu226. This pocket is not found in the glutathione-bound form of GSTO1 (Figure 5b).[38] Binding of probe **1c** to GSTO1 requires repositioning of Trp222 to accommodate the chlorophenyl moiety of **1c**, which adopts a slightly bent or "bowed" conformation within the pocket (Figure S22). A similar repositioning of Trp222 is also observed in a previously-reported structure (PDB 4YQV) of GSTO1 bound to a Cys-reactive covalent inhibitor.[39] We hypothesize that the observed distorted conformation of **1c**, which is supported by omit electron density (Figure S22), may arise due to the constrained, covalent nature of the protein-fluorosulfate conjugate.

Mutagenesis experiments indicate that Lys57 and Lys59, which are proximal to the modified Tyr229 side chain in the binding site of GSTO1, are critical for facilitating arylfluorosulfate conjugation (Figure 3c). While the K57Q mutant showed strongly attenuated conjugate formation, the K57R variant maintained native-like labeling. Mutation of the more distant Arg183 to Gln (R183Q) diminished labeling efficiency to a lesser extent. Interestingly, deletion of the catalytic cysteine (Cys32), which has previously been targeted by other classes of covalent inhibitors,[39–40] did not significantly affect labeling efficiency. Mutation of the labeling site established *via* x-ray crystallography (Y229F) produced some residual labeling; this finding suggests that deletion of the reactive nucleophile may shunt the fluorosulfate into a reaction with a nearby side chain, perhaps Lys57 or Lys59. Carrying out the labeling reaction between probe **1** and recombinant GSTO1(K57Q) or GSTO1(K59Q) at elevated pH (10.5 *vs.* 8.0 used for experiments described above) led to a significant enhancement of labeling efficiency, suggesting that pKa perturbation of Tyr229 by the Lys

residues is an important determinant of conjugate formation. Labeling of wild-type GSTO1 was not affected by elevated pH (Figure S23).

Notably, three of the protein targets that we validated with probe **1** (TIGAR, ALKBH5 and HSDL2) were enriched in the probe **1** Competition Ratio experiment, but not in the probe **2** or **3** Competition Ratio experiments. Since the precise function of HSDL2 is unknown, and given the lack of validated ligands for this protein, HSDL2 could be a compelling "Inverse Drug Discovery" target to pursue *via* medicinal chemistry efforts. This approach could ultimately lead to development of a covalent inhibitor useful for the treatment of glioma.[41] HSDL2 has been identified but not validated as a proteomic target of other, more reactive electrophiles; thus, our studies also provide the first example of a validated ligand that may be useful for modulating TIGAR function, which is closely linked to the progression of cancer.[42–43] To the best of our knowledge, a recent study provides the only other example of a validated TIGAR-directed small-molecule probe. [26]

Validation of Probe 2 Reactive Protein Targets

The probe **2** Competition Ratio experiment (i.e., **2**/(**2**+**2c**)) revealed a high abundance ratio for GSTP1 and thiopurine S-methyltransferase (TPMT), as well as for several members of the nucleoside diphosphate kinase (NME or NDPK) family. Covalent conjugate formation between probe **2** and these three target proteins (GSTP1, TPMT and NME1) was established by LC-ESI-MS (Figures S4, S24, S25). For TPMT, the labeling site was assigned to a Lys residue on the periphery of the substrate binding pocket (Lys32) *via* MS-MS (Figures 4f, S26).

The NME1-probe **2** conjugate was successfully crystallized and an x-ray structure determined (see Table S5 for data processing and model refinement statistics). The structure of the NME-probe **2** conjugate, refined at 2.75 Å resolution, contains six copies of the NME1 monomer in the unit cell (space group P1), arranged in a hexameric assembly. Following structure solution by molecular replacement and partial model refinement, we observed clear density for probe **2** linked to NME1 *via* the side chain of Lys12 in all six symmetry-independent monomers (Figures 6, S27). This finding, which is consistent with the labeling site information we obtained by MS-MS experiment (Figure S28), is intriguing because it clearly establishes that arylfluorosulfates can react with proteins *via* Lys side chains, forming a stable linkage. In all six copies of the NME1-**2** conjugate, the side-chain amine (N_ε) of Lys12 is in close proximity to the guanidine group of Arg105. It seems likely that the positively charged arginine side chain lowers the pKa of Lys12, affording the more nucleophilic amine at neutral pH. Notably, we see no evidence for labeling of NME1 with probe **2** at Tyr52, which is directly adjacent to Lys12. The apparently preferential reactivity between probe **2** and Lys12 may stem from a very specific equilibrium binding mode between NME1 and probe **2** that facilitates the SuFEx reaction[10]. A previous study[44] identified Lys12 in NME2, which is closely homologous to NME1, as a catalytic nucleophile involved in DNA cleavage; thus, this residue in NME1 is likely to be intrinsically nucleophilic, although the structural basis for this property of Lys12 is unclear. Mutation of NME1 to remove the reactive nucleophile (i.e., K12A) completely abolished labeling of this protein by probe **2**, as evidenced by LC-ESI-MS analysis (Figure S29).

Notably, two of the three targets of probe **2** that we validated (TPMT and NME1) were strongly enriched in the probe **2** Competition Ratio experiment, but not by probes **1** or **3**. Since NME1 and other members of the NME family lack effective inhibitors, these proteins are compelling targets for development of selective covalent inhibitors. Secretion of NME1 and NME2 from cancers such as leukemia[45–47], lymphoma[48–49] and breast cancer[50–51] is linked to enhanced growth and metastatic potential, and molecules such as ellagic acid that bind and inhibit these NMEs have shown some ability to inhibit cancer growth. Intriguingly, inclusion of ellagic acid (EA; 30 or 60 μM) in the NME1 (3 μM) + probe **2** (50 μM) labeling reaction blocked conjugate formation, suggesting that EA and probe **2** bind at the same site (Figure S30). Given the functional overlap between EA and probe **2**, we expect that probe **2** or an optimized derivative thereof could prove useful for inhibition of cancer cell growth.

Validation of Probe 3 Reactive Protein Targets

Probe **3** is reactive toward a largely different set of proteins relative to arylfluorosulfate probes **1** and **2**. Target proteins identified by the probe **3** Competition Ratio experiment tended to be proteins that have evolved to bind hydrophobic substrates, such as heme oxygenase 2 (HMOX2), an enzyme responsible for catabolism of heme to biliverdin.[52] GSTO1, a target of probe **1**, was also identified in this experiment. The probe **3** Competition Ratio experiment revealed enrichment of CRABP2 and transthyretin (TTR), both small-molecule carrier proteins (non-enzymes) previously identified to be reactive toward arylfluorosulfates[11, 53]

Covalent conjugate formation between probe **3** and its target proteins (GSTO1, CRABP2, HMOX2, TTR) was assessed by LC-ESI-MS (Figures S31 – S34). Consistent with probe **3** binding in the ligand site of CRABP2, labeling of CRABP2 by probe **3** was strongly attenuated by addition of all-*trans* retinoic acid to the labeling experiment (Figure S35). For TTR labeling by probe **3**, addition of tafamidis (5 μM; a drug that binds with negative cooperativity to the two binding sites, exhibiting dissociation constants of 2 nM and 200 nM) to TTR (3 μM) leads to near-complete inhibition of labeling by **3**, as evidenced by an in-gel fluorescence experiment (Figure S36). Interestingly, hydrolysis of the TTR-probe **3** adduct occurred over a period of days, leading to formation of sulfamated-Lys-15-TTR, as has been observed previously.[53] MS-MS analysis of HMOX2 labeling by probe **3** was indicative of covalent conjugate formation at several sites, both at Tyr and Lys residues (Figure 4i, S37). While MS-MS experiments did not facilitate labeling site assignment for TTR or CRABP2, competition between probe **3** and native ligands of these proteins, as well as previously-reported data, [11, 53] allow for confident assignment of specific labeling sites (Figure 4g–4h).

It is notable that three of the four prevalent target proteins of probe **3** that we validated (i.e., HMOX2, CRABP2, TTR) were not significantly enriched in Competition Ratio experiments with either probe **1** or **2**. These results, as well as the conjugate selectivity exhibited by probes **1** and **2**, support the hypothesis that non-covalent equilibrium binding is an important step in targeting specific arylfluorosulfate-reactive proteins. TTR in particular represents a compelling target for development of second generation kinetic stabilizers that function by selective covalent modification. For instance, if probe **3** or a derivative thereof could be

fashioned to enter the central nervous system or the eye following oral administration, this fluorosulfate would offer benefits over the drug tafamidis in terms of targeting amyloidosis within these organs.[54]

Adaptation of Reactive Docking Methods to Fluorosulfate-Reactive Proteins

To begin to understand the structural determinants underpinning the reactivity and the selectivity of probes **1**–**3**, we adapted the reactive docking method previously used to model electrophile reactions with cysteines[26] to model reactions between arylfluorosulfates and proteins. Structures of the 11 validated target proteins were analyzed to detect Tyr and Lys residues suitable to react (namely, solvent accessible Tyr residues in proximity to Arg/Lys side chains, and solvent accessible Lys residues), identifying a total of 250 residues (73 Tyr, 177 Lys; see Table S6). Then, focused reactive docking calculations (i.e., untethered dockings) were performed, centered on each of these residues. Finally, for each protein complex, a docking score was used to rank the likelihood of each residue reacting with the probe used to target that protein. Docking results are summarized in Table S6.

Overall, modeling predictions showed close agreement with the experimental data, correctly identifying the residue to be covalently modified as the best result in 7 out of 11 cases. The residue that was modified was the second ranked in the remaining cases (4 out of 11). Despite the high relative abundance of potentially targetable Lys over Tyr in the structures examined (Tyr/Lys average ratio: 1/2.3), calculations were successful in isolating the rarer Lys sulfamation (best hits are Lys15 in TTR and Lys32 in TPMT; second best hit is Lys12 in NME1). Compared with Tyr, Lys residues are more likely to be found completely solvent exposed on protein surfaces, rather than in pockets and cavities suitable for small molecule binding. While we did not stipulate the presence of ancillary residues for selection of candidate Lys nucleophiles (as was done for Tyr), the requirement of a suitable binding pocket in the docking protocol likely keeps the Lys false positive rate low.

For GSTO1, initially docking into an *apo* structure did not reproduce the binding interaction observed in the x-ray structure; however, assigning Trp222 as a flexible residue during the docking run produced a docked pose that is strikingly similar to the experimental result (Figure 7a). For the interaction of NME1 and probe **2**, our docking protocol predicted the correct labeling site (Lys12) as the second-most stabilizing interaction (after nearby Tyr52). Intriguingly, the highest-ranked pose for Lys12 differed from the x-ray structure only by a rotation of the propargyl imidazole and methoxybenzyl substituents (Figure 7b).

DISCUSSION

In typical drug discovery efforts, a target protein or cellular phenotype to be modified is probed with up to ~1 million candidate small molecules, until the biochemistry or phenotype is perturbed as desired. The experiments reported here exemplify the “Inverse Drug Discovery” approach, which involves surveying the entire cellular proteome with a latent electrophile attached to an equilibrium binding substructure of intermediate complexity to identify proteins that are capable of facilitating the probe’s reactivity. This approach is intended to facilitate discovery of protein targets that are complementary to the combination of organic substructure and latent electrophile used. The “Inverse Drug Discovery” strategy

was employed with the expectation that some of the proteins identified would be relevant to human diseases, or functionally unannotated, and both expectations were realized experimentally in our study. We posit that the "Inverse Drug Discovery" strategy - once applied to a variety of latent electrophiles - has the potential to efficiently discover selective covalent probes for much of the human proteome.

Several of the proteins identified as targets of arylfluorosulfate probes **1**, **2** or **3** are of high therapeutic interest. For instance, significant research effort has focused on development of inhibitors for GSTP1[37, 55–57] and GSTO1[39–40], which inactivate electrophilic chemotherapeutic agents by conjugating them to glutathione. GSTP1 and GSTO1 are upregulated in drug-resistant cancers.[58] CRABP2 is strongly upregulated during development of osteoarthritis, and its inhibition could be valuable in this context.[59] Inhibition of either TIGAR or HSDL2 using a covalent mechanism could prove beneficial for treatment of gliomas, as described previously.[41, 43] Notably, to the best of our knowledge, we validated a small-molecule probe (probe **1**) for HSDL2 for the first time, which should facilitate further probe/inhibitor design. NME1 represents a promising target for treatment of a range of cancers.[45–51, 60] That ellagic acid, a known inhibitor of NME1 function blocks conjugation of arylfluorosulfate probe **2** to NME1 suggests that arylfluorosulfates could prove highly useful to develop better non-covalent cancer drug candidates, or could be developed into covalent cancer drug candidates (inhibitors). Similarly, our finding that validated ligands for TTR and CRABP2 interfere with arylfluorosulfate-based conjugate formation provides strong evidence that these conjugates will be inactive with respect to carrier protein function. We anticipate that rational targeting of any of these proteins with a fluorosulfate optimized by medicinal chemistry efforts will lead to highly specific conjugate formation with the desired target protein, even in the complex mixture of the human proteome.

The "Inverse Drug Discovery" strategy should be particularly well suited to the exploration of functional groups that possess poorly understood latent or 'activatable' reactivity, and/or functional groups whose reactivity has not been explored in a cellular context. This approach can provide insights into the protein structural features that are required to achieve highly selective protein conjugate formation. In essence, the entire proteome serves as the target library to query small molecule probes that have poorly understood, but nonetheless potentially interesting reactivity. The recently-discovered SuFEx reactivity[10], which requires some form of catalysis to facilitate fluoride ion extraction from the highly kinetically stable arylfluorosulfate group, seems to be an excellent latent electrophile for drug development *via* an "Inverse Drug Discovery" approach; not too reactive, but reactive enough in an activating environment.

Importantly, the "Inverse Drug Discovery" approach is complementary to the very useful activity-based protein profiling (ABPP) method used extensively by numerous groups.[22, 61–65] While ABPP efforts typically employ structurally simple, more-reactive covalent probes that are targeted toward a specific class of enzymes, the results within in the context of "Inverse Drug Discovery" demonstrate that probes of intermediate structural complexity comprising less reactive functional groups have the capacity to form conjugates with a group of proteins exhibiting distinct functions, that are related by their ability to first

bind the probe and then react with it. By varying the aromatic scaffold to which the activatable or latent electrophile is attached, we have engaged distinct sets of proteins from the human proteome. Critically, the "Inverse Drug Discovery" approach has enabled the annotation of specific sites within functionally unrelated proteins as being competent for reaction with the arylfluorosulfate functional group, opening the door to further, focused inhibitor design.

In the eleven cases we examined, the reaction between the recombinant purified protein and arylfluorosulfate appears to occur in a well-defined binding pocket. Typically, these are the same sites used to bind endogenous ligands or cofactors (e.g., ATP, heme, retinoic acid, thyroxine). The reactive sites also tend to contain multiple side chains that are primarily cationic at neutral pH (i.e., those of Arg and Lys). As we have observed previously, and confirmed here with mutagenesis experiments, conjugation with arylfluorosulfates depends on these cationic side chains, which likely lower the pKa of the fluorosulfate-reactive nucleophile (Tyr or Lys side chains) and/or catalyze the SuFEx reaction. pKa perturbation due to Coulombic effects (e.g., proximal like charges) is well documented.[66] For the target protein CRABP2, deprotonation of the reactive nucleophile is not sufficient to achieve reactivity[11], suggesting a role for these cationic residues in lowering the free energy of activation of the conjugation reaction, perhaps by binding to the departing fluoride ion. In the cases of GSTP1, GSTO1, and CRABP2, removal of these supporting cationic side chains from the binding pocket strongly attenuates covalent reactivity. A recent study provides evidence that arylfluorosulfates affixed to optimized equilibrium binding substructures can also undergo covalent reaction with a non-catalytic Ser side chain, affording a stable sulfate ester-based conjugate.[36] Collectively, our experimental findings inform predictive models for what sites/proteins are targetable using the arylfluorosulfate group. The new reactive docking model that we report for arylfluorosulfate-protein reactions closely mirrors the experimental data, and provides further evidence that structural recognition on the protein surface is important for acceleration of covalent bond formation, putatively through transition state stabilization, fluoride abstraction, etc. These factors influence the outcome and the selectivity of the SuFEx reaction.

The human proteins that react with the arylfluorosulfates we investigated likely undergo conjugation because they meet the stringent criteria required for SuFEx reactivity. Stabilization of the departing fluoride ion is assumed to be critical to facilitating this reaction, and may be achieved through interactions with hydrogen bond donors within a protein binding pocket, through the influence of hydrophobic-aqueous interfaces[67–69] or electric fields[70] found within proteins, or through a combination of these factors. Whatever the circumstances responsible for enabling SuFEx reactivity within proteins, fluoride must be extracted from a covalent bond as a leaving group, a process that requires surmounting the high kinetic stability of the arylfluorosulfate.

CONCLUSIONS

Here we identified distinct arylfluorosulfates that are useful for covalently targeting largely non-overlapping groups of human proteins. We validated eleven of these targets, and found that each recombinant protein underwent covalent reaction with a particular arylfluorosulfate

to produce a singly-labeled adduct, in some cases through more than one amino acid side chain. Information gleaned from x-ray crystallography and/or MS-MS analysis points to structural commonalities among these arylfluorosulfate-reactive, functionally unrelated human proteins. Other latent electrophiles attached to organic substructures of intermediate complexity can be probed by the "Inverse Drug Discovery" strategy to match small molecules to human proteins they react with, for the purpose of identifying covalent probes for members of the human proteome, and with further medicinal chemistry effort, covalent and non-covalent drug candidates. In summary, we envision that interesting combinations of latent electrophiles and moderately complex organic substructures render the "Inverse Drug Discovery" strategy an efficient means to link chemical structures not yet used in medicinal chemistry efforts with novel or important protein targets in the context of a living organism or a cell lysate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Skaggs Institute for Chemical Biology, the Lita Annenberg Hazen Foundation, and by National Institutes of Health grants DK046335 (J.W.K.), CA132630 (B.F.C.), GM117145 (K.B.S.) and GM069832 (S.F.). D.E.M. was supported by a grant from the George E. Hewitt Foundation for Medical Research. L.P. was supported by a grant from the Leukemia and Lymphoma Society.

References

1. Singh J, Petter RC, Kluge AF. Curr Opin Chem Biol. 2010; 14:475–480. [PubMed: 20609616]
2. Bauer RA. Drug Discov Today. 2015; 20:1061–1073. [PubMed: 26002380]
3. Singh J, Petter RC, Baillie TA, Whitty A. Nat Rev Drug Discov. 2011; 10:307–317. [PubMed: 21455239]
4. Brown JR. Curr Hematol Malig Rep. 2013; 8:1–6. [PubMed: 23296407]
5. Katzenellenbogen JA, Carlson KE, Heiman DF, Robertson DW, Wei LL, Katzenellenbogen BS. J Biol Chem. 1983; 258:3487–3495. [PubMed: 6833211]
6. Staub I, Sieber SA. J Am Chem Soc. 2008; 130:13400–13409. [PubMed: 18781750]
7. Ekkebus R, van Kasteren SI, Kulathu Y, Scholten A, Berlin I, Geurink PP, de Jong A, Goerdayal S, Neefjes J, Heck AJR, Komander D, Ovaa H. J Am Chem Soc. 2013; 135:2867–2870. [PubMed: 23387960]
8. Shannon DA, Banerjee R, Webster ER, Bak DW, Wang C, Weerapana E. J Am Chem Soc. 2014; 136:3330–3333. [PubMed: 24548313]
9. Shannon DA, Weerapana E. Curr Opin Chem Biol. 2015; 24:18–26. [PubMed: 25461720]
10. Dong J, Krasnova L, Finn MG, Sharpless KB. Angew Chem Int Ed. 2014; 53:9430–9448.
11. Chen W, Dong J, Plate L, Mortenson DE, Brighty GJ, Li S, Liu Y, Galmozzi A, Lee PS, Hulce JJ, Cravatt BF, Saez E, Powers ET, Wilson IA, Sharpless KB, Kelly JW. J Am Chem Soc. 2016; 138:7353–7364. [PubMed: 27191344]
12. Wang L, Li Y, Yan H. J Biol Chem. 1997; 272:1541–1547. [PubMed: 8999826]
13. Furuhashi M, Hotamisligil GS. Nat Rev Drug Discov. 2008; 7:489. [PubMed: 18511927]
14. Maden M. Nat Rev Neurosci. 2007; 8:755–765. [PubMed: 17882253]
15. Adam GC, Cravatt BF, Sorensen EJ. Chem Biol. 2001; 8:81–95. [PubMed: 11182321]
16. Adam GC, Sorensen EJ, Cravatt BF. Nat Biotech. 2002; 20:805–809.

17. Shannon DA, Gu C, McLaughlin CJ, Kaiser M, van der Hoorn RAL, Weerapana E. ChemBioChem. 2012; 13:2327–2330. [PubMed: 23008217]
18. Gu C, Shannon DA, Colby T, Wang Z, Shabab M, Kumari S, Villamor Joji G, McLaughlin Christopher J, Weerapana E, Kaiser M, Cravatt BF, van der Hoorn RAL. Chem Biol. 2013; 20:541–548. [PubMed: 23601643]
19. Zhao Q, Ouyang X, Wan X, Gajiwala KS, Kath JC, Jones LH, Burlingame AL, Taunton J. J Am Chem Soc. 2017; 139:680–685. [PubMed: 28051857]
20. Schardon CL, Tuley A, Er JAV, Swartzel JC, Fast W. ChemBioChem. 2017; 18:1551–1556. [PubMed: 28470883]
21. Clulow JA, Storck EM, Lanyon-Hogg T, Kalesh KA, Jones LH, Tate EW. Chem Comm. 2017; 53:5182–5185. [PubMed: 28439590]
22. Liu Y, Patricelli MP, Cravatt BF. Proc Natl Acad Sci U S A. 1999; 96:14694–14699. [PubMed: 10611275]
23. Chang JW, Cognetta AB, Niphakis MJ, Cravatt BF. ACS Chem Biol. 2013; 8:1590–1599. [PubMed: 23701408]
24. Xiao J, Broz P, Puri AW, Deu E, Morell M, Monack DM, Bogyo M. J Am Chem Soc. 2013; 135:9130–9138. [PubMed: 23701470]
25. Cognetta AB, Niphakis MJ, Lee HC, Martini ML, Hulce JJ, Cravatt BF. Cell Chem Biol. 2015; 22:928–937.
26. Backus KM, Correia BE, Lum KM, Forli S, Horning BD, González-Páez GE, Chatterjee S, Lanning BR, Teijaro JR, Olson AJ, Wolan DW, Cravatt BF. Nature. 2016; 534:570–574. [PubMed: 27309814]
27. Wang C, Weerapana E, Blewett MM, Cravatt BF. Nat Methods. 2014; 11:79–85. [PubMed: 24292485]
28. Rostovtsev VV, Green LG, Fokin VV, Sharpless KB. Angew Chem Int Ed. 2002; 41:2596–2599.
29. Speers AE, Adam GC, Cravatt BF. J Am Chem Soc. 2003; 125:4686–4687. [PubMed: 12696868]
30. Wang Q, Chan TR, Hilgraf R, Fokin VV, Sharpless KB, Finn MG. J Am Chem Soc. 2003; 125:3192–3193. [PubMed: 12630856]
31. Horning BD, Suciu RM, Ghadiri DA, Ulanovskaya OA, Matthews ML, Lum KM, Backus KM, Brown SJ, Rosen H, Cravatt BF. J Am Chem Soc. 2016; 138:13335–13343. [PubMed: 27689866]
32. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, Burkhard PR, Sanchez JC. Anal Chem. 2008; 80:2921–2931. [PubMed: 18312001]
33. Ting L, Rad R, Gygi SP, Haas W. Nat Methods. 2011; 8:937–940. [PubMed: 21963607]
34. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]
35. Zou S, Toh JDW, Wong KHQ, Gao YG, Hong W, Woon ECY. Sci Rep. 2016; 6:25677. [PubMed: 27156733]
36. Fadeyi OO, Hoth LR, Choi C, Feng X, Gopalsamy A, Hett EC, Kyne RE, Robinson RP, Jones LH. ACS Chem Biol. 2017; 12:2015–2020. [PubMed: 28718624]
37. Oakley AJ, Bello ML, Battistoni A, Ricci G, Rossjohn J, Villar HO, Parker MW. J Mol Biol. 1997; 274:84–100. [PubMed: 9398518]
38. Board PG, Coggan M, Chelvanayagam G, Easteal S, Jermiin LS, Schulte GK, Danley DE, Hoth LR, Griffor MC, Kamath AV, Rosner MH, Chrunyk BA, Perregaux DE, Gabel CA, Geoghegan KF, Pandit J. J Biol Chem. 2000; 275:24798–24806. [PubMed: 10783391]
39. Ramkumar K, Samanta S, Kyani A, Yang S, Tamura S, Ziemke E, Stuckey JA, Li S, Chinnaswamy K, Otake H, Debnath B, Yarovenko V, Sebolt-Leopold JS, Ljungman M, Neamati N. Nat Comm. 2016; 7:13084.
40. Tsuboi K, Bachovchin DA, Speers AE, Spicer TP, Fernandez-Vega V, Hodder P, Rosen H, Cravatt BF. J Am Chem Soc. 2011; 133:16605–16616. [PubMed: 21899313]
41. Ruokun C, Yake X, Fengdong Y, Xinting W, Laijun S, Xianzhi L. Tumor Biol. 2016; 37:15065–15077.
42. Bensaad K, Cheung EC, Vousden KH. EMBO J. 2009; 28:3015–3026. [PubMed: 19713938]
43. Wanka C, Steinbach JP, Rieger J. J Biol Chem. 2012; 287:33436–33446. [PubMed: 22887998]

44. Postel EH, Abramczyk BM, Levit MN, Kyin S. Proc Natl Acad Sci U S A. 2000; 97:14194–14199. [PubMed: 11121025]
45. Okabe-Kado J. Leuk Lymphoma. 2002; 43:859–867. [PubMed: 12153176]
46. Lilly AJ, Khanim FL, Hayden RE, Luong QT, Drayson MT, Bunce CM. Cancer Res. 2011; 71:1177–1186. [PubMed: 21169412]
47. Lilly AJ, Khanim FL, Bunce CM. Naunyn-Schmiedeberg's Arch Pharmacol. 2015; 388:225–233. [PubMed: 25119778]
48. Niitsu N, Okabe-Kado J, Okamoto M, Takagi T, Yoshida T, Aoki S, Hirano M, Honma Y. Blood. 2001; 97:1202. [PubMed: 11222361]
49. Niitsu N, Nakamine H, Okamoto M, Akamatsu H, Honma Y, Higashihara M, Okabe-Kado J, Hirano M. Br J Haematol. 2003; 123:621–630. [PubMed: 14616965]
50. Rumjahn SM, Yokdang N, Baldwin KA, Thai J, Buxton ILO. Br J Cancer. 2009; 100:1465–1470. [PubMed: 19367276]
51. Yokdang N, Nordmeier S, Speirs K, Burkin HR, Buxton ILO. Integr Cancer Sci Ther. 2015; 2:192–200. [PubMed: 26413311]
52. Bianchetti CM, Yi L, Ragsdale SW, Phillips GN. J Biol Chem. 2007; 282:37624–37631. [PubMed: 17965015]
53. Baranczak A, Liu Y, Connelly S, Du WGH, Greiner ER, Genereux JC, Wiseman RL, Eisele YS, Bradbury NC, Dong J, Noodleman L, Sharpless KB, Wilson IA, Encalada SE, Kelly JW. J Am Chem Soc. 2015; 137:7404–7414. [PubMed: 26051248]
54. Coelho T, Merlini G, Bulawa CE, Fleming JA, Judge DP, Kelly JW, Maurer MS, Planté-Bordeneuve V, Labaudinière R, Mundayat R, Riley S, Lombardo I, Huertas P. Neurol Ther. 2016; 5:1–25. [PubMed: 26894299]
55. Ang WH, Parker LJ, De Luca A, Juillerat-Jeanneret L, Morton CJ, Lo Bello M, Parker MW, Dyson PJ. Angew Chem Int Ed. 2009; 48:3854–3857.
56. Bräutigam M, Teusch N, Schenk T, Sheikh M, Aricioglu RZ, Borowski SH, Neudörfl JM, Baumann U, Griesbeck AG, Pietsch M. ChemMedChem. 2015; 10:629–639. [PubMed: 25694385]
57. Crawford LA, Weerapana E. Mol BioSys. 2016; 12:1768–1771.
58. Townsend DM, Tew KD. Oncogene. 2003; 22:7369–7375. [PubMed: 14576844]
59. Welch ID, Cowan MF, Beier F, Underhill TM. Arthrit Res Ther. 2009; 11:R14.
60. Otero-Estévez O, De Chiara L, Barcia-Castro L, Páez de la Cadena M, Rodríguez-Berrocal FJ, Cubiella J, Hernández V, Martínez-Zorzano VS. Sci Rep. 2016; 6:26703. [PubMed: 27222072]
61. Speers AE, Cravatt BF. ChemBioChem. 2004; 5:41–47. [PubMed: 14695510]
62. Choe Y, Leonetti F, Greenbaum DC, Lecaille F, Bogyo M, Brömme D, Ellman JA, Craik CS. J Biol Chem. 2006; 281:12824–12832. [PubMed: 16520377]
63. Willems LI, Overkleeft HS, van Kasteren SI. Bioconj Chem. 2014; 25:1181–1191.
64. Hatzios SK, Abel S, Martell J, Hubbard T, Sasabe J, Munera D, Clark L, Bachovchin DA, Qadri F, Ryan ET, Davis BM, Weerapana E, Waldor MK. Nat Chem Biol. 2016; 12:268–274. [PubMed: 26900865]
65. Lentz CS, Ordonez AA, Kasperkiewicz P, La Greca F, O'Donoghue AJ, Schulze CJ, Powers JC, Craik CS, Drag M, Jain SK, Bogyo M. ACS Infect Dis. 2016; 2:807–815. [PubMed: 27739665]
66. Harris TK, Turner GJ. IUBMB Life. 2002; 53:85–98. [PubMed: 12049200]
67. Narayan S, Muldoon J, Finn MG, Fokin VV, Kolb HC, Sharpless KB. Angew Chem Int Ed. 2005; 44(21):3275–3279.
68. Jung Y, Marcus RA. J Am Chem Soc. 2007; 129(17):5492–5502. [PubMed: 17388592]
69. McFearin CL, Richmond GL. J Phys Chem C. 2009; 113(50):21162–21168.
70. Fried SD, Bagchi S, Boxer SG. Science. 2014; 346(6216):1510. [PubMed: 25525245]

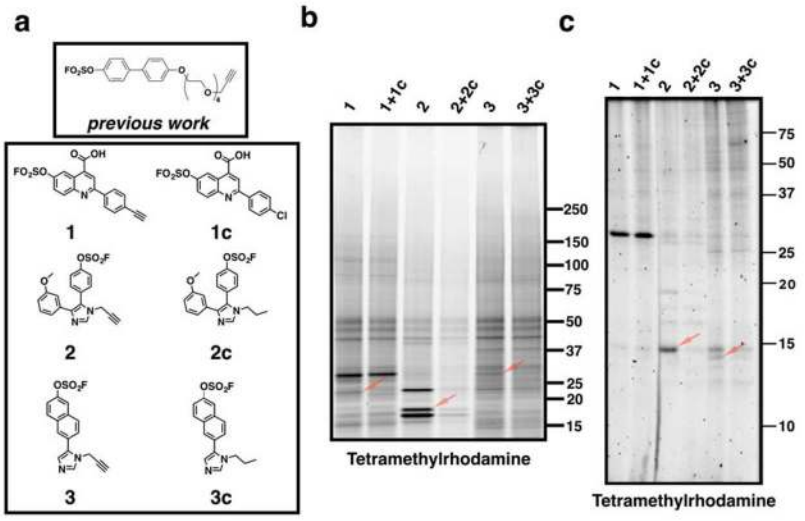


Figure 1.

Chemical structures of arylfluorosulfate probes and gel-based analysis of HEK293T proteome labeling. (a) Chemical structures of a previously reported arylfluorosulfate probe, and three new arylfluorosulfate alkyne probes **1**–**3** and corresponding competitors **1c**–**3c**. (b) SDS-PAGE/Rhodamine analysis of HEK293T lysate labeling (24 h) by alkyne probes (10 μM) in the presence or absence of excess concentration (90 μM) of corresponding competitor probe. (c) Gel analysis of live HEK293T cell labeling (24 h) by alkyne probes (10 μM) in the presence or absence of excess concentration (90 μM) of the corresponding competitor probe. For both (b) and (c), Coomassie-stained gels may be found in Supporting Information (Figure S1). Examples of bands that are diminished in intensity by addition of competitor in either (b) or (c) are highlighted with red arrows.

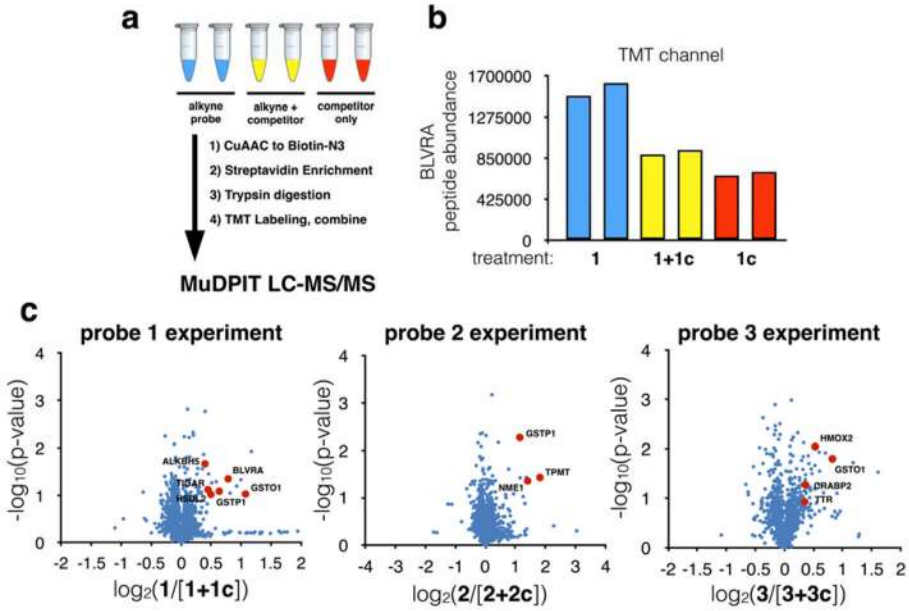


Figure 2.

MuDPIT MS/MS workflow and data derived from proteome treatments with arylfluorosulfate probes. (a) Experimental workflow for CuAAC enrichment and MuDPIT MS/MS analysis of alkyne and competitor probe-labeled lysate. (b) Plot of validated probe 1 target BLVRA peptide abundance in TMT channels corresponding to alkyne-only (blue), alkyne-plus-competitor (yellow) and competitor-only (red) lysate treatments. (c) Volcano plots of $-\log_{10}$(p-value) vs. $\log_2$(alkyne/competition) peptide enrichment for proteins identified via MuDPIT LC-MS/MS using probes **1** and **1c**, **2** and **2c**, and **3** and **3c**. Positions of validated protein targets are indicated in the plots by red circles.

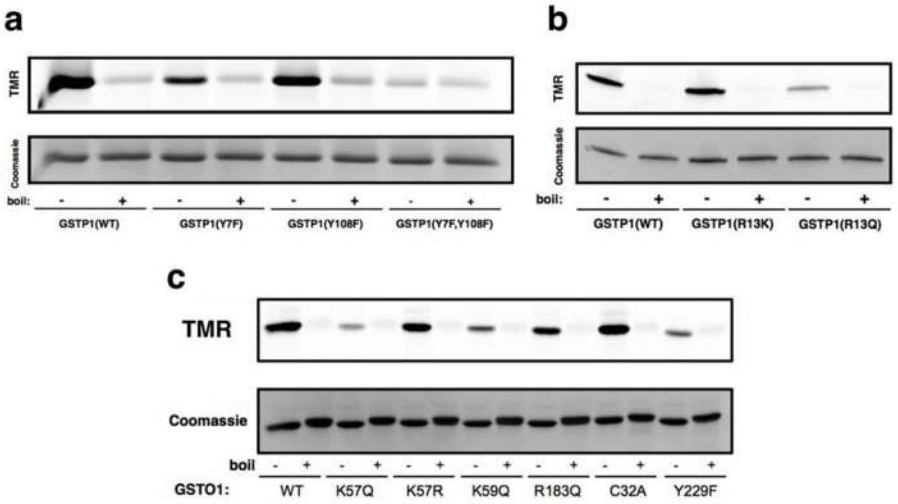


Figure 3.

Gel-based analysis of binding-site mutations on arylfluorosulfate target protein labeling. (a) GSTP1 labeling by probe **2** occurs at both Tyr7 and Tyr108, and mutation of both positions to Phe completely abrogates labeling reaction. GSTP1 variants (3 μM) were incubated 17 h at room temperature with probe 2 (60 μM) in 50 mM Tris pH 8.0. (b) Covalent reaction between GSTP1 and probe **2** depends on the presence of a cationic side chain in the binding pocket. GSTP1 variants (3 μM) were incubated 1 h at room temperature with probe **2** (60 μM) in 50 mM Tris pH 8.0. (c) Effects of binding-site mutagenesis on labeling of GSTO1 by probe **1**. GSTO1 variants (3 μM) were incubated 14 h at room temperature with probe **1** (60 μM) in 50 mM Tris pH 8.0. For all samples, labeling was followed by CuAAC reaction with TMR-N_3 and resolution by SDS-PAGE. Control (boil) samples were pre-boiled 10 min in buffer containing 0.1 % SDS prior to labeling reaction. Protein concentration in each experiment was established via Coomassie staining. “WT” refers to wild-type sequences.

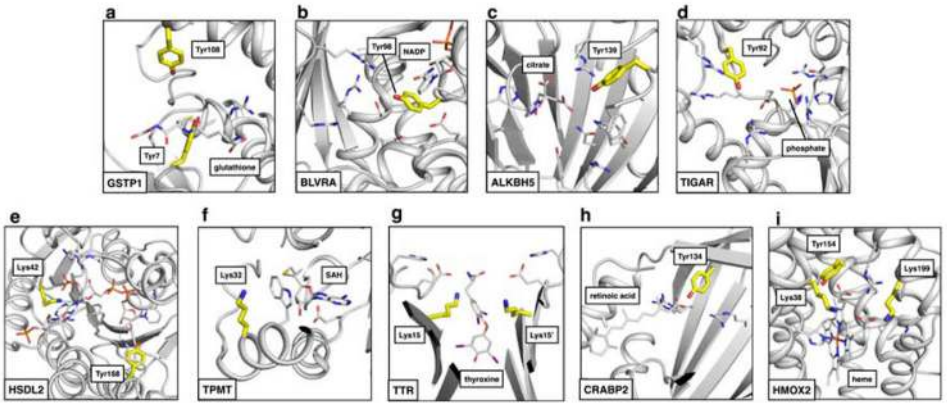


Figure 4.

Validated target proteins with assigned labeling sites in known binding pockets highlighted (ALKBH5: 4O61; BLVRA: 2H63; CRABP2: 2FR3; GSTP1: 5GSS; HSDL2: 3KVO; HMOX2: 2QPP; TIGAR: 3DCY; TPMT: 2BZG; TTR: 2ROX). See Supporting Information for full lists of putative labeling sites identified via MS–MS.

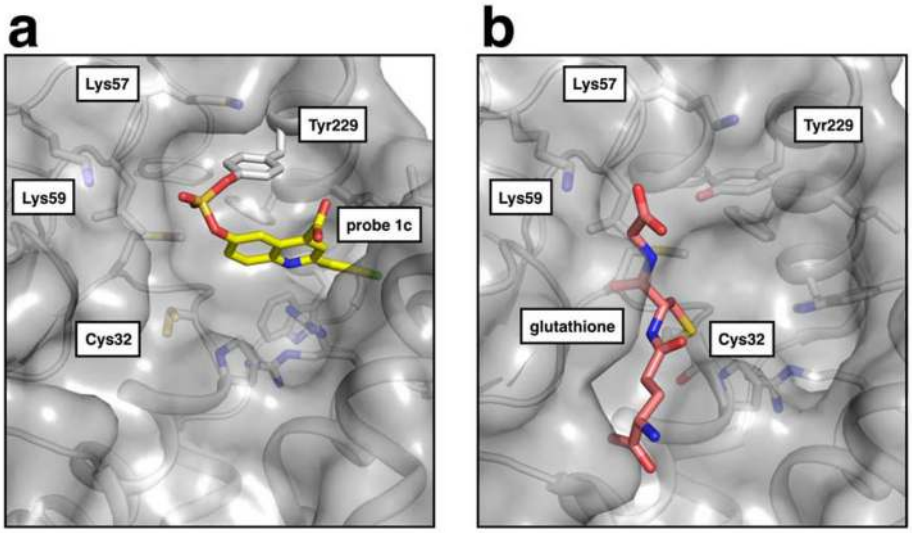


Figure 5.

Structural comparison of arylfluorosulfate-conjugated and glutathione-bound forms of GSTO1. (a) Crystal structure of the GSTO1-probe **1c** conjugate, refined at 2.00 Å resolution. The protein components of the structure, including Tyr229, are shown in grey, while probe **1c** is shown in yellow. (b) Structure of GSTO1 conjugated to glutathione (PDB 1EEM) shown in the same orientation as depicted in panel (a). In both images, the side chains of the fluorosulfate-reactive Tyr229, the catalytic Cys32 and several nearby residues are indicated.

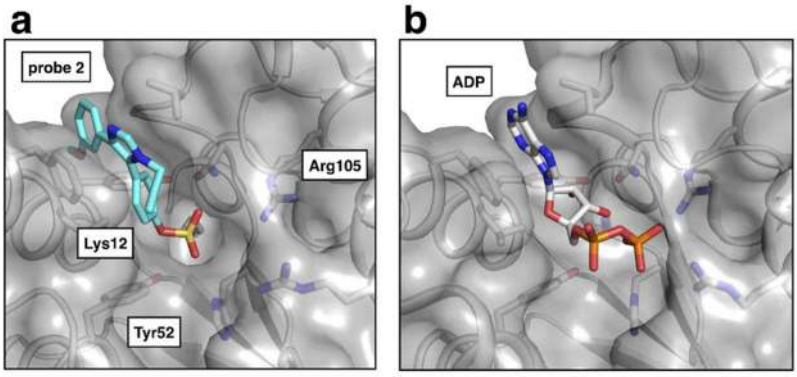


Figure 6.

Structural comparison of arylfluorosulfate-conjugated and ADP-bound forms of NME1. (a) Crystal structure of the NME1-probe 2 conjugate, refined at 2.75 Å resolution. The protein components of the structure, including Lys12, are shown in grey, while probe 2 is shown in cyan. (b) Structure of NME1 bound to ADP (PDB 2HVE) shown in the same orientation as depicted in panel (a).

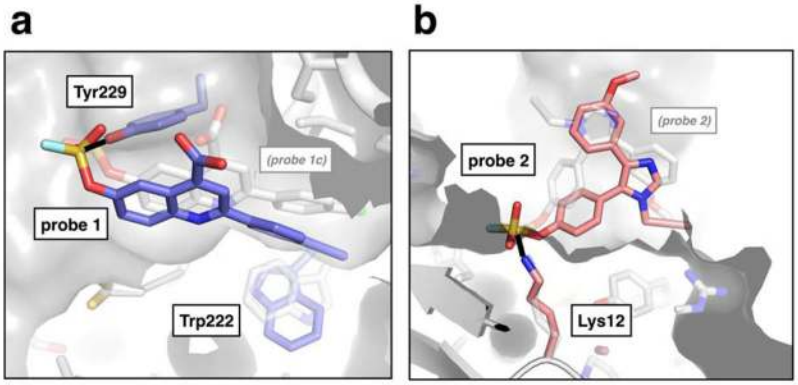


Figure 7.

Structural comparisons of arylfluorosulfate-conjugated forms of GSTO1 and NME1 and docked poses. (a) Docked pose of probe **1** (purple) in binding pocket of GSTO1 superimposed with probe **1c** from GSTO1-**1c** conjugate x-ray structure (gray). For reactive docking run, **1** was docked into PDB structure 1EEM, with side chain of Trp222 treated as flexible, leading to identification of Tyr229 as covalent binding site. (b) Docked pose of probe **2** (pink) in binding pocket of NME1 superimposed with probe **2** from NME1-**2** conjugate x-ray structure (gray). For reactive docking run, **2** was docked into PDB structure 2HVE, leading to assignment of Lys12 as a putative covalent binding site. In both panels, nascent chemical bonds predicted by reactive docking methods are shown as black lines.