Development of chemical probes to profile functional tyrosines in live cells using sulfur triazole exchange chemistry (SuTEx)

Emmanuel Kipruto Toroitich
Eldoret, Kenya

B.S. Chemistry, Bates College, 2015

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Chemistry
University of Virginia
August 2020

Acknowledgement

Foremost, I will like to express my sincere gratitude to my PI and mentor, Dr. Ken Hsu for the support and invaluable guidance over the past 5 years. His motivation, enthusiasm and immense knowledge of the chemical biology field were a source of inspiration. I will also extend my heartfelt thanks to his wife, Sally, for making us feel part of the family.

To my thesis committee, Dr. Andreas Gahlmann, Dr. Cliff Stains, Prof. Donald Hunt and Prof. Gordon Laurie thank you for your encouragement and insightful feedback.

My sincere thanks also go to the many faculty members and staff at UVA who have gone out of their way to support my research. Dr. Adam Libby, Dr. Mark Ross, Dr. Earl Ashcraft, Dr. Diane Dickie, Dr. Jeff Eleanor, Danny Via, Michael Birkhead, Debbie Scott, Seth Matula and Susie Marshall thank you.

To the current and former lab members, it was a pleasure work with you and learn from you all. I do appreciate you all taking the time to train me on new techniques and providing feedback on my projects.

To John manners, Prof. Mike Boit and the entire KenSAP program family, I will forever be indebted to you for your assistance with college admission process and continuous support throughout my career.

Special shoutout to amazing friends I have met in Charlottesville- Wakanda family, Charlottesville premier league fans. Thank you providing avenues to relax and reenergize.

To my family (Including Allen family), I would not be in this position without your continuous encouragement and support. To my Dad, you biking every day for 10+ years to sell milk to ensure we could afford food, school fees and other necessities, will not be in vain. Of course, that will not have been possible without mum on your side. Thank you all for the sacrifices you had to make. I will always work hard to make you proud

Lastly, I want to thank GOD for his blessings and protection.

Abstract

The recent advancement of chemoproteomic technologies has led to discoveries of electrophilic small molecule warheads targeting lysine and cysteine residues that have enormously expanded our understanding of biology by enabling the identification of druggable protein targets and their associated biological functions. Still, chemical ligands have been reported for only a small fraction of the human proteome. Approximately one in three human proteins fall into the "dark genome", of which almost nothing is known regarding their structure and function. To expand the scope of ligandable proteins, chemical probes targeting residues beyond cysteine and lysine are needed.

This work introduces Sulfur-triazole exchange chemistry (SuTEx) as a platform for developing covalent probes for tunable targeted reactivity toward more than 10,000 unique tyrosine sites in ~3700 proteins in cell lysates and live cells. Modifications to the triazole leaving group furnished sulfonyl probes with ~5-fold enhanced chemoselectivity for tyrosines over other nucleophilic amino acids. Approximately 70% of the proteins bound by the probes are not in the DrugBank database and include proteins involved in RNA-recognition and protein-protein interactions, which have historically been challenging to target with small molecules. Additionally, we discovered ~30% of the tyrosine sites labeled by the probes are annotated as phosphotyrosine sites. As a proof of concept, we applied SuTEx as a chemical phosphoproteomics strategy to monitor activation of phosphotyrosine sites.

To expand the utility of the SuTEx platform, we designed small molecule fragments based on the tested probes to target proteins with no effective inhibitors. The small molecule fragments exhibited high selectivity and potency in live cells against two target proteins, Acetyl-CoA acetyltransferase 2 (ACAT2, IC_{50} = 5 μM) and prostaglandin reductase 2 (PTGR2, IC_{50} = 1 μM). PTGR2 is found to be expressed in pancreatic cancer tissues but absent in normal pancreatic tissue,

and knockdown of its expression was found to reduce tumor growth and induce apoptosis. ACAT2 is involved in regulating lipid metabolism by catalyzing the synthesis of acetoacetyl-CoA from two acetyl-CoA molecules, which is later converted to cholesterol. ACAT2 deficiency has been shown to reduce amounts of atherosclerosis.

In summary, we demonstrated that sulfur triazole exchange chemistry is a powerful platform that can provide new biological insights and novel chemical probes for drug discovery.

Table of Contents

Chapter 1: Introduction

1.1 Overview of drug discovery approaches

The two common approaches taken for drug discovery are classic genetics and chemical genetics. Classic genetics is used to study biology by manipulating the biological system at the gene level by mutating the gene and observing the resulting change in phenotype (Fig 1.1). Classic genetics is further subdivided to forward genetics (random mutation of the genes and observe the phenotype change) and reverse genetics (mutation of a specific gene). In contrast, chemical genetics involves the use of small molecules to perturb protein functions. Chemical genetics can be similarly subdivided. Forward chemical genetics involves the use of small molecules to screen for a desired phenotypic effect, and reverse chemical genetics pertains to the screening of small molecules targeting a specific protein[1-2] (Fig 1.1).

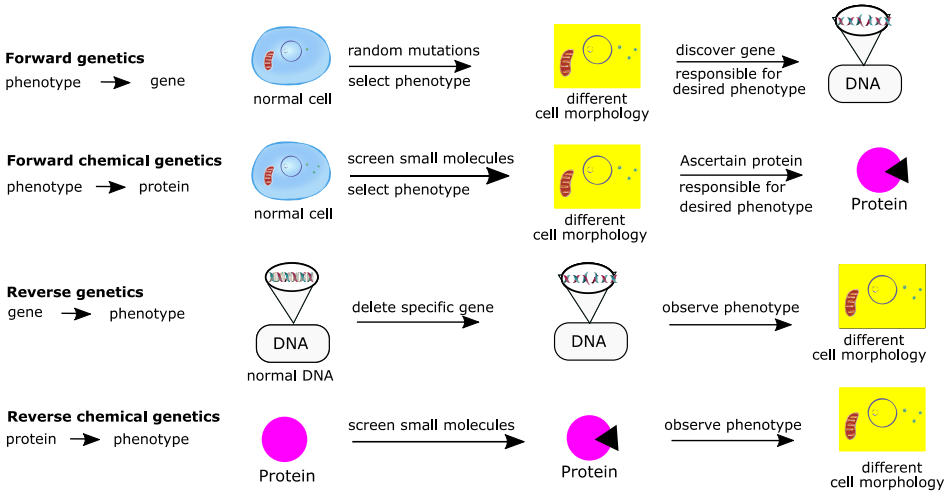


Fig 1.1 Comparison of classic genetics and chemical genetics

Overall, the use of chemical genetics has emerged as a preferred approach over classic genetics for drug discovery for three key reasons. First, genetic models do not represent the physiological systems, and the related genes can at times compensate for the mutated gene leading to observed results not translating to *in situ* or *in vivo models*. Secondly, the effects of mutations are permanent.

They cannot be turned on or off at will. In contrast, the effects of small molecules are reversible due to metabolism and clearing. Lastly, the ease of molecular synthesis and portfolio diversity enables small molecules to be easily and affordably scaled up, which cannot be deployed for classic genetics.[1-2]

The main drawback of the chemical genetics approach is that while any gene, in principle, can be mutated, at present there is no small molecule ligand for most human proteins. Specifically, of the ~20,000 human proteins, only ~11% have been liganded with 1.4-4 % having a quality chemical probe. To qualify as a quality chemical probe, the small molecules must meet these criteria: (1) potency: <100 nm on-target bioactivity, (2) selectivity: at least 10-fold selectivity for the main target over other tested targets, and (3) permeability: cell-active at <10 μM. From > 1.8 million small molecules available in the public database analyzed by Antolin et al. only 355, 304 (20 %) have an acceptable level of biochemical activity (<10 μM *in vitro* activity)[3]. Of these, 14 % meet potency and selectivity criteria and only 2,558 (0.7%, 0.14% of the total compounds) qualify as a quality chemical probe. The compounds that meet the potency and selectivity criteria can probe 795 human proteins (4% of human proteome), and when the permeability criteria are considered, only 250 (1.2 %) proteins can be studied[3].

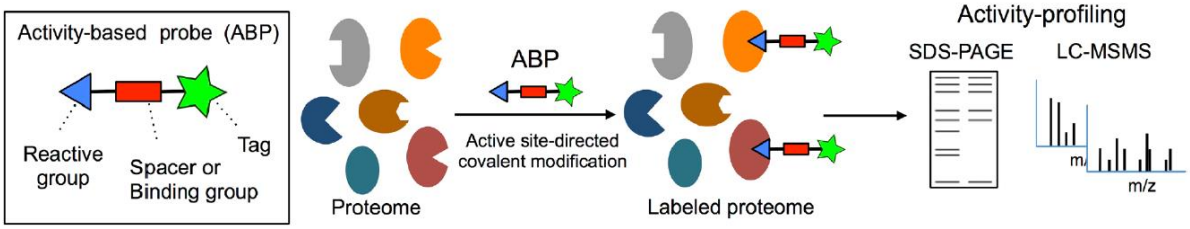
Several factors contribute to such a small fraction of the proteome being ligandable (or, having quality chemical probes). The major one is the lack of effective tools to study the whole proteome. Approximately one in three human proteins fall into the dark human genome, with nearly nothing known about them[4]. As a result, it limits the researchers to remain under the lamp post. For example, if working on a kinase inhibitor, researchers will screen only against kinases because there are no tools to screen against the whole proteome. This leads to researchers overstating compounds selectivity and potentially missing the therapeutic potential and toxicity of

their compounds before costly animal studies, but even then, cannot identify the specific gene responsible for the resulting phenotype.[4]

1.2 Activity-based protein profiling

One of the recent advances in technology that is proving to be key in expanding the ligandable proteome is activity-based protein profiling (ABPP). ABPP is a chemical proteomic technique that enables the study of proteins activity in a complex crowded environment based on enzyme activity rather than expression level[5]. ABPP probes comprise of three main components: reactive group (warhead), spacer, and a reporter tag (Fig 1.2). The warhead covalently binds to a specific active enzyme or a class of enzymes with related functions. The warhead is linked covalently to either a fluorophore for visualization by in-gel fluorescence or biotin groups for enrichment for liquid chromatography-mass spectrometry analysis (LC-MS)[5]. The fluorophores typically used include fluorescein, rhodamine, and BODPIY. Although fluorescein and rhodamine are commonly used fluorophores, due to their large size, they are not cell-permeable, as a result, their use is limited to *ex vivo*.

).



(adapted from JACS.2019, 2782-2799)

Fig 1.2. Activity-based protein profiling (ABPP) overview

The advent of biorthogonal chemistries, such as copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) has revolutionized ABPP in three key ways. First, it enables the replacement of bulky reporter groups with small, cell-permeable affinity handles such as an alkyne or azide. Secondly,

a single probe can be diversified with multiple reporter groups. For example, proteomes from a live cell treatment can be split into two fractions. The first fraction can be conjugated with rhodamine azide for in-gel fluorescence imaging for a fast readout, and the second fraction can be conjugated with desthiobiotin azide for a detailed analysis of probe targets using mass spectrometry. Thirdly, the small size of the clickable probes minimizes steric interaction in enzyme active sites, which improves the binding affinity[6].

Beside CuAAC, other biorthogonal chemistries include a traceless Staudinger ligation which couples azides with triarylphosphines to produce an amide linkage[7], (b) Diels-Alder cycloaddition of 1,2,4,5-tetrazine and a strained diene[8], and (c) copper-free azide-alkyne addition, which uses strained alkyne to accelerate the reaction[9].

ABPP offers several advantages that make it a superior method for probe/inhibitor protein engagement elucidation. ABPP probes react based on enzyme activity rather than expression levels, providing access to low abundant proteins in a crowded cellular environment. Additionally, ABPP enables the elucidation of post-translational reactions that regulate enzyme activity, and finally, selectivity and reactivity can be assessed in parallel[5], which is not only convenient but also saves research time and cost.

1.2.1 Gel-Based ABPP

The gel-based method is a high-throughput strategy to elucidate protein activity. Briefly, proteomes are first treated with the probe either already appended with a fluorophore, such as FP-rhodamine, or conjugated by CuAAC. In competitive gel-based ABPP, the proteomes are treated with an inhibitor compound or vehicle (control, e.g., DMSO) before being treated with the probe. The probe-treated samples are then denatured with sodium dodecyl sulfate (SDS) and 2-mercaptoethanol, thereby denaturing the folded proteins. SDS also serves to coat the proteins with

a negative charge. The negatively-charged proteomes are loaded to polyacrylamide gel, and when the electrical field is applied, the proteins are separated based on their molecular weights, with the lower molecular weight molecules moving faster toward the positive cathode. The probe-labeled proteins are visualized by in-gel fluorescence scanning, and any changes in band intensity, inhibitor- relative to vehicle-treated control, indicate inhibition of probe binding and therefore protein activity. (Fig 1.2). Gel-based ABPP is widely used because of its affordability, robustness, and throughput. In our lab, ~100 samples can be analyzed in a day. Gel-ABPP, however, is limited in its inability to provide the identity of the protein/peptide modified by the probe or inhibitor. In addition, due to low resolution, less abundant proteins are often masked if they have similar molecular weight with more abundant proteins, which leads to over estimating inhibitor selectivity.

1.2.2 Liquid chromatography-mass spectrometry ABPP

The advances in mass spectrometry technology have revolutionized the field of chemical biology and drug discovery. It solves the shortcoming of gel-ABPP by providing a platform to identify and quantify probe/inhibitor protein/peptide targets and post-translational modifications. As a general overview of LC-MS ABPP, proteomes are treated with the ABPP probe and conjugated to biotin or desthiobiotin by CuAAC, followed by treatment with dithiothreitol (DTT) to break disulfide bonds, and the free thiols are alkylated with iodoacetamide. For protein target identification, the biotinylated probes are enriched with (strept)avidin beads, and the enriched proteins digested with trypsin, and the resulting peptides analyzed by LC-MS/MS. For peptide target identification, the biotinylated proteins are first trypsin digested, and the probe-modified peptides enriched with (strept) avidin, eluted with organic solvents and analyzed by LC-MS/MS.

So why should you use desthiobiotin instead of biotin? Desthiobiotin binds with lower affinities to avidin and streptavidin than biotin (Kd=10^{-11} M vs 10^{-15} M respectively) while still

providing excellent specificity[10]. The lower binding affinity enables the desthiobiotin-labeled proteins/peptides to be eluted with mild conditions, typically 1:1 acetonitrile/water. In contrast, biotin-(strept)avidin interactions (binding) cannot be dissociated without harsh conditions (boiling, extreme PH, or denaturants), which could affect protein integrity and downstream applications[10].

1.2.3 ABPP-SILAC

ABPP in combination with stable isotope labeling by amino acids (SILAC) provides a platform to identify and semi-quantify probe and inhibitor target engagement. In SILAC, cells are grown in media containing either "light" (($^{12}C_6$,$^{14}N_2$) L-lysine and ($^{12}C_6$,$^{14}N_4$) L-arginine)) or "heavy" (($^{13}C_6$,$^{15}N_2$) L-lysine and ($^{13}C_6$,$^{15}N_4$) L-arginine)), resulting in the synthesis of proteins with different masses of lysines and arginines[11]. The light cells are treated with vehicle control and the heavy cells are treated with the experimental compound, and after CuAAC of the alkyne probe with desthiobiotin azide the "light" and "heavy" proteomes are mixed 1:1. Digestion of the proteomes with trypsin protease results in peptides with identical sequences and with at least one lysine or arginine with different masses (nominal 8 and 10 Da mass difference between "light" and "heavy" lysine and arginine respectively). Because the peptides in each light/heavy pair have the same sequence, they will have the same chromatographic retention time but can be differentiated by their mass-to-charge ratios (m/z). The relative intensity of light and heavy peptide molecular ions (MS1) is used to quantify the effect of the treatment conditions.

The use of ABPP-SILAC, however, suffers from two major limitations. First, is the cost of the heavy isotopic amino acids. 1 gram each of heavy lysine and arginine both costs ~$2,000 (Sigma Aldrich), with academic discount and perplexing $9,200 without a discount. I will not delve into the pricing difference of the two options, but the general high cost limits the use of ABPP-SILAC especially new researchers with limited funding. Secondly, because ABPP-SILAC

relies on cells being grown in media supplemented with the amino acids, it is not a viable option for research on *in vivo* animal models or patient-derived cells. To address the latter limitation, researchers use, for click chemistry, an azide functionalized with isotopically-labeled tobacco etch virus biotin tag[12]. Another option is Tandem Mass Tags (TMT), which is used to isotopically label peptides after trypsin digestion. TMT are isobaric chemical tags, with varying numbers and combinations of ^{13}C and ^{15}N. The tag contains an amine-reactive group which binds to the N-terminus of a peptide or a lysine residue. Besides enabling analysis of cells or tissues, TMT enables analysis of up to 16 different samples[13] in one LC-MS experiment.

The ability to identify and quantify chemical probe/inhibitor-protein engagement in lysates and live cells has transformed the field of chemical biology and has uncovered new biology. Over the last decade, chemical probes to study enzymes such as lysine, methionine, cysteine and post-translational modifications, including methylation and crotonylation have been developed. I will highlight a few of the probes with their important implications

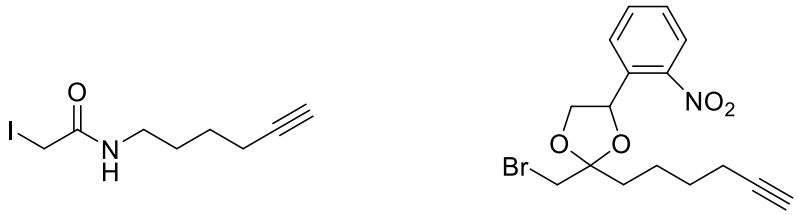
1.3 Activity-based protein profiling probes

1.3.1 Cysteine targeting probes

Cysteine is the most targeted amino acid residue partly because of its high reactivity (nucleophilicity) and redox sensitivity of cysteine thiol[14]. Although the average pKa of surface-exposed thiol is ~8.5, the value might vary based on the protein microenvironment. For example, at the active site residues of cysteine proteases and protein tyrosine phosphatases, the pKa value is as low as 2.5. The perturbed pKa renders the cysteine more nucleophilic.

A vast array of probes has been developed to identify and quantify functional cysteines in proteomes[14-17]. Weerapana, Cravatt, and colleagues used alkynylated iodoacetamide probe (Fig.

1.3), which labeled 1,082 of the total 8,910 cysteine sites on 890 human proteins, to profile the intrinsic reactivity of cysteine residues in native proteomes. They discovered cysteines with heightened reactivity ("hyperreactivity") in both catalytic and non-catalytic sites, and that hyper-reactivity can predict cysteine function in both native and designed proteins. They highlighted a highly conserved C93 site in FAM96B, a previously uncharacterized protein, and showed that it is involved in iron-sulphur protein biogenesis and is vital for yeast viability[18]. Unfortunately, the high toxicity of the iodoacetamide (IA) probe limits its use at high concentration in cells. To circumvent this, Weerapana group developed a caged bromomethyl ketone electrophile (Fig. 1.3) with similar coverage as IA probes but low cytotoxicity[19]. They used the probe to monitor cysteine reactivity changes upon epidermal growth factor (EGF)-stimulated release of cellular reactive oxygen species, and they found that cysteines known to form sulfenic acids and redox-active disulfide were less reactive.



Iodoacetamide Bromomethyl ketone

Fig 1.3 Cysteine targeting probes

Because cysteine is involved in many cellular signaling and regulatory processes, such as synthesis of antioxidant glutathione[20-21], metal ion binding[22], and post-translational modifications such as oxidation, S-nitrosylation, palmitoylation, sulfenylation, prenylation and S-GlcNAcylation [23-25], more diverse probes need to be developed to elucidate and distinguish the functions of cysteine residues in the biological system.

1.3.2 Lysine targeting probes

Lysine is the 4th most abundant amino acid residue on protein surfaces (5.9% of all sites in human proteins), after serine, leucine, and alanine[26]. However, lysine residues are more difficult to target, compared to cysteine residues, because the average pKa of the surface-exposed amino group is ~10.4, which means that at physiological pH the lysine side chain amine is entirely protonated. Despite that, several lysine reactive probes have been developed including dichlorotriazines[27-28], imidoesters[29], 2-acetyl- or 2-formyl-benzeneboronic acids[30], isothiocyanates[31], pyrazolecarboxamidines[32], sulfonyl fluorides[33-34], and vinyl sulfonamides[35]. To highlight one example, in 2017 Hacker et al. developed a sulfotetraflouro-phenyl ester (STP) probe that displayed broad and selective reactivity with lysines compared with other amino acids[36]. Using the probe, they quantified a total of more than 9,000 lysines in human cell proteomes, and, similar to the Weerapana et al. cysteine work, they identified hyperreactive lysines enriched at protein functional sites and that can be targeted by electrophilic small molecules. Guided by the STP probe, they designed lysine-reactive fragments that inhibit enzymes in active sites (K89 of NUDT2 and K171 of G6PD), allosteric sites (K688 of PFKP, Fig. 1.4), and also disrupt protein-protein interactions in transcriptional complexes (SIN3A-TGIFs, Fig. 1.4). Of note, SIN3A-TGIFs interaction has been implicated in triple negative breast cancer[37], therefore further optimization of the fragment targeting SIN3A can have important implications in discovering drugs targeting the disease.

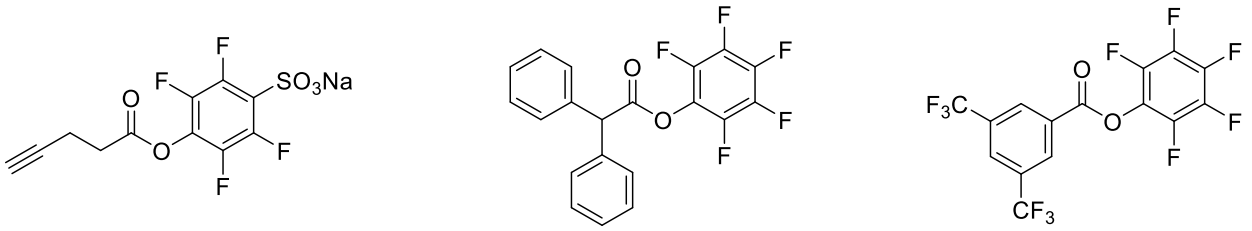
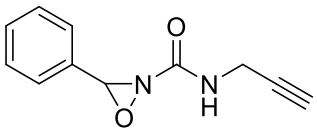


Fig 1.4 Lysine probe sulfotetraflouro-phenyl ester (STP) probe and fragments targeting lysine residues

1.3.4 Methionine targeting probes

Methionine is the 2nd least abundant and among the most hydrophobic amino acid residues, and therefore developing a methionine-selective probe has been challenging. Against the odds, in 2017 Chang group reported an efficient strategy for chemoselective methionine bioconjugation through a redox reactivity using oxaziridine-based reagents (ReACT, Fig. 1.5)[38]. They treated HeLa cells with 1 mM of oxaziridine probe for 10 minutes, and they observed labeling of 235 methionine residues and a single lysine residue. They demonstrated that the ReACT platform can be used to synthesize antibody-drug conjugates (ADCs) with a defined drug-to-antibody ratio in excellent purity, which has been challenging for bioconjugation methods using cysteine and lysine ligation. Moreover, they identified 116 hyperreactive methionines, which, as shown above, are predictive of functional sites. As a representative example, they highlighted a new hyperreactive site Met169 on enolase (corresponding to Met171 on yeast enolase 1), which is a central enzyme in glycolysis pathway and is important in regulating diseases such as cancer via the Warburg effect. They observed that yeast cells carrying Met171 mutation are more resistant to oxidative stress-induced cell death compared to the wild-type strains, corroborating the functional role of that site.



oxaziridine ReACT probe

Fig. 1.5 Methionine Oxaziridine probe

1.3.5 Serine hydrolase targeting probes

Serine hydrolases constitute one of the largest and most diverse enzyme classes in nature, and account for approximately 1 % of proteins in eurkaryotes[39]. They represent enzyme classes such as proteases, peptidases, lipases, esterases, and amidase[5], which are involved in important roles in the human body including inflammation, angiogenesis, neural plasticity, peptide hormone processing, and T-lymphocyte-mediated cytotoxicity[40-41]. Serine hydrolases have been used as a clear illustration of how protein microenvironment influences protein functions. Serine (pKa ~13) at physiological pH is protonated and it will not participate in catalytic reactions. To increase nucleophilicity, all enzymes in the serine hydrolase family contain a catalytic dyad (e.g. Ser-Lys or Ser-Asp) or triad (e.g. Ser-His-Asp or Ser-Ser-Lys) which form an acyl-enzyme intermediate at the active site serine, followed by water hydrolysis to form the final product and regenerate free serine to participate in the next reaction cycle[42]. This phenomenon in which amino acid residues interact with each other leading to some amino acid residues being more nucleophilic and therefore significantly more reactive, broadly defined as "hyperreactive", has also been observed in lysines, cysteines and methionines, and will be further described in chapter 2.

The development of phosphonate probes, i.e fluorophosphonates and arylphosphonates (Fig. 1.6), has led to the identification of more than 80 distinct serine hydrolases that are uncharacterized in terms of their endogenous substrates and targets. Additionally, the probes have been vital to

profiling the activity of serine hydrolases and screening of inhibitors, which has uncovered specialized inhibitor chemotypes, such as lactones/lactams, carbamates, and ureas that inactivate the enzyme by covalently binding to the conserved serine nucleophile[43]. Several drugs targeting serine hydrolases have been approved by the FDA, including rivaroxaban that targets factor Xa for thrombosis[44], Orlistat targeting pancreatic/gastric lipases for obesity(Fig. 1.6)[45-46], dabigatran etexilate targeting thrombin for thrombosis[47-48], and sivelestat that targets human neutrophil elastase for respiratory diseases[49]. Notable inhibitors targeting serine hydrolases in clinical trial include ABX-1431, monoacylglycerol lipase (MGLL) inhibitor in phase 1 and 2a for neuropathic pain and Tourette syndrome (Fig. 1.6)[50-51].

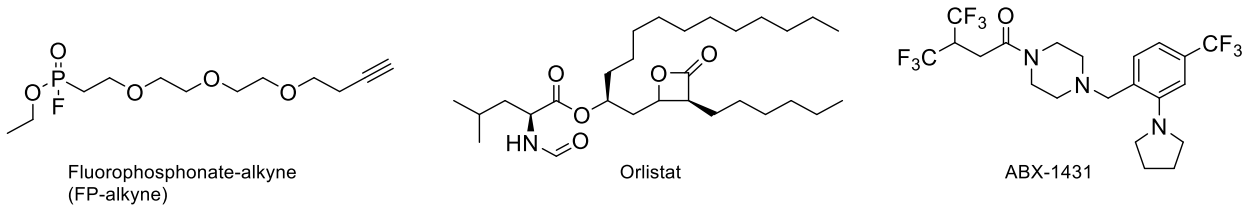


Fig. 1.6. Chemical structure of serine hydrolase probe, FP-alkyne, Orlistat a pancreatic/gastric lipase inhibitor used to treat obesity, and ABX-1431, monoacylglycerol lipase (MGLL) inhibitor in phase 1 and 2a for neuropathic pain and Tourette syndrome

1.3.6 Targeting electrophilic sites on proteins

The above examples highlight ABPP probes targeting nucleophilic functional groups using electrophilic probes. In general, ABPP probes targeting electrophilic residues are largely underexplored. In 2017, Matthews et al. introduced reverse polarity ABPP hydrazine probes as a versatile method to detect and profile electrophilic post-translational modifications (PTMs) and in particular oxidative PTMs[60]. The probes enabled the discovery of N-terminal glyoxylyl modification on the poorly characterized protein secernin-3 (SCRN3). While N-terminal glyoxylyl

modification is shared by other N-terminal nucleophile hydrolases[61-62], they were the first to experimentally verify this modification of SCRN3. Additionally, they showed that the pyruvoyl cofactor of S-adenosyl-L-methionine decarboxylase (AMD1) is dynamically controlled by intracellular methionine concentrations.

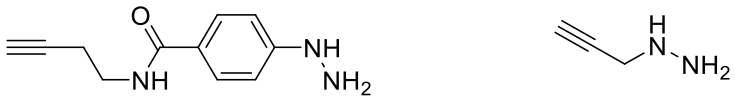


Fig. 1.7 Reverse polarity ABPP hydrazine probes to identify electrophilic protein residues

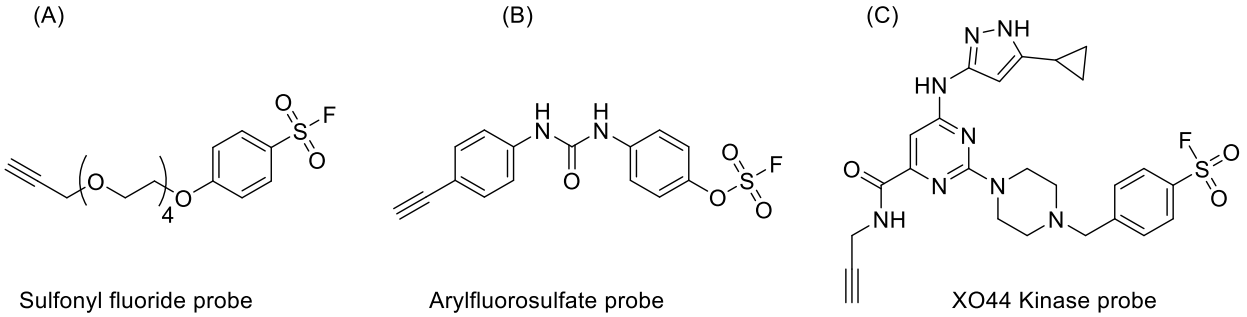
1.4 Sulfur fluoride exchange chemistry (SuFEx) probes

SuFEx and arylfluorosulfate probes (Fig 1.8 A and B) have emerged as versatile tools to profile tyrosine, lysine, serine, threonine, cysteine, and histidine residues. SuFEx chemistry, pioneered by Steinkopf in 1927[52] and revived by Barry Sharpless in 2014[53], is gaining huge interest from drug discovery programs because of its unique properties that enable reactivity with diverse protein nucleophiles in a context-dependent manner. First, unlike sulfonyl chloride, sulfonyl fluorides are resistant to oxidation, reduction, acid/base hydrolysis, and are stable even when exposed to temperatures as high as 130 °C for 3 hours (sulfonyl chloride decomposes after 30 seconds at 130 °C). Secondly, sulfonyl fluoride is involved only in addition-elimination and direct substitution reactions, and for those reactions to proceed the leaving fluoride ion has to be stabilized by either H^+ or R_3Si^+. Moreover, reactions with sulfonyl fluoride occur exclusively at the sulfur atom leading to the formation of only sulfonylation products. In contrast, due to the high polarizability of the chlorine center in sulfonyl chloride, both one- and two-electron pathways are possible leading to a mixture of chlorination and sulfonylation products[53]. Due to the aforementioned

reasons, many applications of SuFEx chemistry have emerged in various fields ranging from chemical biology, synthetic chemistry, to materials science.

SuFEx chemistry has been used to develop covalent small molecule inhibitors targeting a vast array of proteins such as mRNA decapping scavenger enzyme DcpS[54],transthyretin[33], splicing kinase SRPK1[55], and adenosine A_1 receptor[56-57], all of which have been implicated in human disease. SuFEx has also been key for late stage functionalization of drug candidates[58-59]. Recently, Wolan, Sharpless, and coworkers presented a one-step, overnight reaction, high-throughput hit-to-optimization process based on SuFEx chemistry[59]. As a proof of principle, they diversified a hit compound for a bacterial cysteine protease SpeB, benzyl (cyanomethyl)carbamate (IC_{50} = 14 μM, Fig. 1.8D) to 460 analogs in overnight reactions resulting in a drug-like inhibitor with 480-fold higher potency (IC_{50} = 29 nM, Fig 1.8E).

In addition to inhibitors, SuFEx has been used to develop probes to study kinases. Previously, the standard tools to profile kinases were reversible bead-immobilized kinase inhibitors ("kinobeads") or irreversible ATP-biotin probes, which can interrogate 150−200 kinases from a single cell line but are not cell-permeable. Zhao et al. designed a SuFEx probe, XO44, which can profile up to 133 endogenous kinases (Fig 1.8C). Remarkably, 50 kinases captured by XO44 in live cells were not captured by kinobeads in cell lysates[34].



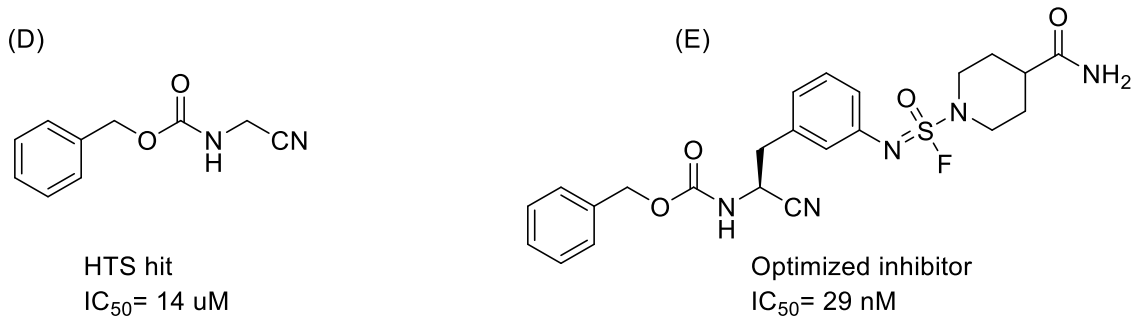


Fig. 1.8 SuFEx probes and applications. (A) and (B) Examples of sulfonyl fluoride and arylfluorosulfate probes to profile tyrosine, lysine, serine, threonine, cysteine, and histidine residues. (C) XO44 probe, a cell-permeable kinase probe. (D) and (E) A high-throughput screening hit against bacterial cysteine protease SpeB and optimized using SuFEx chemistry to yield 480-fold more potent inhibitor.

1.5 Research goals

The probes highlighted above have expanded our understanding of biological processes by serving as tools to ascribe function to the vast number of uncharacterized proteins in the human protein. However, a large fraction of the human proteome remains underexplored. My research focus was to expand chemical tools to improve the understanding of human biology. In particular, I believe the potential of the SuFEx chemistry is limited by the reliance on the fluoride leaving group. To optimize SuFEx probes only the staying group can be modified, which limits the scope of proteins that can be explored. Additionally, the poor chemoselectivity of SuFEx probes (targeting a wide range of amino acid residues including tyrosine, serine, histidine, and lysine) makes SuFEx non-ideal to design chemical inhibitors due to potential off-target effects. In Chapter 2, I will discuss how we replaced the fluorine with triazole to develop tunable sulfur-triazole exchange chemistry (SuTEx) probes to globally profile tyrosine and lysine residues in live cells and lysates. Chapter 3

will focus on how we advanced SuTEx chemistry to develop small molecule inhibitors for proteins involved in important biological functions in humans. Chapter 4 will summarize the findings and outline future directions to improve this work.

References

1. Spring, D. R., Chemical genetics to chemical genomics: small molecules offer big insights. *Chemical society reviews* 2005, *34* (6), 472-482.
2. Kawasumi, M.; Nghiem, P., Chemical genetics: elucidating biological systems with small-molecule compounds. *Journal of Investigative Dermatology* 2007, *127* (7), 1577-1584.
3. Antolin, A. A.; Tym, J. E.; Komianou, A.; Collins, I.; Workman, P.; Al-Lazikani, B., Objective, Quantitative, Data-Driven Assessment of Chemical Probes. *Cell Chem Biol* 2018, *25* (2), 194-205.e5.
4. Oprea, T. I.; Bologa, C. G.; Brunak, S.; Campbell, A.; Gan, G. N.; Gaulton, A.; Gomez, S. M.; Guha, R.; Hersey, A.; Holmes, J., Unexplored therapeutic opportunities in the human genome. *Nature reviews Drug discovery* 2018, *17* (5), 317.
5. Cravatt, B. F.; Wright, A. T.; Kozarich, J. W., Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* 2008, *77*, 383-414.
6. Martell, J.; Weerapana, E., Applications of copper-catalyzed click chemistry in activity-based protein profiling. *Molecules* 2014, *19* (2), 1378-1393.
7. Köhn, M.; Breinbauer, R., The Staudinger ligation—a gift to chemical biology. *Angewandte Chemie International Edition* 2004, *43* (24), 3106-3116.
8. Karver, M. R.; Weissleder, R.; Hilderbrand, S. A., Synthesis and evaluation of a series of 1, 2, 4, 5-tetrazines for bioorthogonal conjugation. *Bioconjugate chemistry* 2011, *22* (11), 2263-2270.
9. Lutz, J. F., Copper-Free Azide–Alkyne Cycloadditions: New Insights and Perspectives. *Angewandte Chemie International Edition* 2008, *47* (12), 2182-2184.
10. Hirsch, J. D.; Eslamizar, L.; Filanoski, B. J.; Malekzadeh, N.; Haugland, R. P.; Beechem, J. M.; Haugland, R. P., Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Analytical biochemistry* 2002, *308* (2), 343-57.
11. Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics* 2002, *1* (5), 376-386.
12. Speers, A. E.; Cravatt, B. F., A tandem orthogonal proteolysis strategy for high-content chemical proteomics. *Journal of the American Chemical Society* 2005, *127* (28), 10018-10019.
13. TMT Quantitation. https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/proteomics-mass-spectrometry/quantitative-proteomics-mass-spectrometry/tmt-quantitation.html (accessed 06/07/2020).
14. Chalker, J. M.; Bernardes, G. J.; Lin, Y. A.; Davis, B. G., Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chemistry–An Asian Journal* 2009, *4* (5), 630-640.

15. Toda, N.; Asano, S.; Barbas III, C. F., Rapid, Stable, Chemoselective Labeling of Thiols with Julia–Kocieński-like Reagents: A Serum-Stable Alternative to Maleimide-Based Protein Conjugation. *Angewandte Chemie International Edition* 2013, *52* (48), 12592-12596.
16. Gunnoo, S. B.; Madder, A., Chemical protein modification through cysteine. *ChemBioChem* 2016, *17* (7), 529-553.
17. Bernardim, B.; Cal, P. M.; Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Albuquerque, I. S.; Perkins, E.; Corzana, F.; Burtoloso, A. C.; Jiménez-Osés, G., Stoichiometric and irreversible cysteine-selective protein modification using carbonylacrylic reagents. *Nature communications* 2016, *7* (1), 1-9.
18. Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F., Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* 2010, *468* (7325), 790-5.
19. Abo, M.; Weerapana, E., A caged electrophilic probe for global analysis of cysteine reactivity in living cells. *Journal of the American Chemical Society* 2015, *137* (22), 7087-7090.
20. Chawla, R. K.; Lewis, F. W.; Kutner, M. H.; Bate, D. M.; Roy, R. G.; Rudman, D., Plasma cysteine, cystine, and glutathione in cirrhosis. *Gastroenterology* 1984, *87* (4), 770-776.
21. Sekhar, R. V.; Patel, S. G.; Guthikonda, A. P.; Reid, M.; Balasubramanyam, A.; Taffet, G. E.; Jahoor, F., Deficient synthesis of glutathione underlies oxidative stress in aging and can be corrected by dietary cysteine and glycine supplementation1–4. *The American Journal of Clinical Nutrition* 2011, *94* (3), 847-853.
22. Baker, D. H.; Czarnecki-Maulden, G. L., Pharmacologic Role of Cysteine in Ameliorating or Exacerbating Mineral Toxicities. *The Journal of Nutrition* 1987, *117* (6), 1003-1010.
23. Maynard, J. C.; Burlingame, A. L.; Medzihradszky, K. F., Cysteine S-linked N-acetylglucosamine (S-GlcNAcylation), a new post-translational modification in mammals. *Molecular & Cellular Proteomics* 2016, *15* (11), 3405-3411.
24. Mieyal, J. J.; Chock, P. B., Posttranslational modification of cysteine in redox signaling and oxidative stress: focus on s-glutathionylation. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA: 2012.
25. Chung, H. S.; Wang, S.-B.; Venkatraman, V.; Murray, C. I.; Van Eyk, J. E., Cysteine oxidative posttranslational modifications: emerging regulation in the cardiovascular system. *Circulation research* 2013, *112* (2), 382-392.
26. Tamura, T.; Hamachi, I., Chemistry for covalent modification of endogenous/native proteins: from test tubes to complex biological systems. *Journal of the American Chemical Society* 2018, *141* (7), 2782-2799.
27. Crawford, L.; Weerapana, E., A tyrosine-reactive irreversible inhibitor for glutathione S-transferase Pi (GSTP1). *Molecular biosystems* 2016, *12* (6), 1768-1771.
28. Shannon, D. A.; Banerjee, R.; Webster, E. R.; Bak, D. W.; Wang, C.; Weerapana, E., Investigating the proteome reactivity and selectivity of aryl halides. *Journal of the American Chemical Society* 2014, *136* (9), 3330-3333.
29. Hunter, M.; Ludwig, M., The reaction of imidoesters with proteins and related small molecules. *Journal of the American Chemical Society* 1962, *84* (18), 3491-3504.
30. Bandyopadhyay, A.; Gao, J., Iminoboronate-based peptide cyclization that responds to pH, oxidation, and small molecule modulators. *Journal of the American Chemical Society* 2016, *138* (7), 2098-2101.

31. Zhang, Y.; Kensler, T. W.; Cho, C.-G.; Posner, G. H.; Talalay, P., Anticarcinogenic activities of sulforaphane and structurally related synthetic norbornyl isothiocyanates. *Proceedings of the National Academy of Sciences* 1994, *91* (8), 3147-3150.
32. Kapp, T. G.; Fottner, M.; Maltsev, O. V.; Kessler, H., Small cause, great impact: modification of the guanidine group in the RGD motif controls integrin subtype selectivity. *Angewandte Chemie International Edition* 2016, *55* (4), 1540-1543.
33. Grimster, N. P.; Connelly, S.; Baranczak, A.; Dong, J.; Krasnova, L. B.; Sharpless, K. B.; Powers, E. T.; Wilson, I. A.; Kelly, J. W., Aromatic sulfonyl fluorides covalently kinetically stabilize transthyretin to prevent amyloidogenesis while affording a fluorescent conjugate. *Journal of the American Chemical Society* 2013, *135* (15), 5656-5668.
34. Zhao, Q.; Ouyang, X.; Wan, X.; Gajiwala, K. S.; Kath, J. C.; Jones, L. H.; Burlingame, A. L.; Taunton, J., Broad-spectrum kinase profiling in live cells with lysine-targeted sulfonyl fluoride probes. *Journal of the American Chemical Society* 2017, *139* (2), 680-685.
35. Asano, S.; Patterson, J. T.; Gaj, T.; Barbas III, C. F., Site-Selective Labeling of a Lysine Residue in Human Serum Albumin. *Angewandte Chemie International Edition* 2014, *53* (44), 11783-11786.
36. Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F., Global profiling of lysine reactivity and ligandability in the human proteome. *Nature chemistry* 2017, *9* (12), 1181-1190.
37. Kwon, Y.-J.; Leibovitch, B. A.; Bansal, N.; Pereira, L.; Chung, C.-Y.; Ariztia, E. V.; Zelent, A.; Farias, E. F.; Waxman, S., Targeted interference of SIN3A-TGIF1 function by SID decoy treatment inhibits Wnt signaling and invasion in triple negative breast cancer cells. *Oncotarget* 2017, *8* (51), 88421.
38. Lin, S.; Yang, X.; Jia, S.; Weeks, A. M.; Hornsby, M.; Lee, P. S.; Nichiporuk, R. V.; Iavarone, A. T.; Wells, J. A.; Toste, F. D.; Chang, C. J., Redox-based reagents for chemoselective methionine bioconjugation. *Science* 2017, *355* (6325), 597-602.
39. Zuhl, A. M.; Mohr, J. T.; Bachovchin, D. A.; Niessen, S.; Hsu, K.-L.; Berlin, J. M.; Dochnahl, M.; López-Alberca, M. P.; Fu, G. C.; Cravatt, B. F., Competitive Activity-Based Protein Profiling Identifies Aza-β-Lactams as a Versatile Chemotype for Serine Hydrolase Inhibition. *Journal of the American Chemical Society* 2012, *134* (11), 5068-5071.
40. Liu, Y.; Patricelli, M. P.; Cravatt, B. F., Activity-based protein profiling: the serine hydrolases. *Proceedings of the National Academy of Sciences* 1999, *96* (26), 14694-14699.
41. Long, J. Z.; Cravatt, B. F., The metabolic serine hydrolases and their functions in mammalian physiology and disease. *Chem Rev* 2011, *111* (10), 6022-63.
42. Rauwerdink, A.; Kazlauskas, R. J., How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: the Serine-Histidine-Aspartate Catalytic Triad of α/β-Hydrolase Fold Enzymes. *ACS catalysis* 2015, *5* (10), 6153-6176.
43. Bachovchin, D. A.; Cravatt, B. F., The pharmacological landscape and therapeutic potential of serine hydrolases. *Nature reviews. Drug discovery* 2012, *11* (1), 52-68.
44. Mueck, W.; Lensing, A. W.; Agnelli, G.; Decousus, H.; Prandoni, P.; Misselwitz, F., Rivaroxaban. *Clinical pharmacokinetics* 2011, *50* (10), 675-686.
45. Kridel, S. J.; Axelrod, F.; Rozenkrantz, N.; Smith, J. W., Orlistat is a novel inhibitor of fatty acid synthase with antitumor activity. *Cancer research* 2004, *64* (6), 2070-2075.
46. Yang, P.-Y.; Liu, K.; Ngai, M. H.; Lear, M. J.; Wenk, M. R.; Yao, S. Q., Activity-based proteome profiling of potential cellular targets of Orlistat− an FDA-approved drug with anti-tumor activities. *Journal of the American Chemical Society* 2010, *132* (2), 656-666.

47. Stangier, J., Clinical pharmacokinetics and pharmacodynamics of the oral direct thrombin inhibitor dabigatran etexilate. *Clinical pharmacokinetics* 2008, *47* (5), 285-295.
48. Hankey, G. J.; Eikelboom, J. W., Dabigatran etexilate: a new oral thrombin inhibitor. *Circulation* 2011, *123* (13), 1436-1450.
49. Zeiher, B. G.; Matsuoka, S.; Kawabata, K.; Repine, J. E., Neutrophil elastase and acute lung injury: prospects for sivelestat and other neutrophil elastase inhibitors as therapeutics. *Critical care medicine* 2002, *30* (5), S281-S287.
50. Fraser, I.; Blankman, J.; Clapper, J.; Grice, C.; O'Neill, G.; Ezekowitz, A.; Thurston, A.; Geenens, E.; Vandermeulen, C.; De Hoon, J. In *Preclinical characterization and first-in-human administration of a selective monoacylglycerol lipase inhibitor, ABX-1431*, Front. Pharmacol. Conference Abstract: EUFEMED, 2017.
51. Cisar, J. S.; Weber, O. D.; Clapper, J. R.; Blankman, J. L.; Henry, C. L.; Simon, G. M.; Alexander, J. P.; Jones, T. K.; Ezekowitz, R. A. B.; O'Neill, G. P., Identification of ABX-1431, a selective inhibitor of monoacylglycerol lipase and clinical candidate for treatment of neurological disorders. *Journal of medicinal chemistry* 2018, *61* (20), 9062-9084.
52. Steinkopf, W., Aromatic sulphuric fluoride. *J. prakt. Chem* 1927, *117*, 1-82.
53. Dong, J.; Krasnova, L.; Finn, M.; Sharpless, K. B., Sulfur (VI) fluoride exchange (SuFEx): another good reaction for click chemistry. *Angewandte Chemie International Edition* 2014, *53* (36), 9430-9448.
54. Hett, E. C.; Xu, H.; Geoghegan, K. F.; Gopalsamy, A.; Kyne Jr, R. E.; Menard, C. A.; Narayanan, A.; Parikh, M. D.; Liu, S.; Roberts, L., Rational targeting of active-site tyrosine residues using sulfonyl fluoride probes. *ACS Chemical Biology* 2015, *10* (4), 1094-1098.
55. Hatcher, J. M.; Wu, G.; Zeng, C.; Zhu, J.; Meng, F.; Patel, S.; Wang, W.; Ficarro, S. B.; Leggett, A. L.; Powell, C. E., SRPKIN-1: A covalent SRPK1/2 Inhibitor that potently converts VEGF from pro-angiogenic to anti-angiogenic ISOFORM. *Cell chemical biology* 2018, *25* (4), 460-470. e6.
56. Beauglehole, A. R.; Baker, S. P.; Scammells, P. J., Fluorosulfonyl-substituted xanthines as selective irreversible antagonists for the A1-adenosine receptor. *Journal of medicinal chemistry* 2000, *43* (26), 4973-4980.
57. Glukhova, A.; Thal, D. M.; Nguyen, A. T.; Vecchio, E. A.; Jörg, M.; Scammells, P. J.; May, L. T.; Sexton, P. M.; Christopoulos, A., Structure of the adenosine A1 receptor reveals the basis for subtype selectivity. *Cell* 2017, *168* (5), 867-877. e13.
58. Liu, Z.; Li, J.; Li, S.; Li, G.; Sharpless, K. B.; Wu, P., SuFEx Click Chemistry Enabled Late-Stage Drug Functionalization. *Journal of the American Chemical Society* 2018, *140* (8), 2919-2925.
59. Kitamura, S.; Zheng, Q.; Woehl, J. L.; Solania, A.; Chen, E.; Dillon, N.; Hull, M. V.; Kotaniguchi, M.; Cappiello, J. R.; Kitamura, S., Sulfur (VI) Fluoride Exchange (SuFEx)-Enabled High-Throughput Medicinal Chemistry.
60. Matthews, M. L.; He, L.; Horning, B. D.; Olson, E. J.; Correia, B. E.; Yates III, J. R.; Dawson, P. E.; Cravatt, B. F., Chemoproteomic profiling and discovery of protein electrophiles in human cells. *Nature chemistry* 2017, *9* (3), 234.
61. Brannigan, J. A.; Dodson, G.; Duggleby, H. J.; Moody, P. C.; Smith, J. L.; Tomchick, D. R.; Murzin, A. G., A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* 1995, *378* (6555), 416-419.

62. Aronson Jr, N. N., Lysosomal glycosylasparaginase: a member of a family of amidases that employ a processed N-terminal threonine, serine or cysteine as a combined base-nucleophile catalyst. *Glycobiology* 1996, *6* (7), 669-675.

Chapter 2: Global targeting of functional tyrosines using sulfur triazole exchange chemistry

Adapted from: Heung Sik Hahm‡, **Emmanuel K. Toroitich**‡, Adam L. Borne‡, Jeffrey W. Brulet‡, Adam H. Libby, Kun Yuan, Timothy B. Ware, Rebecca L. McCloud, Anthony M. Ciancone, and Ku-Lung Hsu

Nature Chemical Biology 16, 150–159(2020)

2.1 Abstract

Covalent probes serve as valuable tools for global investigation of protein function and ligand binding capacity. Despite efforts to expand coverage of residues available for chemical proteomics (e.g. cysteine and lysine), a large fraction of the proteome remains inaccessible with current activity-based probes. Here, we introduce sulfur-triazole exchange (SuTEx) chemistry as a tunable platform for developing covalent probes with broad applications for chemical proteomics. We show modifications to the triazole leaving group can furnish sulfonyl probes with ~5-fold enhanced chemoselectivity for tyrosines over other nucleophilic amino acids to investigate, for the first time, more than 10,000 tyrosine sites in lysates and live cells. We discover tyrosines with enhanced nucleophilicity are enriched in enzymatic, protein-protein interaction, and nucleotide recognition domains. We apply SuTEx as a chemical phosphoproteomics strategy to monitor activation of phosphotyrosine sites. Collectively, we describe SuTEx as a biocompatible chemistry for chemical biology investigations of the human proteome.

2.2 Introduction

Chemical proteomics is a powerful technology for ascribing function to the vast number of uncharacterized proteins in the human proteome[1, 2]. This proteomic method employs probes designed with reactive groups that exploit accessibility and reactivity of binding sites to covalently label active proteins with reporter tags for function assignment and inhibitor development[3]. Selective probes resulting from competitive screening efforts serve as enabling, and often first-in-class, tools for uncovering biochemical and cellular functions of proteins (e.g. serine hydrolases[4], proteases[5], kinases[6], phosphatases[7], and glycosidases[8]) and their roles in contributing to human physiology and disease. The basic and translational opportunities afforded by chemical proteomics has prompted exploration of new biocompatible chemistries for broader exploration of the proteome.

Covalent probes used for chemical proteomics range from highly chemoselective fluorophosphonates for catalytic serines[9] to general thiol alkylating agents and amine-reactive esters of cysteines[10] and lysines[11], respectively. The ability to globally measure protein functional states and selectively perturb proteins of interest has substantially augmented our basic understanding of protein function in cell and animal models[1, 3]. Exploration of new redox-based oxaziridine chemistry, for example, identified a conserved hyper-reactive methionine residue (M169) in redox regulation of mammalian enolase[12]. Hydrazine probes revealed a novel N-terminal glyoxylyl post-translational modification on the poorly characterized protein SCRN3[13]. More recent exploration of photoaffinity probes facilitate global evaluation of reversible small molecule-protein interactions to expand the scope of proteins available for chemical proteomic profiling[14].

Sulfonyl-fluorides[15] (-SO_2F) and fluorosulfates[16, 17] (-OSO_2F) have emerged as a promising scaffold for covalent probe development because of the wide range of amino acids (e.g. serine[18, 19], tyrosine[20], lysine[21], histidine[22]) and diverse protein targets (proteases[18, 19], kinases[21], GPCRs[23]) available for sulfur-fluoride exchange chemistry (SuFEx[24]). Reactivity of SuFEx is driven largely through stabilization of the fluorine leaving group (LG) at protein sites during covalent reaction[25, 26]. The sensitivity of SuFEx to protein microenvironments allows, for example, the ability to target orthogonal nucleophilic residues in the same nucleotide-binding site of decapping enzymes[27]. The broad reactivity and context-dependent activation of SuFEx present opportunities for modulating the sulfur electrophile to target novel, and potentially functional, sites of proteins[21, 25, 26, 28]. The reliance on fluorine, while key for activating SuFEx chemistry, is limiting in terms of LG modifications to modify reactivity, specificity, and binding affinity at protein sites across the proteome.

Here, we introduce sulfur-triazole exchange chemistry (dubbed SuTEx) for development of phenol-reactive probes that can be tuned for tyrosine chemoselectivity in proteomes (>10,000 distinct sites in ~3,700 proteins) through modifications to the triazole LG. We use these probes to discover a subset of tyrosines with enhanced reactivity that are localized to functional protein domains and to apply SuTEx for global phosphotyrosine profiling of pervanadate-activated cells. Our findings illustrate the broad potential for deploying SuTEx to globally investigate tyrosine reactivity, function, and post-translational modification state in proteomes and live cells.

2.3 Results

2.3.1 Design and synthesis of sulfonyl-triazole probes

We reasoned that triazoles could serve as a suitable replacement for the fluorine LG used to promote SuFEx[24]. Previous studies demonstrated that triazoles activate ureas for covalent protein modification with a significant advantage of tunability[29], which is not possible with fluorine as a LG by comparison. We envisioned that a sulfonyl-triazole scaffold would permit evaluation, and potentially control, of reactivity and specificity of the sulfur electrophile through structural modifications to the triazole LG (Fig. 2.1a). Our hybrid probe strategy is further bolstered by the broad functional group tolerance of 1,2,3- and 1,2,4-triazoles as a LG for development of covalent serine hydrolase inhibitors[29, 30].

We developed a general strategy for synthesizing sulfonyl-triazole probes for testing in chemical proteomic assays. To add an alkyne reporter tag for downstream detection, we coupled propargyl-amine to 4-(chlorosulfonyl)benzoic acid to produce an alkyne-modified sulfonyl chloride intermediate (S1) that could be further coupled to either unsubstituted or substituted triazoles (Fig. 2.2). Initially, we synthesized an unsubstituted triazole analog HHS-465 as a starting point for testing LG effects on proteome reactivity (Fig. 2.1b). The N2 isomeric state of HHS-465 was confirmed by NMR and x-ray crystallography (Fig. 2.3). Purity of the N2-isomer was confirmed to be >95% as measured by HPLC. See Supplementary Information for full synthetic procedures and characterization of all sulfonyl probes reported.

2.3.2 Chemical proteomic evaluation of SuTEx chemistry

We established a chemical proteomic method to assess the reactivity of HHS-465 with amino acid residues in proteomes. HEK293T cell proteomes were treated with HHS-465 (100 μM, 1 hr, 25 °C) followed by copper-catalyzed azide-alkyne cycloaddition (CuAAC) coupling with a desthiobiotin-azide tag. Proteomes were digested with trypsin protease and desthiobiotin-modified

peptides enriched by avidin affinity chromatography, released, and analyzed by high-resolution liquid chromatography-mass spectrometry (LC-MS, Fig. 2.1c). Probe-modified peptide-spectrum matches (PSMs) that met our quality control confidence criteria of ≥300 Byonic score[31] and ≤5 ppm mass accuracy were selected for further manual evaluation (see Methods for additional details).

We predicted, based on our proposed reaction mechanism, that amino acid residues modified by HHS-465 would be identified by differential modification with a sulfonyl-desthiobiotin adduct that is the product of SuTEx reaction (Fig. 2.1c). We synthesized and included a 1,2,4-triazole counterpart, HHS-475, for testing to demonstrate SuTEx as a common mechanism among triazole regioisomers (Fig. 2.1b and Fig. 2.2). Initial evaluation of our data assigned >60% of HHS-465- and HHS-475-labeled peptides as uniquely modified tyrosines (Fig. 2.4). Evaluation of MS2 spectra showed confident identification of all major *y*-ions and a large fraction of *b*-ions, including fragment ions (*y* and *b*) that allowed identification of the tyrosine site of HHS-465 and HHS-475 binding (mass adduct of 635.2737 Da, Fig. 2.1d and Fig. 2.5-2.7). The remaining probe-modified peptides were assigned largely to lysines, which after removal of incorrect search algorithm matches to C-terminal modified peptides represented a minor fraction of total modified residues (<25%; Supplementary Fig. 2.4, 2.8, and 2.9). We evaluated additional human cell proteomes to determine the number and type of tyrosines amenable to SuTEx reaction. On average, we reliably identified >2,800 tyrosines per data set and in aggregate, ~8,000 tyrosine sites from ~3,000 proteins with diverse enzymatic and non-catalytic functions across 5 cell proteomes evaluated with HHS-465- and -475 (Fig. 2.10a and b, Supplementary Dataset 1). A large fraction of HHS-465/475-modified sites were also annotated as phosphorylation sites as reported in the PhosphoSitePlus database[32] (Fig. 2.10c).

We next tested whether SuTEx probes exhibit sufficient stability and cell permeability to permit global tyrosine profiling in living systems. We observed robust proteome labeling that was concentration- and time-dependent in fluorescence gel-based analyses of proteomes from HEK293T cells treated with HHS-465 or HHS-475 (Fig. 2.11 and 2.12). Using a saturating probe labeling condition (100 µM, 2 hr, 37 °C) for our live cell studies, we consistently measured ~3,500 distinct tyrosine sites (corresponding to ~1,700 proteins), in total, across membrane and soluble fractions in each cell line tested (HEK293T, Jurkat; Supplementary Dataset 1). For comparison, recent reports using sulfonyl-fluorides showed probe modifications of ~70-130 protein targets in live cell studies[21, 25]. HHS-465- and HHS-475-labeled proteins from live cell profiling were largely absent from the DrugBank database[33] (77%, Fig. 2.10d). Evaluation of probe-enriched domains (Q-values < 0.01) from the non-DrugBank protein (non-DBP) group revealed highly enriched functions that include proteins involved in RNA recognition (RRM domain[34]) and protein-protein interactions (PCI/PINT and SH3 domains[35], Fig. 2.10d). By comparison, the DrugBank protein group (DBP) was largely overrepresented with domains found in enzymes (kinases and redox enzymes, Fig. 2.10d).

2.3.3 Discovery of hyper-reactive tyrosines in human proteomes

Previous studies identified a subset of hyper-reactive cysteine and lysine residues that specify function and are susceptible to binding with electrophilic ligands[10, 11]. Whether tyrosines differ in intrinsic reactivity and the functional implications of heightened nucleophilicity remain largely underexplored on a proteome-wide scale. Here, we used HHS-465 and quantitative chemical proteomics to evaluate tyrosine reactivity directly in human cell proteomes derived from isotopically light and heavy amino acid-labeled HEK293T cells (i.e. stable isotope labeling with

amino acids in cell culture; SILAC[36]). We measured concentration-dependent HHS-465 labeling where nucleophilic tyrosines are expected to exhibit comparable labeling intensity at low and high concentrations of HHS-465 while less nucleophilic tyrosines show concentration-dependent increases in probe labeling. We treated HEK293T proteomes with high versus low concentrations of HHS-465 (250 versus 25 μM; 10:1 comparison) for 1 hr (25 °C) and then analyzed samples by quantitative LC-MS (Fig. 2.13). Tyrosine nucleophilicity was segregated into low, medium, and high groups based on their respective SILAC ratios (SR >5, 2 < SR ≤5, SR ≤ 2, respectively; Fig. 2.14a). We also verified in a control experiment (25 vs 25 μM) that SR values were ~1 in a 1:1 comparison (Fig. 2.15).

In total, we quantified ~2,400 tyrosine residues from >1,100 proteins in soluble proteomes from HEK293T cells that showed consistent SILAC ratios across replicate experiments (n = 4, Fig. 2.14a). The majority of quantified tyrosines showed concentration-dependent increases in HHS-465 labeling, which is indicative of low intrinsic nucleophilicity (Fig. 2.14a). Similar to cysteines and lysine residues, a subset of tyrosines (~5%, 127 sites in total; Fig. 2.14a) demonstrated enhanced nucleophilicity (i.e. hyper-reactivity[10, 11]) as evidenced by SR ≤ 2 for 10:1 conditions (Fig. 2.14a). The majority of proteins contained a single hyper-reactive tyrosine among several tyrosines quantified (Fig. 2.18). Reactive tyrosines (SR < 5) were enriched in domains of enzymes while tyrosines with lower reactivity (SR > 5) were localized at small molecule binding sites (Fig. 2.14b). Comparison of tyrosine reactivity and evidence of phosphorylation revealed a marked inverse correlation. Specifically, tyrosines with low reactivity (SR > 5) were significantly overrepresented for phosphotyrosine sites compared with medium- and hyper-reactive groups (SR ≤ 5, Fig. 2.14c).

We verified our tyrosine reactivity annotations by comparing SuTEx probe labeling of recombinant wild-type (WT) and tyrosine-to-phenylalanine mutants of human proteins with tyrosine sites identified as high (Y8, GSTP1; Y475, EDC3), low/medium (Y417, DPP3), or low hyper-reactivity (Y92, PGAM1). Proteins like glutathione *S*-transferase Pi (GSTP1) with a single hyper-reactive tyrosine, among several modified tyrosines, showed robust HHS-475 labeling that was largely abolished in recombinant Y8F mutant (Fig. 2.14d). Mutation of the hyper-reactive tyrosine in the Yjef-N domain of enhancer of mRNA decapping protein 3 (EDC3) also resulted in near-complete loss of probe labeling (Y475F, Fig. 2.14d). In contrast, mutation of a tyrosine with low nucleophilicity in PGAM1 resulted in negligible alterations in probe labeling (Y92F, Fig. 2.14d). A notable exception was dipeptidyl peptidase 3 (DPP3), which contains a single modified tyrosine (Y417) that, despite a low/medium nucleophilicity ratio (*SR* ~6), showed near-complete blockade of probe labeling in corresponding tyrosine mutants (Y417F, Fig. 2.14d).

Finally, we confirmed the catalytic role of GSTP1 tyrosine 8, located in the GSH binding site (G-site), by mutating this residue (Y8F) and demonstrating abolished biochemical activity (Fig. 2.19a and b, 2.20). In comparison, recombinant DPP3 WT- and Y417F mutant-overexpressed cell lysates showed comparable catalytic activity in a peptidase substrate assay, supporting a non-catalytic role for Y417 (Supplementary Fig. 2.19c and d, 2.21). Future studies will focus on testing whether the moderate reactivity of the non-catalytic Y417 (Fig. 2.14d) can be exploited for DPP3 inhibitor development.

2.3.4 Tuning the triazole LG for tyrosine chemoselectivity

A key advantage of SuTEx technology is the capacity for modifying the triazole LG to tune chemoselectivity of resulting probes. Here, we tested whether we could enhance the specificity of HHS-465/475 for tyrosine modification through addition of functional groups to the triazole (Fig. 2.16a). To globally evaluate probe reactivity and specificity in parallel, we compared the total number of probe-modified sites (Y and K combined) as a function of the ratio of modified tyrosines to lysines (Y/K ratio), respectively, for each SuTEx analog. First, we synthesized a sulfonyl-fluoride counterpart to HHS-465/475, termed HHS-SF-1 (Fig. 2.16a), to directly compare fluoro- and triazole-LGs with respect to proteome specificity and reactivity. HHS-SF-1 exhibited a ~4-fold reduction in the total number of modified sites and lower specificity for tyrosine compared with HHS-465 and HHS-475 (Y/K of 2.3 versus 2.5 and 2.8, respectively; Fig. 2.16a and Fig. 2.17).

In light of the improved tyrosine specificity of HHS-475, we synthesized and evaluated a series of 1,2,4-triazole analogs bearing different substituents at the R2 position (Fig. 2.1a and 2.16a). Addition of a phenyl group improved both tyrosine specificity (Y/K = 3.5) and overall proteome reactivity of the resulting HHS-481 probe (~4,000 total sites; Fig. 2.16a). Modification of the phenyl-triazole resulted in further alterations in proteome activity of SuTEx probes. Addition of a *para*-fluoro substituent (HHS-483) resulted in comparable reactivity and slightly lowered tyrosine specificity compared with HHS-481 (Fig. 2.16a). In contrast, the *para*-methoxy probe HHS-482 showed the highest tyrosine specificity (Y/K ratio of ~5) while maintaining good overall proteome reactivity (~3,000 probe-modified sites, HHS-482; Fig. 2.16a). Evaluation of HHS-482 reactivity against other amino acids revealed high tyrosine specificity with ~75% of probe-modified residues assigned to tyrosines (Fig. 2.16b).

Comparison of tyrosine sites modified by HHS-SF-1 and HHS-482 revealed high overlap (>90%) indicating that substitution of fluorine for a triazole LG did not result in loss of tyrosine coverage (Fig. 2.16c). In contrast, LG modifications to 1,2,4-SuTEx probes furnished analogs that each expanded tyrosine coverage via detection of unique-modified sites (HHS-475, 391 sites; HHS-482, 112 sites; HHS-483, 433 sites; HHS-481, 445 sites; Fig. 2.16d). In summary, our studies highlight a key difference between sulfonyl-fluoride compared with -triazole chemistry; the latter reaction dramatically enhances overall reactivity and through LG modifications can be tuned for enhanced tyrosine chemoselectivity and coverage in proteomes (Fig. 2.16b and d).

2.3.5 Triazole LG enhances phenol reactivity of probes

Next, we compared solution reactivity of sulfonyl probes to evaluate whether the enhanced tyrosine reactivity of SuTEx is a function of the LG or protein microenvironment. We established an HPLC assay to test reactivity of SuTEx and SuFEx probes with nucleophiles that model side chain groups of tyrosine (*p*-cresol) and lysine (*n*-butylamine). We synthesized the predicted products from *p*-cresol (KY-2-48) and *n*-butylamine (KY-2-42) reaction with sulfonyl probes to establish HPLC conditions for monitoring this covalent reaction in solution (Fig. 2.22). We incubated *p*-cresol with a mixture of all three sulfonyl probes and monitored time-dependent reaction by depletion of respective SuTEx (HHS-475, HHS-482) and SuFEx (HHS-SF-1) probe signal. Our probe competition studies were performed with increasing tetramethylguanidine (TMG[37]) base to compare probe reactivity as a function of increasing phenol nucleophilicity. We also measured stability and found that all three sulfonyl probes showed negligible hydrolysis in aqueous and organic solvents even after incubation for 48 hours at room temperature (Fig. 2.23).

At lower TMG (1.1 equivalents), HHS-475 (peak 3) was the most reactive probe as evidenced by consumption by 30 minutes while unreacted HHS-SF-1 (peak 4) and HHS-482 (peak 7) was still detectable (Fig. 2.24a). The difference in reactivity between SuTEx and SuFEx was apparent at higher TMG (2.2 equivalents) conditions. Both SuTEx probes (HHS-475 and HHS-482) were consumed by 10 minutes while HHS-SF-1 was still detectable even after 90 minutes of reaction (Fig. 2.24a); depletion of HHS-SF-1 signal was only observed at the highest TMG tested (3.3 equivalents, Fig. 2.24a and Fig. 2.25). We also verified a similar trend in reactivity when *p*-cresol was incubated with individual sulfonyl probes. The reactivity of all three sulfonyl probes for *n*-butylamine was substantially reduced compared with *p*-cresol even at high TMG (3.3 equivalents) conditions (Fig. 2.24b). Reaction of HHS-475 with *n*-butylamine required 6 hours to complete and HHS-482 and HHS-SF-1 were not consumed even after 24 hours (Fig. 2.24b and Fig. 2.25). To investigate selectivity further, we incubated sulfonyl probes with *n*-butylamine and *p*-cresol mixed in a 5:1 ratio and demonstrated minimal *n*-butylamine- compared with *p*-cresol-probe adduct formation for HHS-475 as well as HHS-482 and HHS-SF-1 (Fig. 2.26).

Collectively, we show the triazole LG enhances intrinsic reactivity of sulfonyl probes for phenol without compromising stability in solvents commonly used for biological experiments (i.e. DMSO). While our solution findings agree with the enhanced reactivity of SuTEx compared with SuFEx observed by proteomics, the differences in tyrosine chemoselectivity between HHS-482 and HHS-475 are likely a function of the protein microenvironment and a feature of probe reactivity that has been reported for other electrophiles[38].

2.3.6 Chemoproteomic profiling of phosphotyrosine activation

Considering the overlap of SuTEx-modified tyrosines with reported phosphotyrosine sites (pY, Fig 2.14c), we investigated whether we could apply this methodology for a "chemical" phosphoproteomics approach. We hypothesized that tyrosine accessibility by SuTEx probes would be inversely correlated with modification status and could be used to identify changes in pY sites (Fig. 2.27). Given the low abundance of phospho-tyrosine (1%) compared with -serine (88%) and -threonine (11%) detected in cell[39] and tissue proteomes[40], we activated global phosphorylation using cell permeable tyrosine phosphatase inhibitors to increase pY signals for our LC-MS studies. Previous live cell studies demonstrated the high efficiency of pervanadate for global inhibition of tyrosine phosphatase activity[41]. We treated live A549 cells with pervanadate at varying concentrations (0 – 500 μM) and time (0 – 30 min) and measured global changes in tyrosine phosphorylation by western blot using a pY-specific antibody (P-Tyr-100[42]). We observed robust increases in global tyrosine phosphorylation as judged by a massive increase in pY-antibody signals that appeared to saturate at 100 μM and 30 min of pervanadate treatment (Fig. 2.28 and 2.29).

Proteomes from cells treated with our pervanadate activation conditions (100 μM, 30 min) were labeled with HHS-475 or HHS-482 (100 μM, 30 min) followed by CuAAC with desthiobiotin and quantitative LC-MS to evaluate how phosphorylation status affected SuTEx probe labeling. Pervanadate blockade of tyrosine phosphatases should activate endogenous phosphorylation and compete for SuTEx probe labeling at phosphorylated- but not unmodified-tyrosine sites that can be differentiated by SILAC ratios of vehicle- (light) versus pervanadate (heavy)-treated cells (Fig. 2.27 and Fig. 2.30a). We detected in total ~2,200 probe-modified tyrosine sites across ~1,000 proteins using both HHS-475 and HHS-482 that were further separated

into pervanadate-sensitive (PerS, *SR* ≥ 2) and -insensitive groups (PerI, *SR* < 2, Fig. 2.30b and Fig. 2.31). In support of our hypothesis, the probe-modified tyrosines found in the PerS group appeared to be enriched for annotated phosphotyrosine sites (HTP ≥10 in PhosphoSitePlus, Fig. 2.30c and Fig. 2.31) and represented only a small fraction of all unique HHS-475- and HHS-482-modified tyrosines detected by chemical proteomics (~3%, 67 sites). The overall median *SR* of all probe-modified tyrosines was ~1 for both HHS-475 and HHS-482 datasets, which supports tyrosine phosphorylation as a rare post-translational event and the ability of our platform to capture subtle changes in the tyrosine phosphoproteome.

To further validate our chemical phosphoproteomics strategy, we tested whether tyrosine sites identified as pervanadate sensitive were also directly phosphorylated under the same treatment conditions. For our studies, we chose several proteins from the PerS group based on a high phosphotyrosine annotation score (HTP >100, PhosphoSitePlus) and evidence for a role in signaling in human cancer cells like A549. We identified signal transducer and activator of transcription 3 (STAT3) as a target protein with reduced SuTEx probe labeling at Y705 (*SR* = 2.3, Fig. 2.30d) that corresponded with enhanced phosphorylation at this site upon pervanadate activation (Fig. 2.30e). Our data are in agreement with previous findings reporting STAT3 Y705 as a phosphorylation site for activation by tyrosine kinases in human non-small cell lung cancer lines including A549[43]. We validated another tyrosine kinase-targeted site (Y228) on catenin δ-1 (CTNND1[44]) and showed blockade of SuTEx probe labeling (*SR* = 3.3, Fig. 2.30d) coincided with direct phosphorylation at this tyrosine site by western blot analysis (Fig. 2.30e).

In contrast, we identified Y105 as a pervanadate insensitive site (*SR* = 1.1, Fig. 2.30d) on pyruvate kinase (PKM) that showed negligible changes in phosphorylation at this tyrosine upon pervanadate activation (Fig. 2.30e). Our proteomic findings support previous reports of substantial

basal levels of phosphorylated-Y105 on PKM in A549 cells[45], which could explain why pervanadate activation did not further enhance pY levels. As a control, we showed that SuTEx probe treatment of pervanadate-activated cell proteomes did not result in non-specific displacement of phosphates from tyrosines (Fig. 2.32). In summary, we applied SuTEx technology as a chemical strategy that exploits probe labeling as a site-specific readout of changes in pY levels upon global activation of the phosphoproteome.

2.4 Discussion

We describe sulfur-triazole exchange chemistry for development of covalent probes that are compatible with biological systems, easily accessible via modern synthetic chemistry, and can be adapted for diverse chemical proteomic applications. We demonstrate, on a proteomic scale, that addition of a triazole LG introduces key capabilities to the sulfur electrophile including tunability for protein reaction, robust cellular activity, and capacity for directing amino acid specificity. Compared with more widely used sulfonyl-fluorides, the triazole LG dramatically enhanced overall reactivity of sulfonyl probes in solution (Fig. 2.24) that can, through modest structural modifications, be optimized for high tyrosine chemoselectivity in proteomes (Fig. 2.16a and b). Key to success is a general synthetic strategy for introducing a common mass spectrometry-stable enrichment tag (Fig. 2.1d) and incorporating diverse triazole LGs to enable global structure-activity relationship (SAR) studies of SuTEx probes directly in lysates and live cells (Fig. 2.10).

We exploit these features of SuTEx for functional studies of >10,000 unique tyrosine sites from ~3,700 protein targets detected in human cell proteomes. While previous chemical proteomic studies have shown promise for functional tyrosine profiling[20, 25, 26, 46], the broad coverage of SuTEx permitted global tyrosine quantitation with unprecedented depth and breadth. A striking

discovery from our studies was high enrichment of tyrosine sites in nucleotide-binding domains from *in vitro* and *in situ* probe-labeling experiments using HHS-465 and HHS-475 (Fig. 2.10b and d). We identified prominent labeling of tyrosines localized in RNA-recognition motifs (RRMs) of serine/arginine-rich protein splice factors (SRSF1-12, ~70% coverage of members by SuTEx) involved in regulation of mRNA splicing, export, and translation[47]. Several probe-labeled tyrosines including Y13 of SRSF3 RRM have been shown through structural studies to directly mediate RNA binding[48]. Combined with prominent *in situ* labeling at domains mediating protein-protein interactions (e.g. PCI/PINT and SH3[35]), SuTEx offers a valuable resource for developing chemical probes against proteins that have been historically challenging to target with small molecules (Fig. 2.10d).

Our functional profiling studies led to the discovery of intrinsically nucleophilic tyrosines that are enriched in enzyme sites but also prominent in domains mediating protein-small molecule and protein-protein interactions (*SR* < 5, Fig. 2.14b). The rare nature of hyper-reactive tyrosines (~5% of all quantified sites) are in agreement with previous chemical proteomic studies that identified minor subsets of cysteine and lysine residues that demonstrate enhanced reactivity[10, 11]. We demonstrated that hyper-reactive residues like Y8 of GSTP1 are key for catalytic function and mutation of this site (Y8F) abolished biochemical activity (Supplementary Fig. 2.19a and b). We also identified a non-catalytic tyrosine near the zinc-binding region of DPP3 (Y417) that exhibited moderate nucleophilicity (*SR* ~6) and may offer future opportunities for developing site-specific ligands (Fig. 2.14d and Fig. 2.19d). We find it noteworthy that several arginines (R548 and R572, Fig. 2.19c) are in close proximity to Y417 and these positively-charged residues may play a role in perturbing the p*Ka* of neighboring tyrosine residues as previously reported for alanine racemase[49]. In contrast with GSTP1 and DPP3 enzymes, the discovery of a hyper-reactive tyrosine

(Y475, Fig. 2.14d) in the Yjef-N domain of the scaffolding protein EDC3 is intriguing given the role of this domain in assembly of cytoplasmic RNA–protein (RNP) granules known as P-bodies involved in post-transcriptional regulation[50]. Future studies will test whether the hyper-reactive nature of the Y475 site can be exploited for developing ligands to modulate EDC3 function.

We applied SuTEx for development of a chemical phosphoproteomics platform to identify and quantitatively measure tyrosine sites whose probe modification status is competed by activation of phosphorylation. As proof of concept, we studied global changes in the tyrosine phosphoproteome under pervanadate activation of A549 cells to identify pervanadate-sensitive (PerS) sites that represented putative phosphotyrosines (Fig. 2.30 and 2.31). Across >2,000 quantified sites, we identified a small subset of PerS sites (67 sites, Supplementary Dataset 1), which is in agreement with the low frequency of tyrosine phosphorylation (1%) compared with more abundant phospho-serines and -threonines[39, 40]. We verified that SuTEx probe labeling is anticorrelated with phosphorylation at Y705 and Y228 of STAT3 and CTNND1, respectively (Fig. 2.30d and e). Both sites are highly annotated phosphotyrosines and reported substrates for tyrosine kinases in cancer cell signaling[43, 44]. In contrast, the pervanadate-insensitive Y105 site of PKM did not show changes in phosphotyrosine signals with pervanadate activation and further supports the ability of SuTEx to differentiate probe labeling of tyrosines based on phosphorylation state (Fig. 2.30d and e). Future studies will focus on further refinement, e.g. improvements to LC-MS method and use of SuTEx probe cocktails, to expand the number and type of phosphotyrosine sites quantified.

In summary, we deployed SuTEx for development of a quantitative chemical proteomics platform to globally profile tyrosine nucleophilicity and post-translational modification state in human cell proteomes. We believe our current findings serve as a blueprint for design of activity-

based probes that can be synthetically modulated to meet the proteomic demands of chemical biology applications. Expansion of our chemical phosphoproteomics to other activation paradigms should afford additional opportunities for studying and potentially targeting tyrosine post-translational modifications in future studies (Fig. 2.10c). The latter effort will be expedited by conversion of SuTEx probes into inhibitors or ligands to reveal the inventory of tyrosine (and potentially phospho-tyrosine) sites that are “druggable” in proteomes.

2.5 Figures

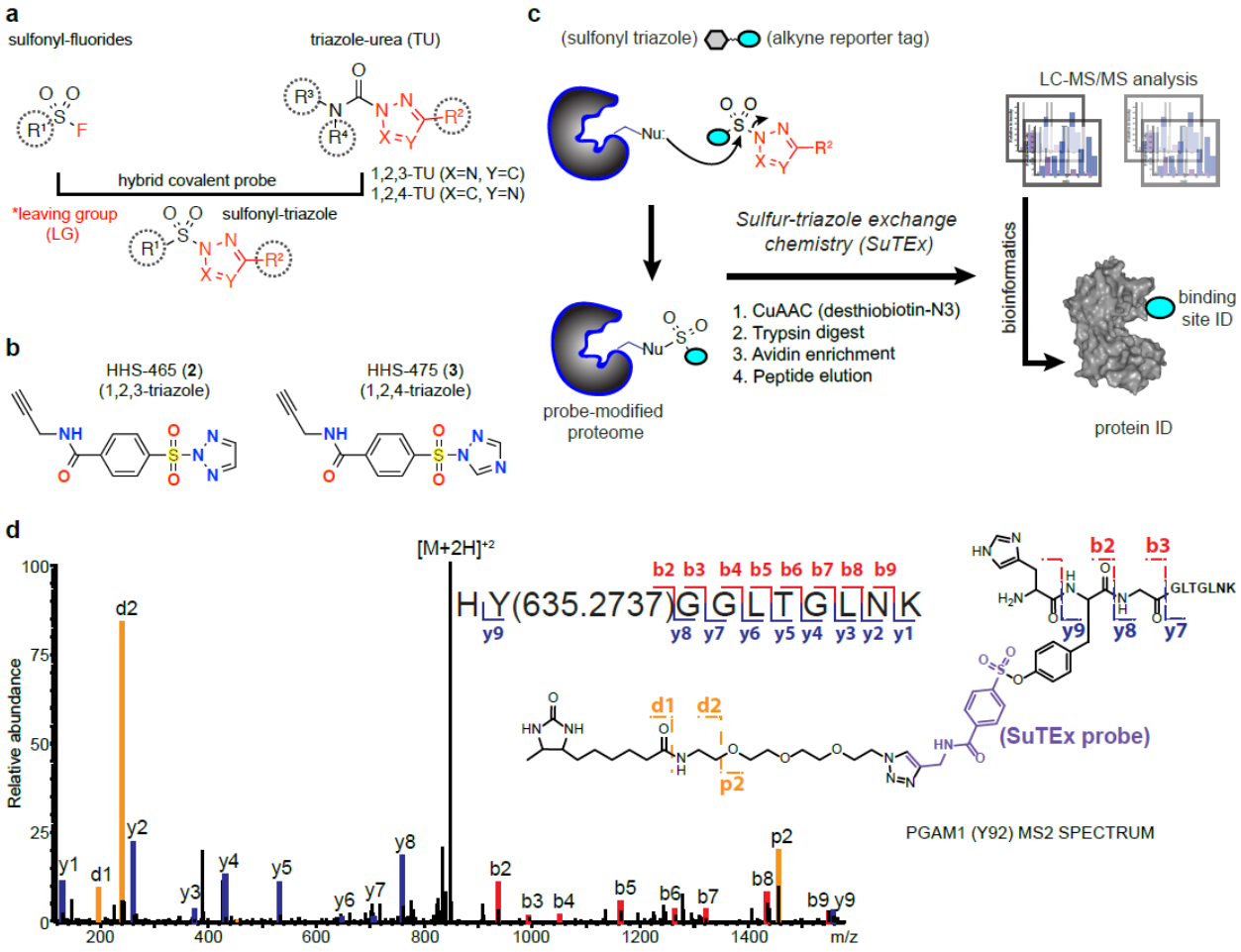


Figure 2.1. **Development of sulfur-triazole exchange (SuTEx) chemistry for chemical proteomics**. a) Sulfonyl-triazoles are a hybrid of sulfonyl-fluoride and triazole-ureas for developing covalent probes with reactivity that can be modulated through the triazole leaving group (LG). b) Chemical structures of 1,2,3- and 1,2,4-sulfonyl triazoles HHS-465 and HHS-475, respectively. c) Proposed reaction mechanism of sulfur-triazole exchange (SuTEx) chemistry and LC-MS/MS workflow to identify proteins and corresponding binding sites from SuTEx reaction. See Methods for additional details. d) MS2 spectrum annotation of an HHS-475-modified tyrosine site (Y92) found in PGAM1. Covalent reaction with HHS-465 and HHS-475 adds +635.2737 Da

to the modified amino acid (Y92 from PGAM1 shown as a representative example) and supports the proposed SuTEx reaction mechanism. Data shown are representative of two experiments (*n*=2 biologically independent experiments).

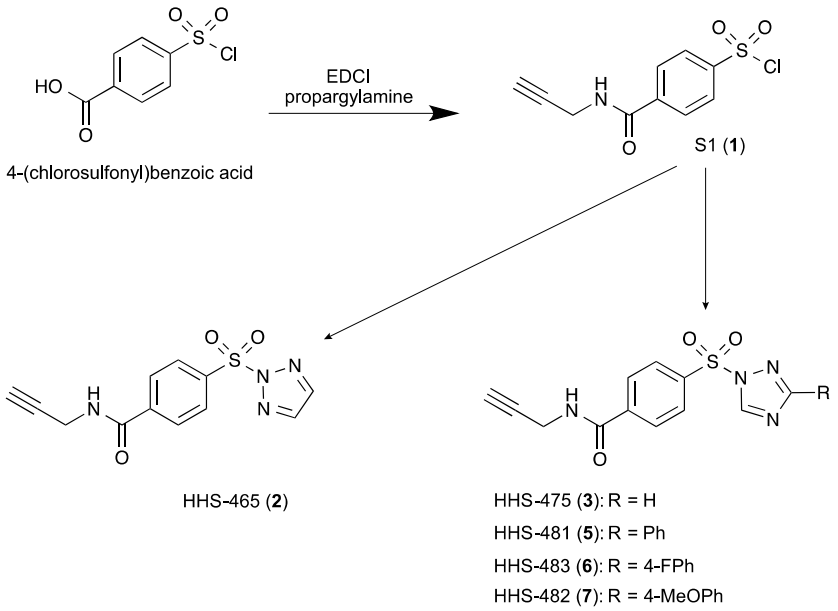


Figure 2.2. Synthetic scheme showing general strategy for developing alkyne-modified sulfonyl-triazole probes.

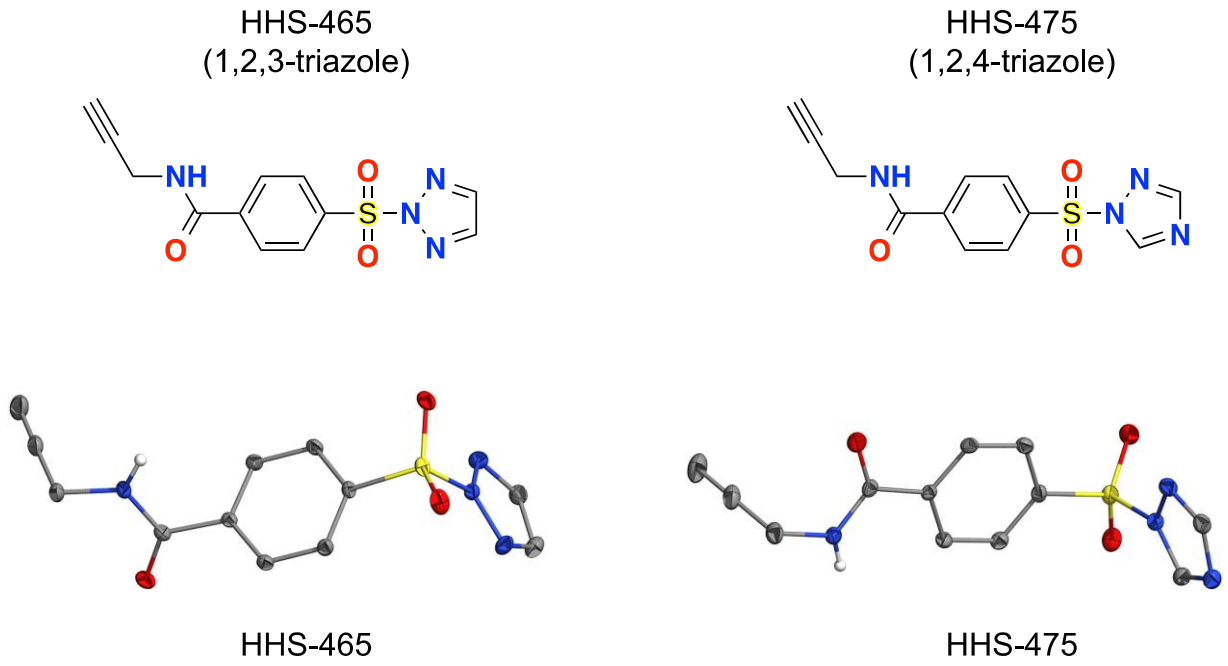


Figure 2.3. Crystal structures of HHS-465 and HHS-475.

We developed a general strategy for synthesizing sulfonyl-triazole probes for testing in chemical proteomic assays. To add an alkyne reporter tag for downstream detection, we coupled propargyl-amine to 4-(chlorosulfonyl)benzoic acid to produce an alkyne-modified sulfonyl chloride intermediate (S1) that could be further coupled to either unsubstituted or substituted triazoles (Fig 2.2). Initially, we synthesized an unsubstituted triazole analog HHS-465 as a starting point for testing LG effects on proteome reactivity (Fig. 2.1b). The N2 isomeric state of HHS-465 was confirmed by NMR and x-ray crystallography (Fig. 2.3). Purity of the N2-isomer was confirmed to be >95% as measured by HPLC.

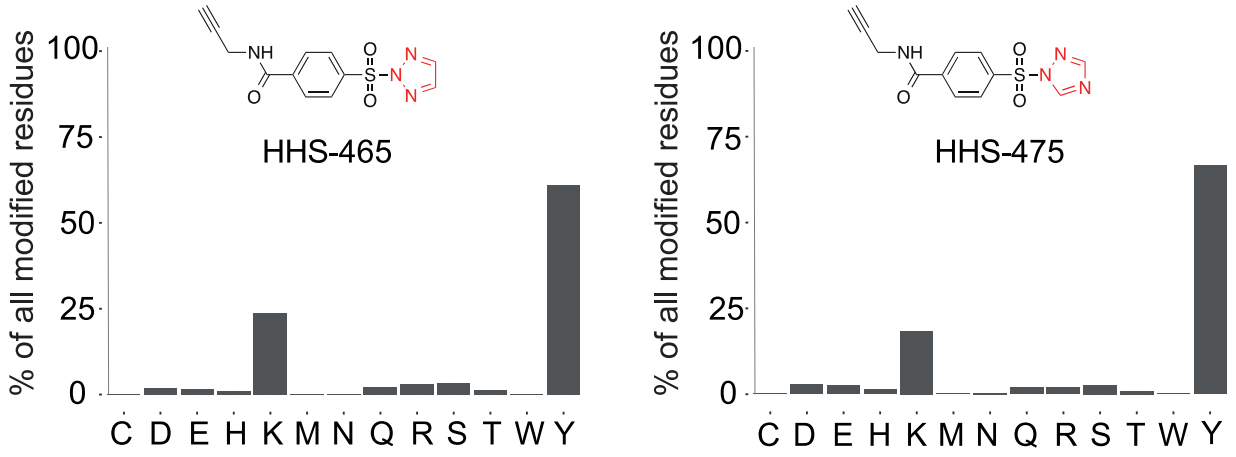


Figure 2.4. Bar plot showing distribution of HHS-465 and HHS-475-modified sites (high confidence sites; Byonic score > 600) against nucleophilic amino acid residues detected in proteomes. Data shown are representative of two experiments (*n*=2 biologically independent experiments).

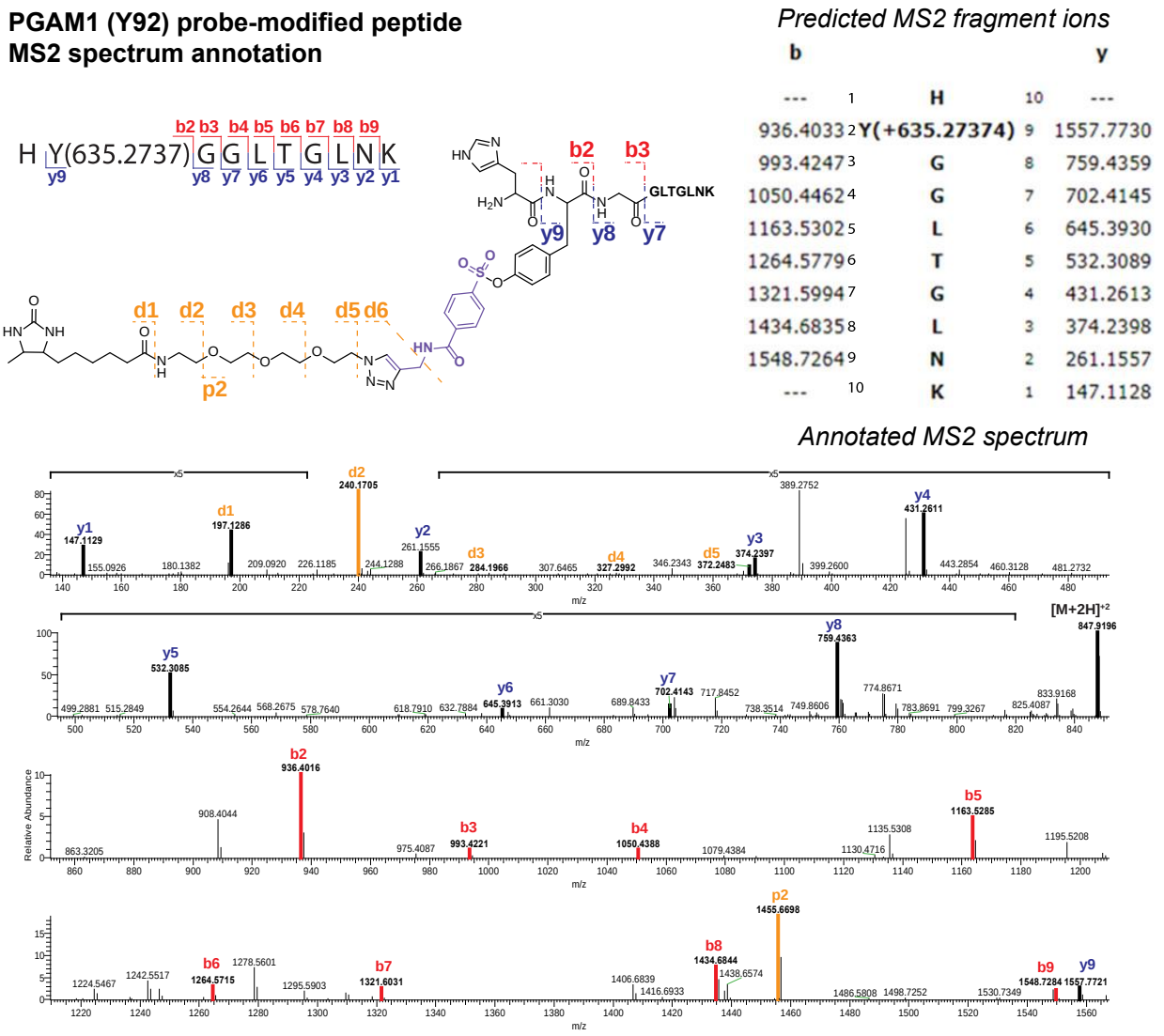


Figure 2.5. **MS2 annotation of the PGAM1 (Y92) HHS-475-modified tryptic peptide**. Left panel: Modified sequence and fragment ion notation of HY*GGLTGLNK (residue 91-100) peptide from PGAM1. Covalent reaction of HHS-475 with Y92 results in a modified tyrosine (Y*) with the addition of +635.2737 Da. Fragmentation of the desthiobiotin-containing tag is also shown. Right panel: predicted MS2 b- and y-fragment ions from collision-induced dissociation (CID) as determined using Protein Prospector software (http://prospector.ucsf.edu/prospector/mshome.htm). Bottom panel: annotation of the MS2 spectrum for the HHS-475-modified PGAM1 Y92 tryptic peptide including fragment ions

containing the probe binding site (modified tyrosine). Data shown are representative of two experiments (*n*=2 biologically independent experiments).

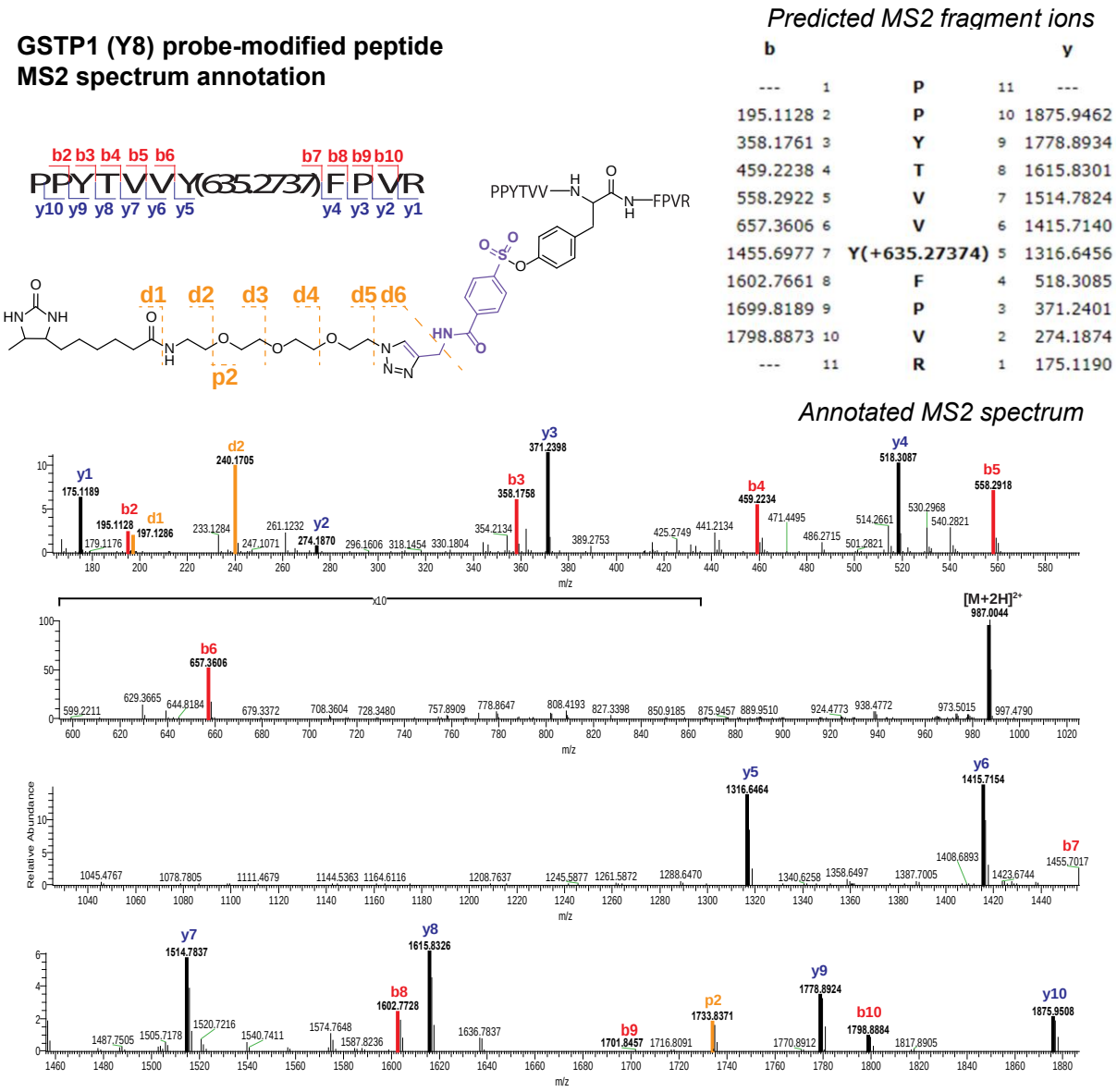
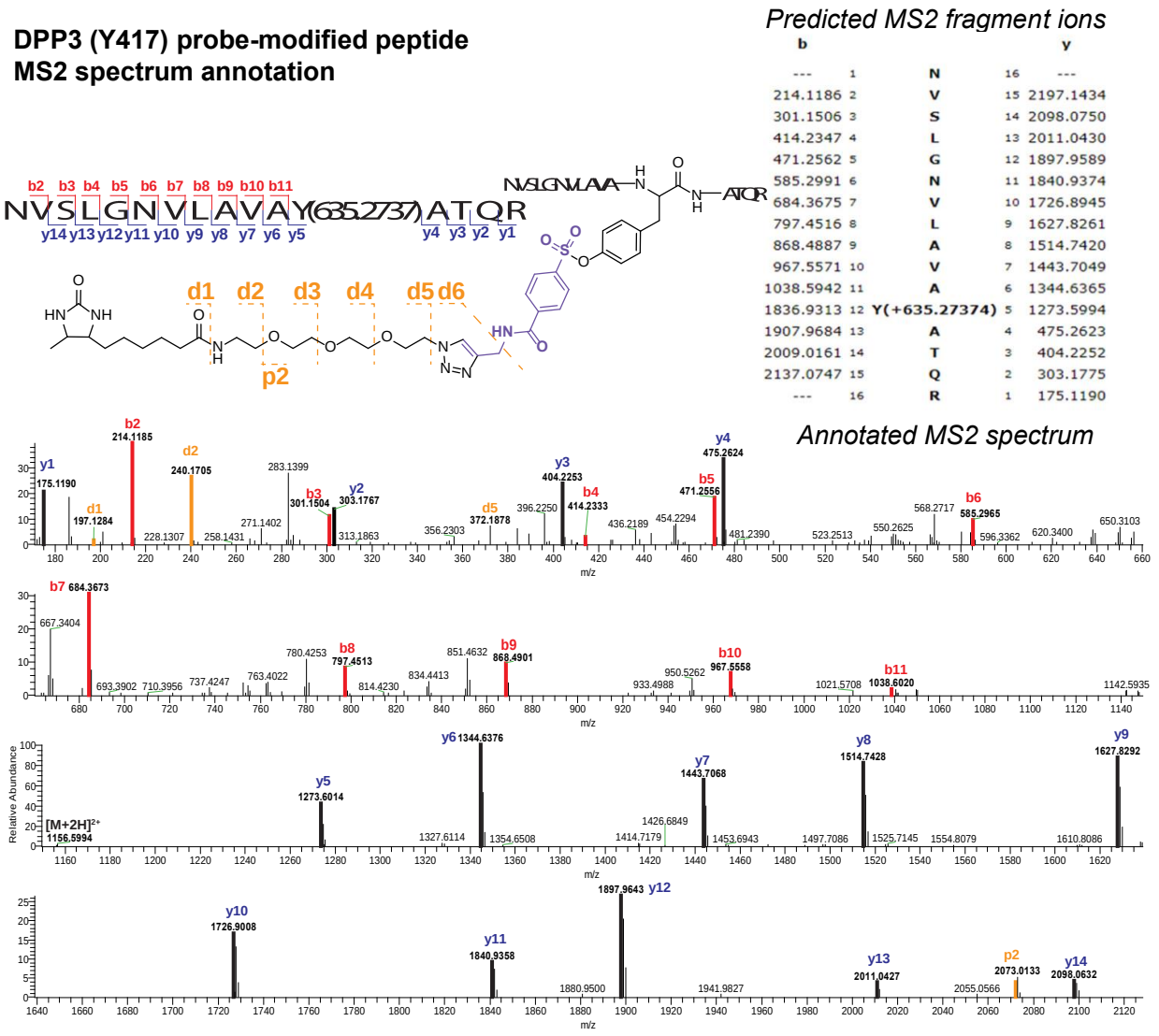


Figure 2.6. **MS2 annotation of the GSTP1 (Y8) HHS-475-modified tryptic peptide**. Left panel: Modified sequence and fragment ion annotation of PPYTVVY*FPVR (residue 2-12) peptide from PGAM1. Covalent reaction of HHS-475 with Y8 results in a modified tyrosine (Y*) with the addition of +635.2737 Da. Fragmentation of the desthiobiotin-containing tag is also shown. Right panel: predicted MS2 b- and y-fragment ions from CID as determined using Protein Prospector software. Bottom panel: annotation of the MS2 spectrum for the HHS-475-modified GSTP1 Y8

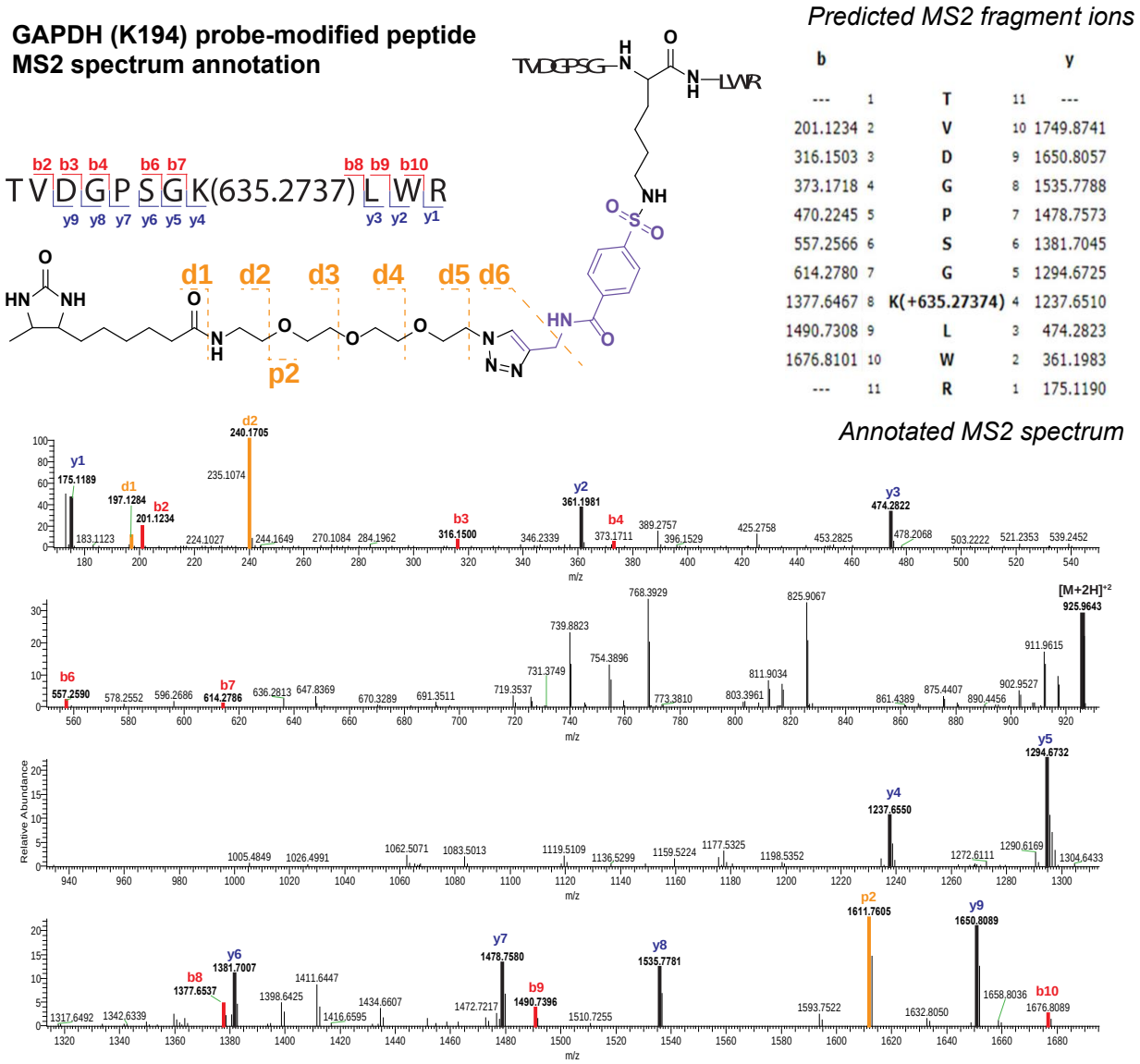
tryptic peptide including fragment ions containing the probe binding site (modified tyrosine). Data shown are representative of two experiments (*n*=2 biologically independent experiments).



b				y
---	1	N	16	---
214.1186	2	V	15	2197.1434
301.1506	3	S	14	2098.0750
414.2347	4	L	13	2011.0430
471.2562	5	G	12	1897.9589
585.2991	6	N	11	1840.9374
684.3675	7	V	10	1726.8945
797.4516	8	L	9	1627.8261
868.4887	9	A	8	1514.7420
967.5571	10	V	7	1443.7049
1038.5942	11	A	6	1344.6365
1836.9313	12	Y(+635.27374)	5	1273.5994
1907.9684	13	A	4	475.2623
2009.0161	14	T	3	404.2252
2137.0747	15	Q	2	303.1775
---	16	R	1	175.1190

Figure 2.7. **MS2 annotation of the DPP3 (Y417) HHS-475-modified tryptic peptide**. Left panel: Modified sequence and fragment ion notation of NVSLGNVLAVAY*ATQR (residue 406-421) peptide from DPP3. Covalent reaction of HHS-475 with Y417 results in a modified tyrosine (Y*) with the addition of +635.2737 Da. Fragmentation of the desthiobiotin-containing tag is also shown. Right panel: predicted MS2 b- and y-fragment ions from CID as determined using Protein Prospector software. Bottom panel: annotation of the MS2 spectrum for HHS-475-modified DPP3

Y417 tryptic peptide including fragment ions containing the probe binding site (modified tyrosine). Data shown are representative of two experiments (*n*=2 biologically independent experiments).



b			y	
---	1	T	11	---
201.1234	2	V	10	1749.8741
316.1503	3	D	9	1650.8057
373.1718	4	G	8	1535.7788
470.2245	5	P	7	1478.7573
557.2566	6	S	6	1381.7045
614.2780	7	G	5	1294.6725
1377.6467	8	K(+635.27374)	4	1237.6510
1490.7308	9	L	3	474.2823
1676.8101	10	W	2	361.1983
---	11	R	1	175.1190

Figure 2.8. **MS2 annotation of the GAPDH (K194) HHS-475-modified tryptic peptide**. Left panel: Modified sequence and fragment ion notation of TVDGPSGK*LWR (residue 187-197) peptide from GAPDH. Covalent reaction of HHS-475 with K194 results in a modified lysine (K*) with the addition of +635.2737 Da. Fragmentation of the desthiobiotin-containing tag is also shown. Right panel: predicted MS2 b- and y-fragment ions from CID as determined using Protein Prospector software. Bottom panel: annotation of the MS2 spectrum for HHS-475-modified

GAPDH K194 tryptic peptide including fragment ions containing the probe binding site (modified lysine). Data shown are representative of two experiments (*n*=2 biologically independent experiments).

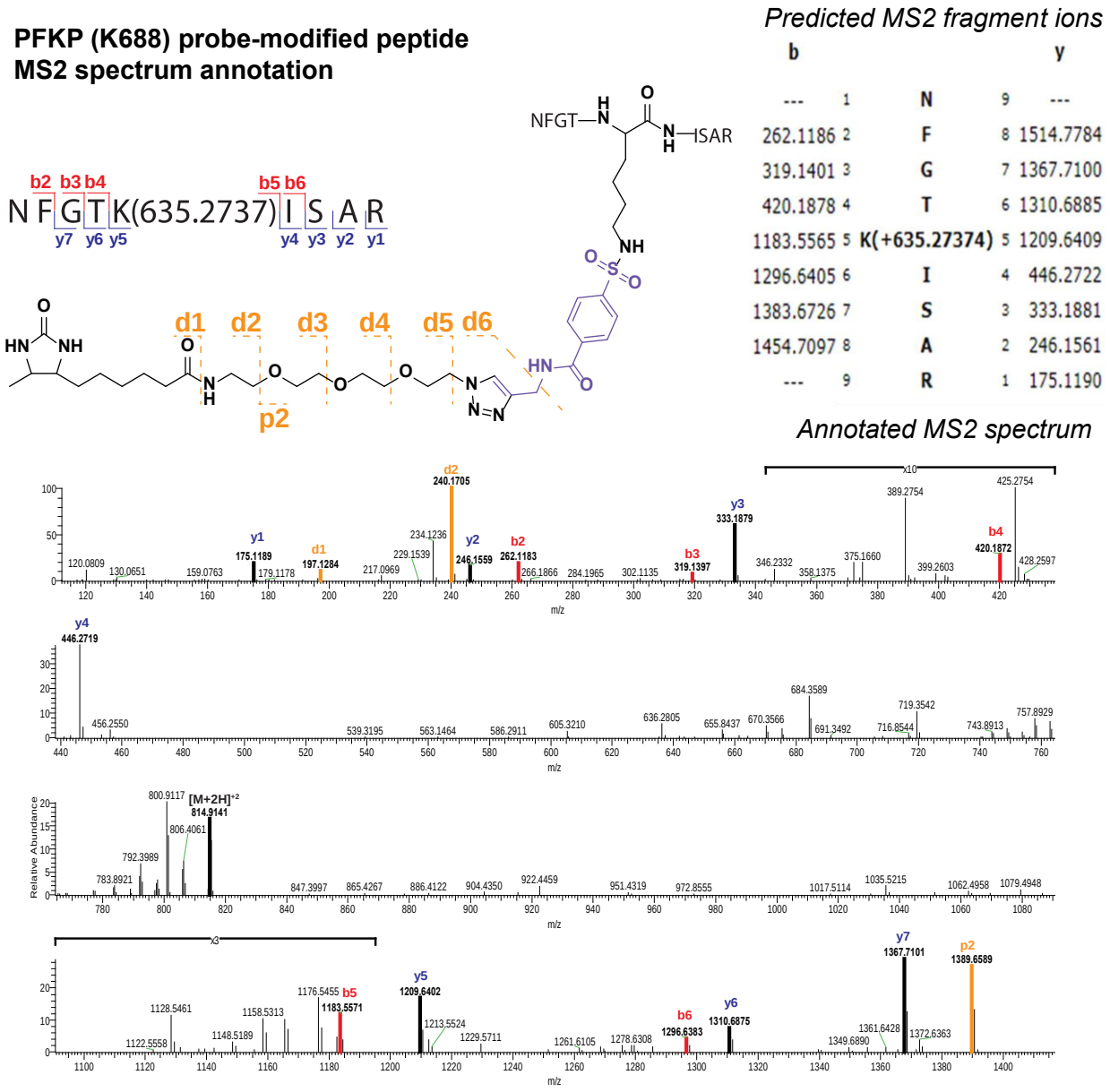


Figure 2.9. **MS2 annotation of the PFKP (K688) HHS-475-modified tryptic peptide**. Left panel: Modified sequence and fragment ion notation of NFGTK*ISAR (residue 684-692) peptide from PFKP. Covalent reaction of HHS-475 with K688 results in a modified lysine (K*) with the addition of +635.2737 Da. Fragmentation of the desthiobiotin-containing tag is also shown. Right panel: predicted MS2 b- and y-fragment ions from CID as determined using Protein Prospector

software. Bottom panel: annotation of the MS2 spectrum for the HHS-475-modified PFKP K688 tryptic peptide including fragment ions containing the probe binding site (modified lysine). Data shown are representative of two experiments (*n*=2 biologically independent experiments).

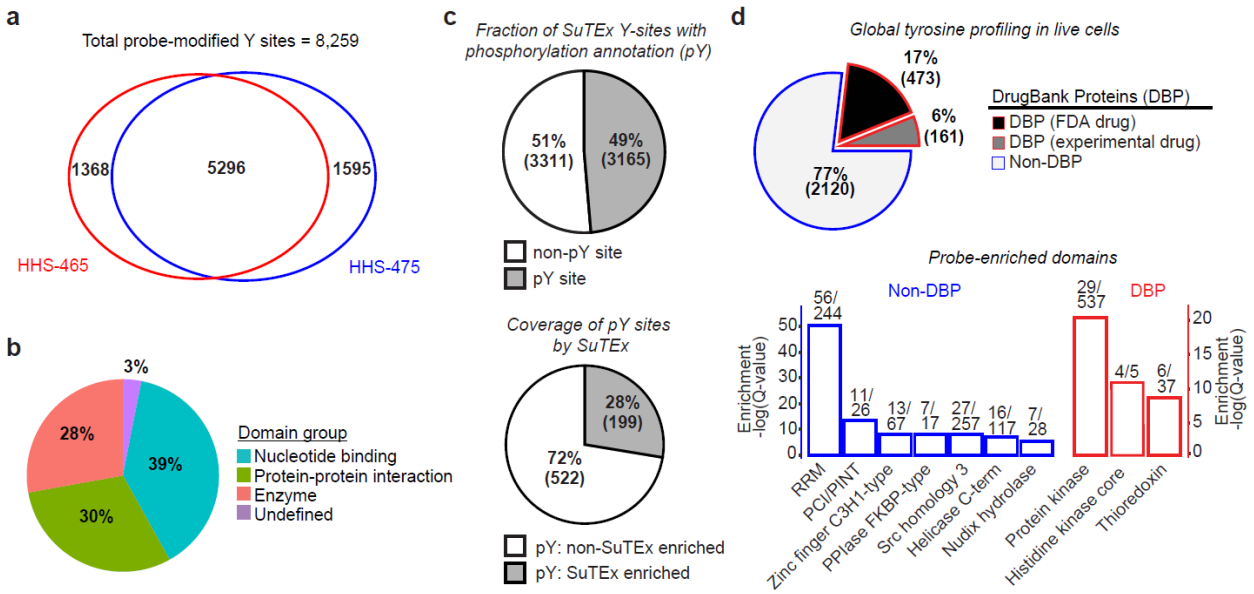


Figure 2.10. **Functional tyrosine profiling in proteomes and live cells**. a) Comparison of HHS-465- and HHS-475-tyrosine modified sites identified from human cell proteomes (HEK293T, A549, DM93, H82, and Jurkat cells) treated with SuTEx probes (100 µM, 1 hr, 25 °C). b) Distribution of protein domain groups that are significantly overrepresented using probe-modified tyrosine sites from *in situ* chemical proteomic studies. Enriched domain annotations are those with a Q-value < 0.01 after Benjamini–Hochberg correction of a two-sided binomial test (see Methods for details). c) Top panel: Overlap between *in situ* HHS-465- and HHS-475-modified tyrosine sites that are also phosphorylation sites (number of phosphotyrosine high throughput annotation on PhosphoSitePlus (HTP score); HTP ≥1). Bottom panel: coverage of phospho-tyrosine sites (HTP ≥10) that were detected by *in situ* chemical proteomics of HEK293T and Jurkat cells (HHS-465 and -475). d) Top panel: Comparison of HHS-465 and HHS-475 *in situ* probe-modified proteins with DrugBank proteins (DBP group). The Non-DBP group consists of proteins that did not match a DrugBank entry. Bottom panel: probe-enriched domains from DBP and non-DBP groups. Enriched domain annotations are those with a Q-value < 0.01 after Benjamini–Hochberg

correction of a two-sided binomial test. All data shown are representative of two experiments (*n*=2 biologically independent experiments).

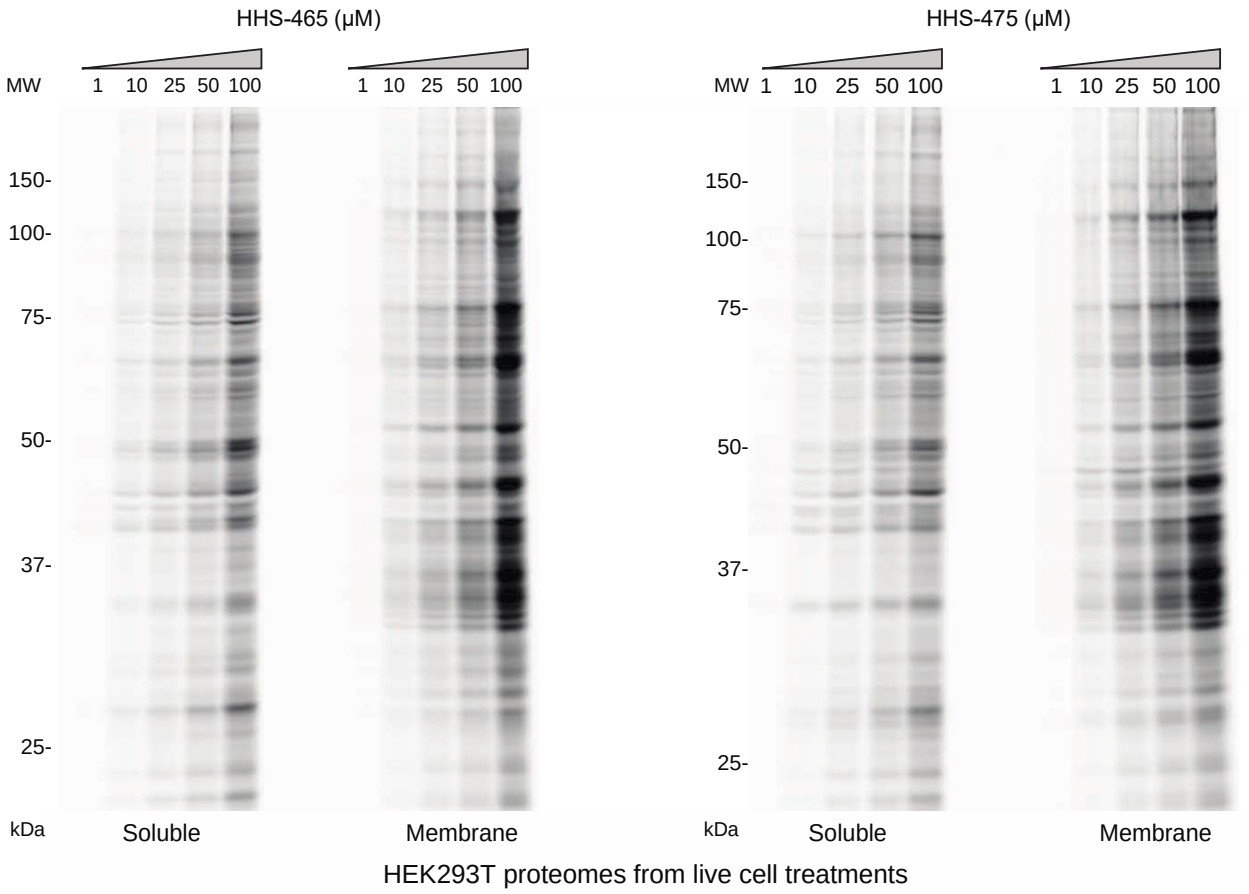


Figure 2.11. **Concentration-dependent labeling of live HEK293T cells treated with SuTEx probes**. HEK293T cells were treated with indicated concentrations of HHS-465 (left panel) or HHS-475 (right panel) for 2 h at 37 °C. After treatment, cells were lysed, probe-modified proteomes (1 mg/mL) subjected to CuAAC with rhodamine-azide followed by SDS-PAGE analysis and in-gel fluorescence scanning. Data shown are representative of two experiments (*n*=2 biologically independent experiments).

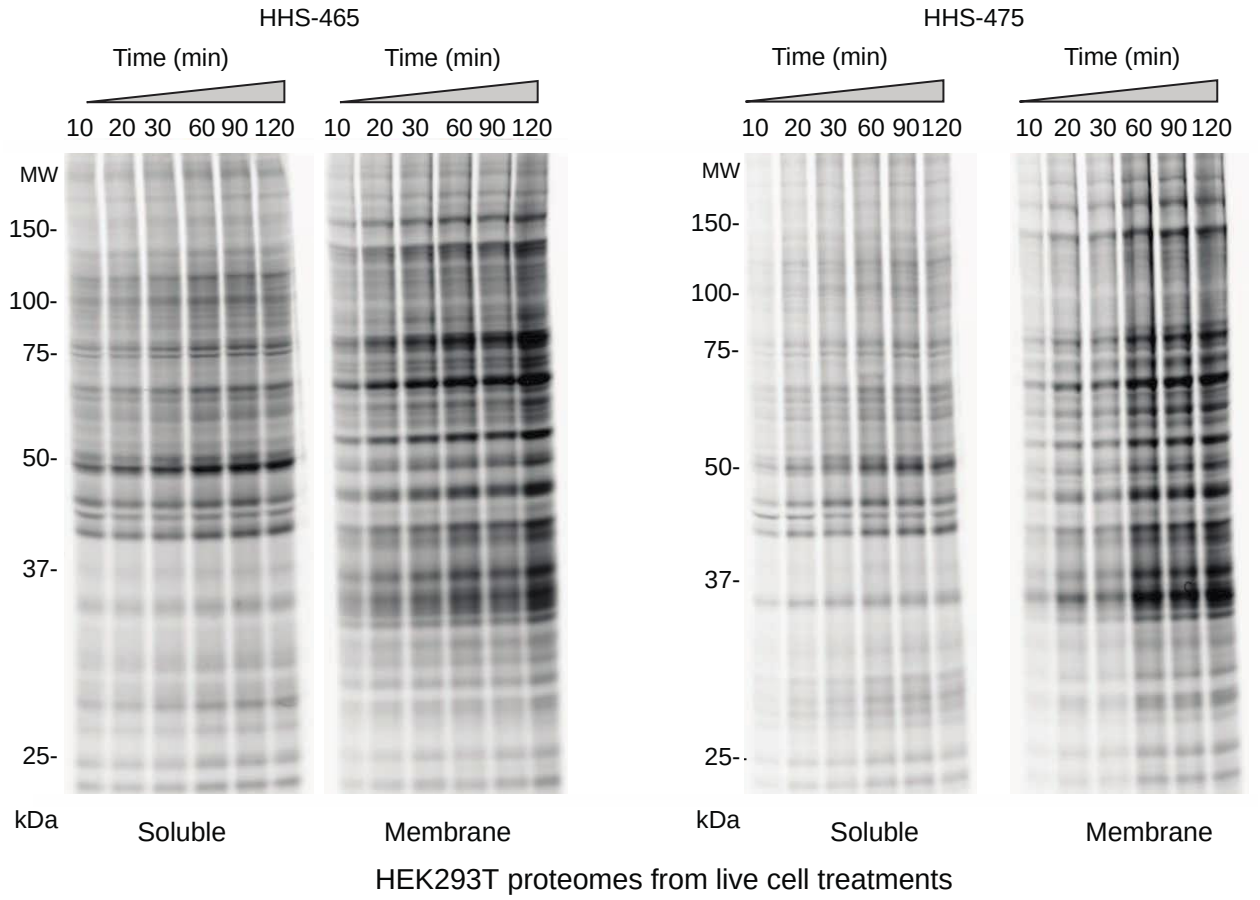


Figure 2.12. **Time-dependent labeling of live HEK293T cells treated with SuTEx probes**. HEK293T cells were treated with 25 μM of HHS-465 (left panel) or HHS-475 (right panel) for the indicated times at 37 °C. After treatment, cells were lysed, and the probe-modified proteomes (1 mg/mL) were subjected to CuAAC with rhodamine-azide followed by SDS-PAGE analysis and in-gel fluorescence scanning. Data shown are representative of two experiments (*n*=2 biologically independent experiments).

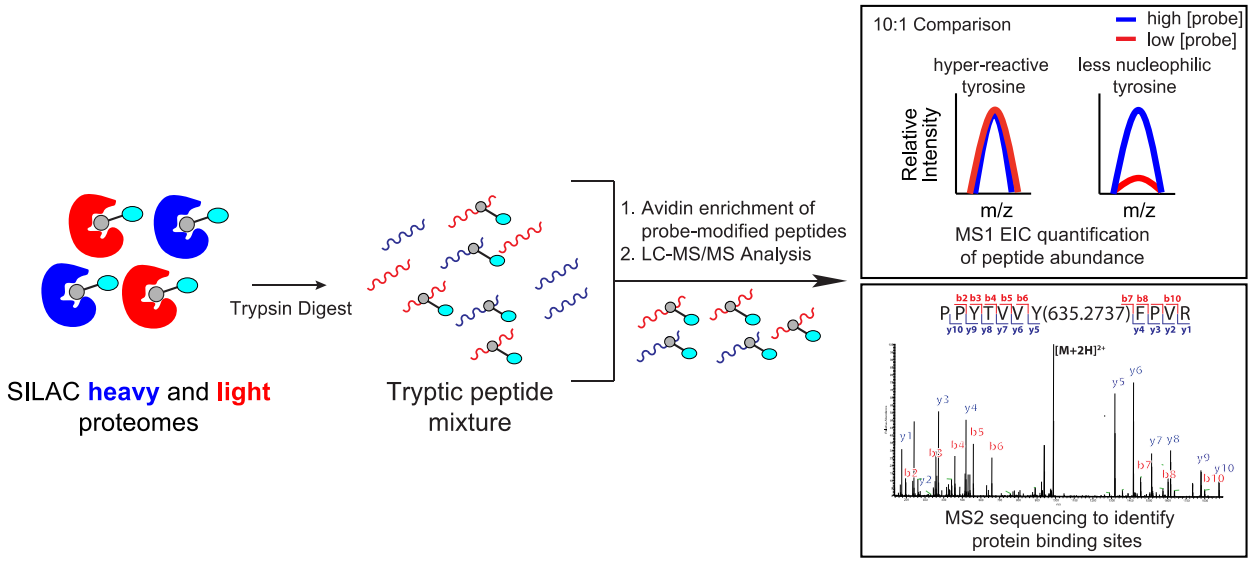


Figure 2.13. **Quantitative chemical proteomics for profiling tyrosine reactivity**. Experimental workflow for quantitative chemical proteomics to measure intrinsic tyrosine nucleophilicity (i.e. reactivity). HEK293T cells were cultured in SILAC media supplemented with either "light" ^{12}C, ^{14}N-labeled lysine and arginine (denoted in red) or "heavy" ^{13}C, ^{15}N-labeled lysine and arginine (denoted in blue). Heavy and light HEK293T proteomes were treated with 250 (high [probe]) or 25 µM (low [probe]) HHS-465, respectively (10:1 comparison). The resulting SILAC ratios (*SR*) were quantified using the area under the curve of MS1 extracted ion chromatograms. Hyper-reactive tyrosines are expected to show equivalent probe labeling intensity at high and low [probe] (left MS1, *SR*~1) while less nucleophilic tyrosines show concentration dependent probe labeling (right MS1, *SR* >>1). A separate experiment where heavy and light proteomes are treated with equivalent [probe] (1:1 comparison) is used as a control for potential false quantifications. Peptide sequencing and validation of the site of probe binding are determined using MS2 (fragmentation) spectra (bottom panel).

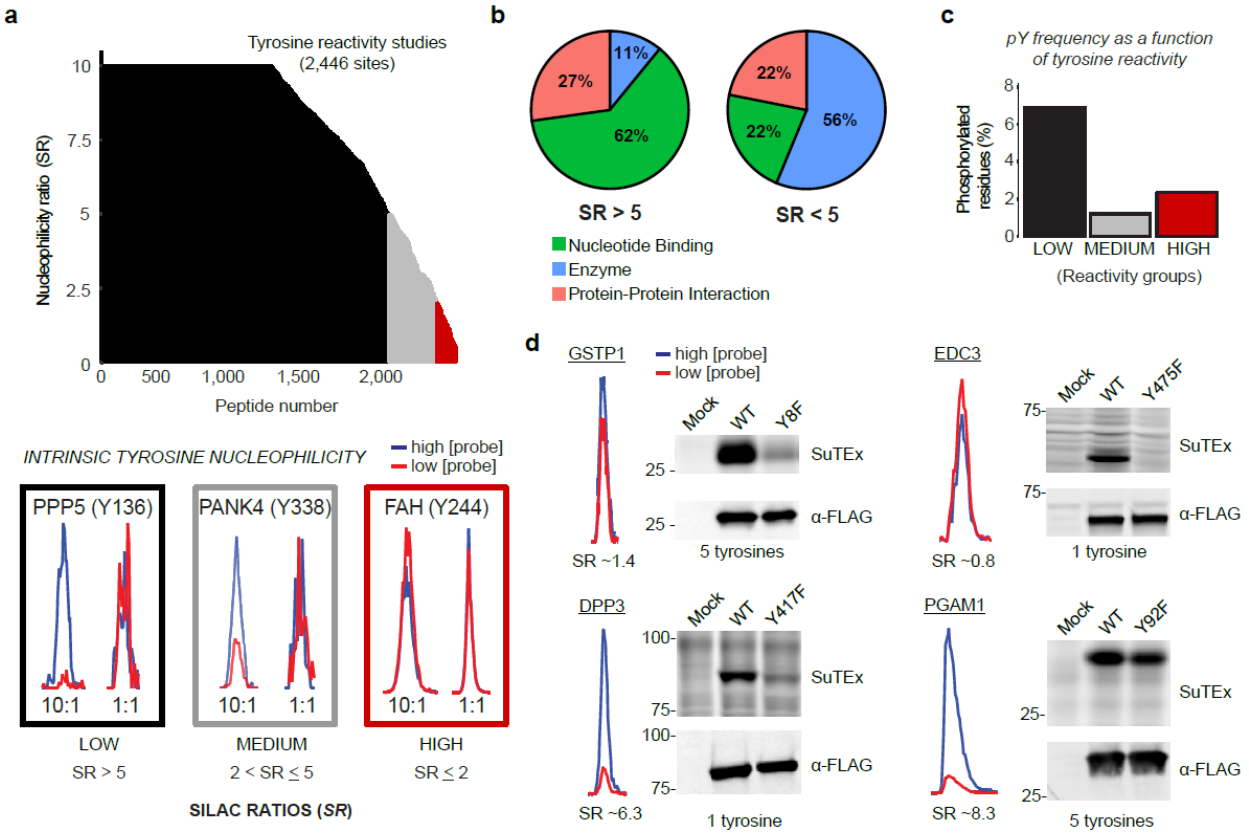


Figure 2.14. **SuTEx-enabled discovery of intrinsically nucleophilic tyrosines in human cell proteomes**. HEK293T SILAC heavy and light soluble proteomes were treated with 250 or 25 µM HHS-465 (10:1 comparison), respectively. The resulting SILAC ratios (*SR*) were quantified using the area under the curve of MS1 extracted ion chromatograms (EIC) to determine tyrosine nucleophilicity. a) A waterfall plot of nucleophilicity ratio (median *SR* values) as a function or probe-modified tyrosine sites to quantitate tyrosine reactivity across the proteome. A MS1 EIC is shown for *SR* values that represent each nucleophilicity group (low-black, medium-grey, and high-red). b) Distribution of protein domain groups that contain tyrosines quantified as low (*SR* >5) or medium/high (*SR* <5) reactivity. Domain annotations shown were significantly enriched (Q-value

< 0.01 after Benjamini–Hochberg correction of a two-sided binomial test) with HHS-465. c) Bar plot depicting tyrosines with medium to high nucleophilicity are less likely to be phosphorylated (HTP ≥10, PhosphoSitePlus) compared with less reactive tyrosines. d) Proteins containing a hyper-reactive tyrosine (GSTP1 Y8, EDC3 Y475) or single probe-modified tyrosine (DPP3 Y417) can be site-specifically labeled with SuTEx probes (50 µM, 30 min, 37 °C). Recombinant wild-type (WT) protein or corresponding tyrosine (Y)-to-phenylalanine (F) mutant HEK293T proteomes were treated with HHS-475 (GSTP1, DPP3, PGAM1) or HHS-465 (EDC3) and analyzed by gel-based chemical proteomics. Proteins that contain less nucleophilic tyrosines (PGAM1 Y92) are labeled at multiple sites and show negligible differences in probe labeling between WT and tyrosine mutant. Western blots show equivalent expression of recombinant WT and mutant proteins. All data shown are representative of two experiments (*n*=2 biologically independent experiments).

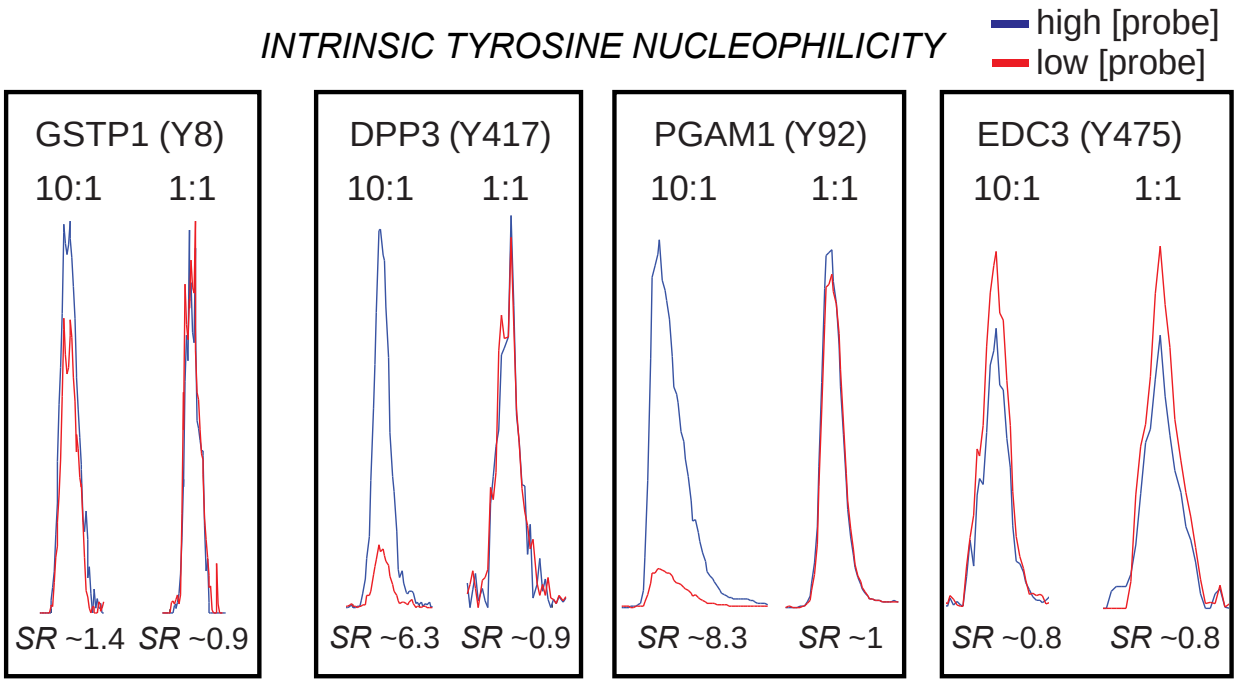


Figure 2.15. **Quantitative analysis of tyrosine reactivity**. Quantitative comparison of tyrosine nucleophilicity between sites detected in human GSTP1 (Y8), DPP3 (Y417), PGAM1 (Y92), and EDC3 (Y475). Heavy and light MS1 extracted ion chromatograms were used to calculate the SILAC ratio (*SR*) for 10:1 and 1:1 probe (HHS-465) comparisons. Data shown are representative of two experiments (n=2 biologically independent experiments).

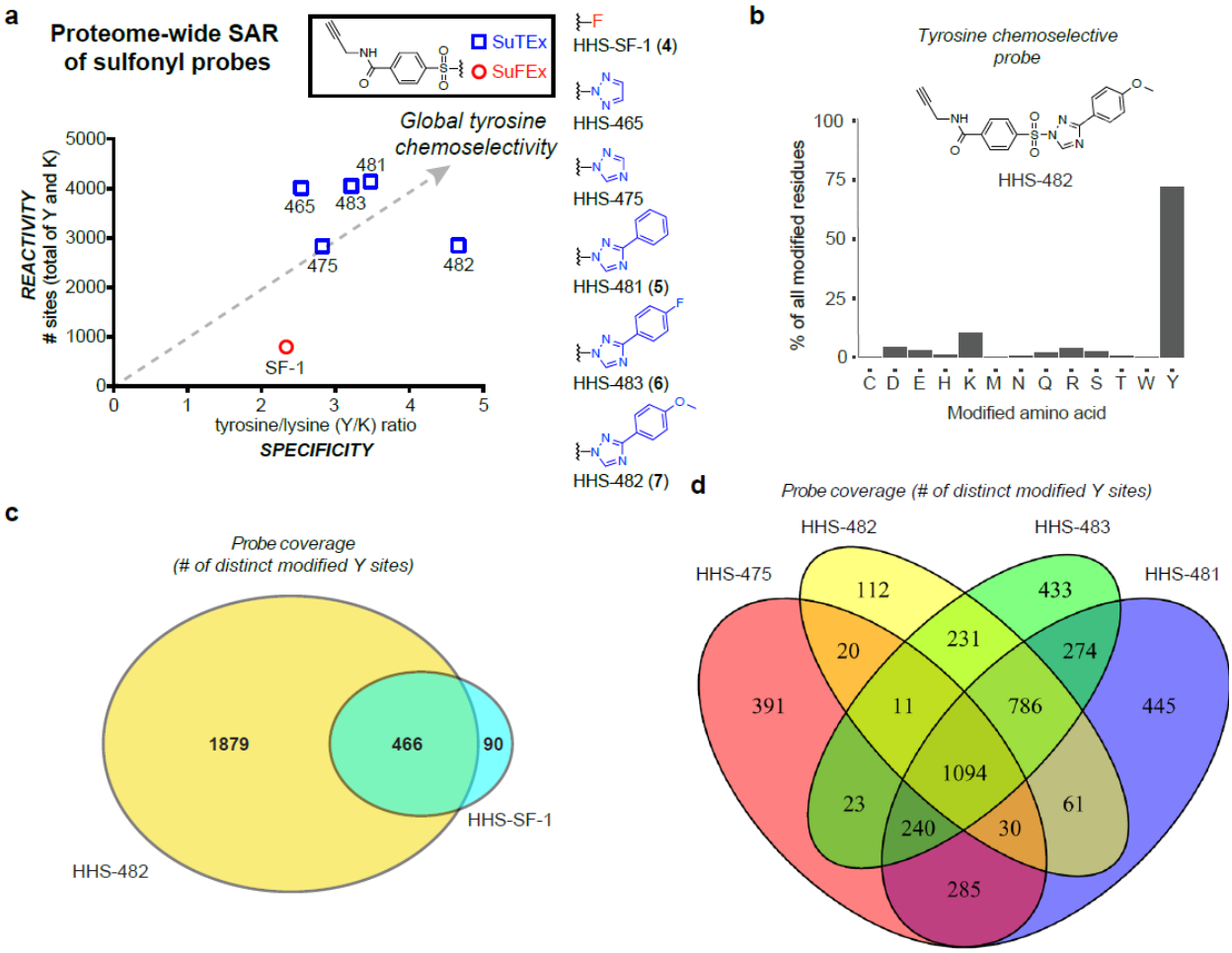


Figure 2.16. **Tuning SuTEx probes for tyrosine chemoselectivity in cell proteomes**. HEK293T soluble proteomes were treated with SuFEx and SuTEx probes. a) Global reactivity [total number of tyrosine (Y) and lysine (K) sites] and specificity (Y/K ratio) of probe-labeled sites from LC-MS chemical proteomic experiments. A bar graph depiction of reactivity and selectivity data can be found in Fig. 2.17b) Bar plot showing distribution of HHS-482-modified sites (high confidence sites; Byonic score > 600) against nucleophilic amino acid residues detected in proteomes. c) High overlap of tyrosine-modified sites from proteomes treated with sulfonyl-triazoles (HHS-482) compared with -fluorides (HHS-SF-1). d) Comparison of probe-modified tyrosine sites from LC-MS chemical proteomic studies using 1,2,4-sulfonyl-triazoles. Each 1,2,4-sulfonyl-triazole probe

was able to modify unique tyrosine sites to increase overall tyrosine coverage. All data shown are representative of two experiments (*n*=2 biologically independent experiments).

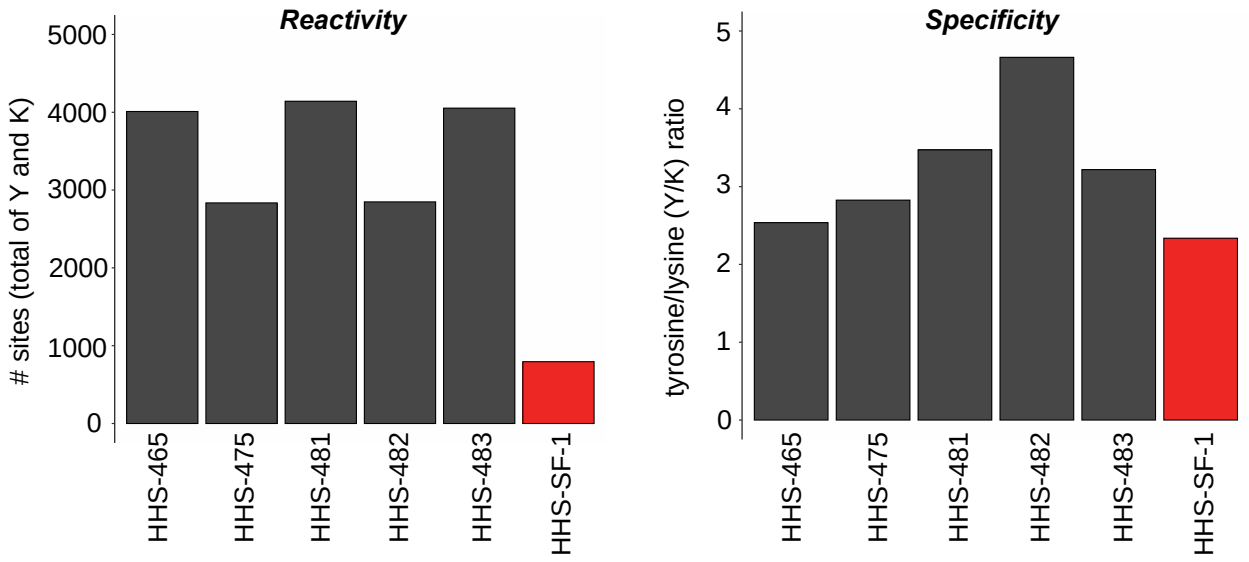


Figure 2.17. Bar plot depiction of global tyrosine reactivity and selectivity of SuTEx and SuFEx probes shown in Fig 2.16a.

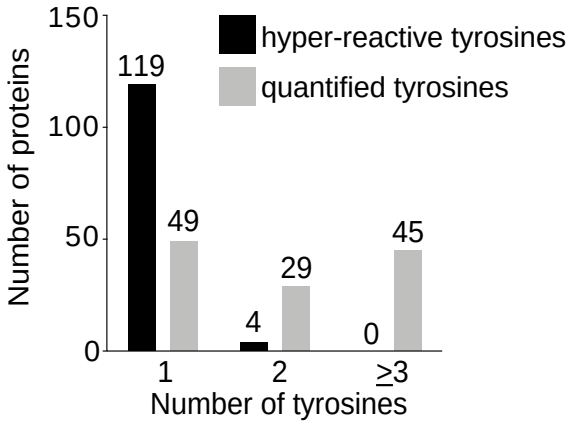


Figure 2.18. Bar plot of the number of hyper-reactive (high nucleophilicity) and quantified tyrosines per protein that contained at least a single hyper-reactive tyrosine. Data shown are representative of two experiments (*n*=2 biologically independent experiments).

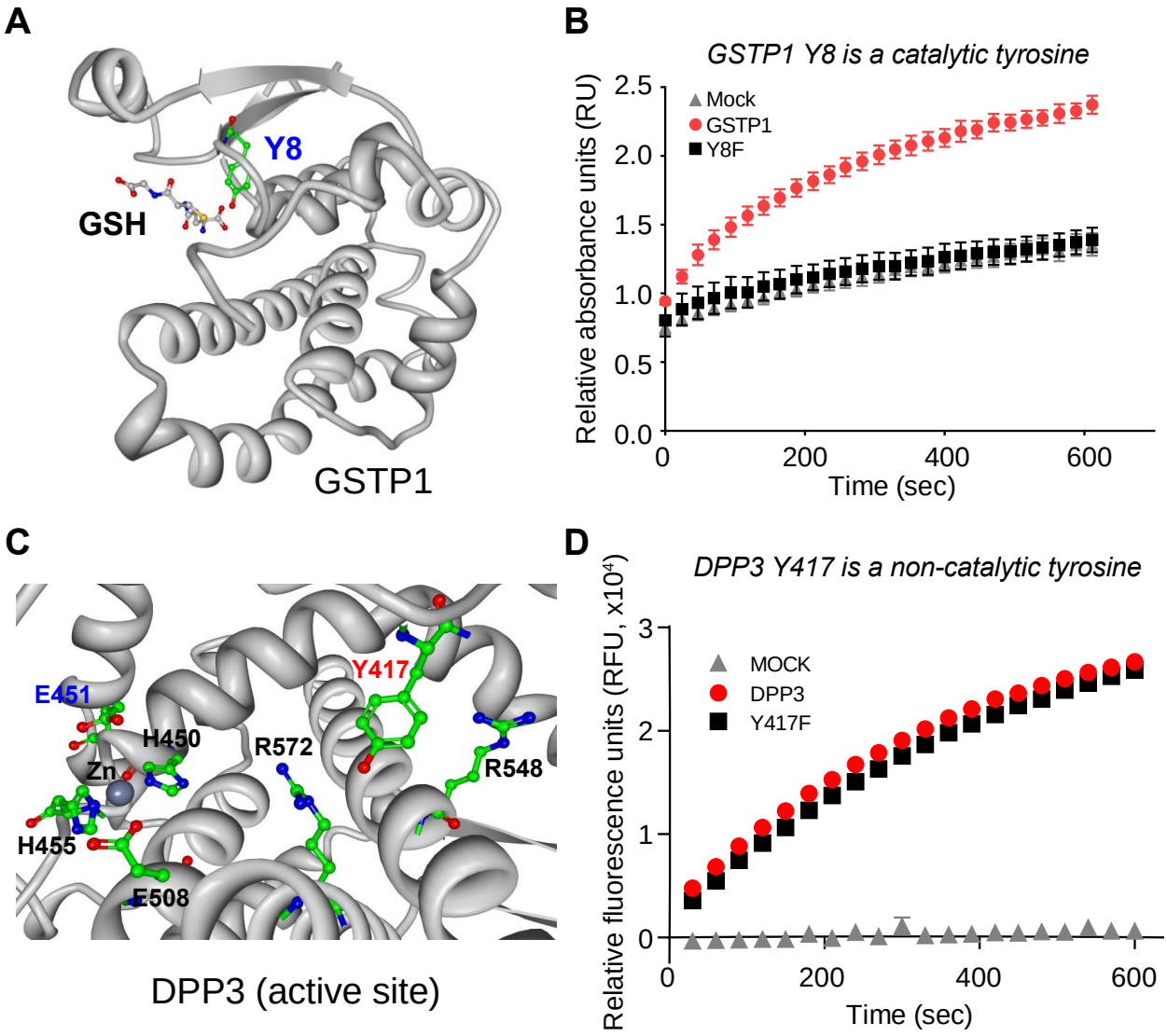


Figure 2.19. **SuTEx probes target reactive catalytic and non-catalytic tyrosines of enzymes**. A) Crystal structure of human GSTP1 (grey, PDB accession code 6GSS) shows tyrosine 8 (Y8) is located in the GSH binding site. B) Loss of biochemical activity in GSTP1 Y8F mutant supports tyrosine 8 as a catalytic residue. Biochemical activity of recombinant GSTP1-HEK293T proteomes (1 mg/mL) was assessed using a substrate assay measuring GSTP1-catalyzed conjugation of GSH to BDNB (10 min, 37 °C). See Fig. 2.20 for additional details. Data are shown as mean ± s.e.m,; *n*=7 biologically independent experiments. C) Crystal structure of human DPP3

(grey, PDB accession code 3FVY) showing location of residues involved in zinc metal binding (H450, H455, E508), the catalytic glutamate (E451), and a non-catalytic tyrosine 417 (Y417) identified by SuTEx. Positively-charged arginines (R548, R572) are found in close proximity to Y417. D) Recombinant DPP3- and Y417F mutant-HEK293T soluble proteomes (1 mg/mL) showed comparable activity in a peptidase substrate assay supporting Y417 as a non-catalytic tyrosine. Data are show as mean $\pm$ s.e.m,; *n*=4 biologically independent experiments.

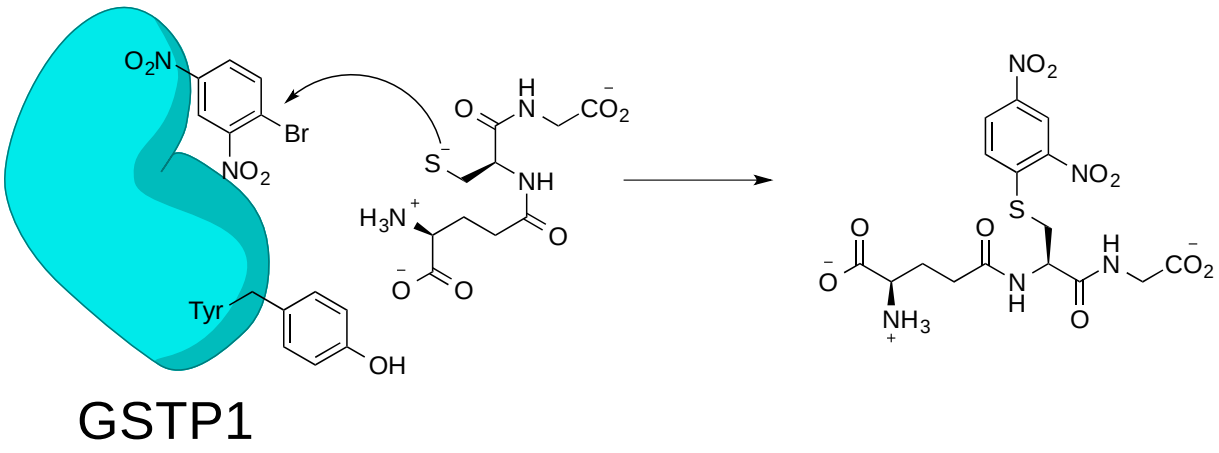


Figure 2.20. GSTP1 biochemical substrate assay. GSTP1 catalytic activity was evaluated by monitoring transfer of glutathione (GSH) to 1-bromo-2,4-dinitrobenzene (BDNB), which produces a dinitrophenyl thioether that can be detected spectrophotometrically by measuring absorbance at 340 nm.

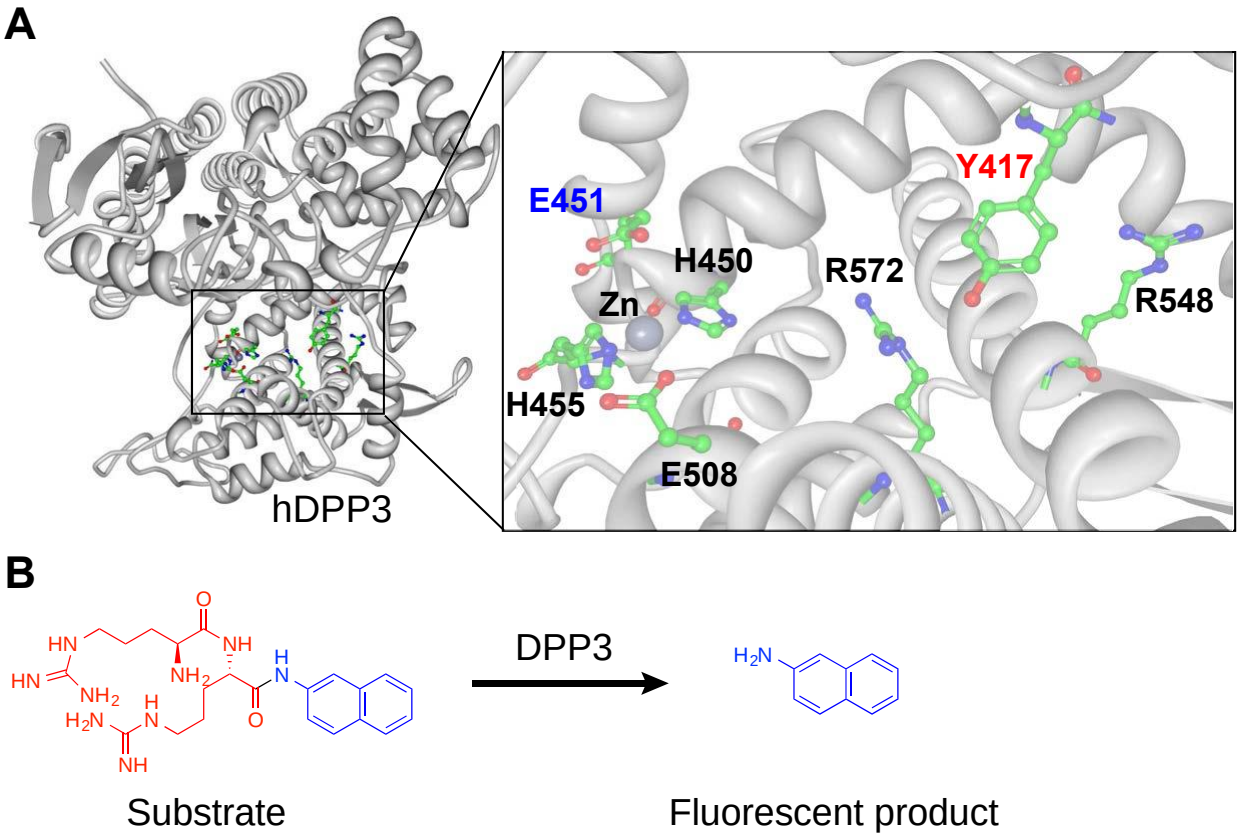


Figure 2.21. Substrate assay for evaluating DPP3 tyrosine 417 mutant. A) Crystal structure of human DPP3 (grey, PDB accession code 3FVY) showing location of residues involved in zinc metal binding (H450, H455, E508), the catalytic glutamate (E451), and a non-catalytic tyrosine 417 (Y417) identified by SuTEx. Positively-charged arginines (R548, R572) are found in close proximity to Y417. B) DPP3 cleaves Arg-Arg β-naphthylamide substrate to release the colored naphthylamide product that can be detected spectrophotometrically by measuring fluorescence at 450 nm.

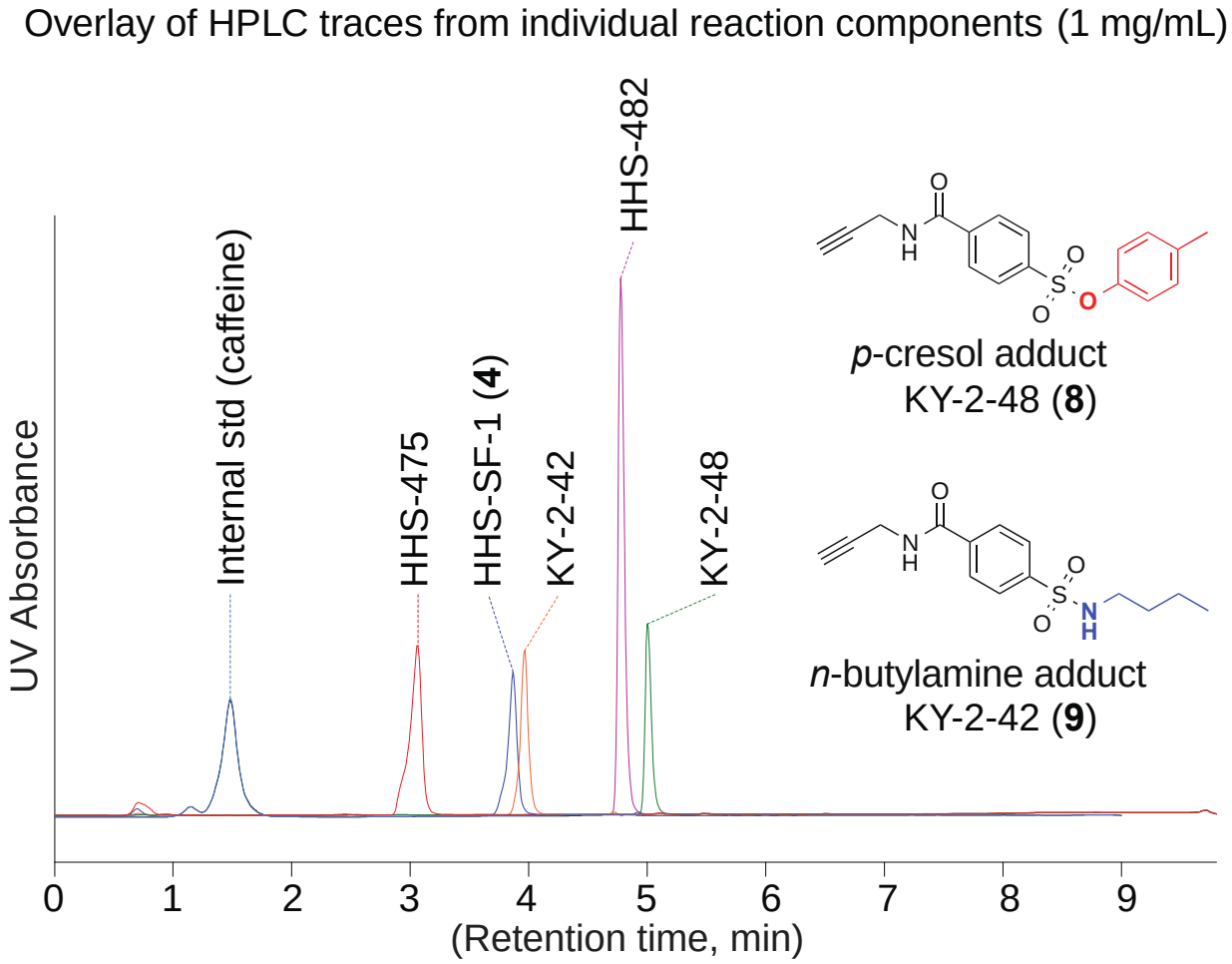


Figure 2.22. Overlay of individual HPLC traces at 1 mg/mL concentrations to show chromatographic resolution of reaction components: caffeine (sky blue), HHS-475 (red), HHS-SF-1 (blue), KY-2-42 (orange, *n*-butylamine-probe adduct), HHS-482 (pink), KY-2-48 (green, *p*-cresol-probe adduct). Caffeine was spiked into each sample as an internal standard to control for run-to-run variations in HPLC analysis (UV detection at 254 nm) of SuTEx and SuFEx reactions. Data shown are representative of two independent experiments (*n*=2).

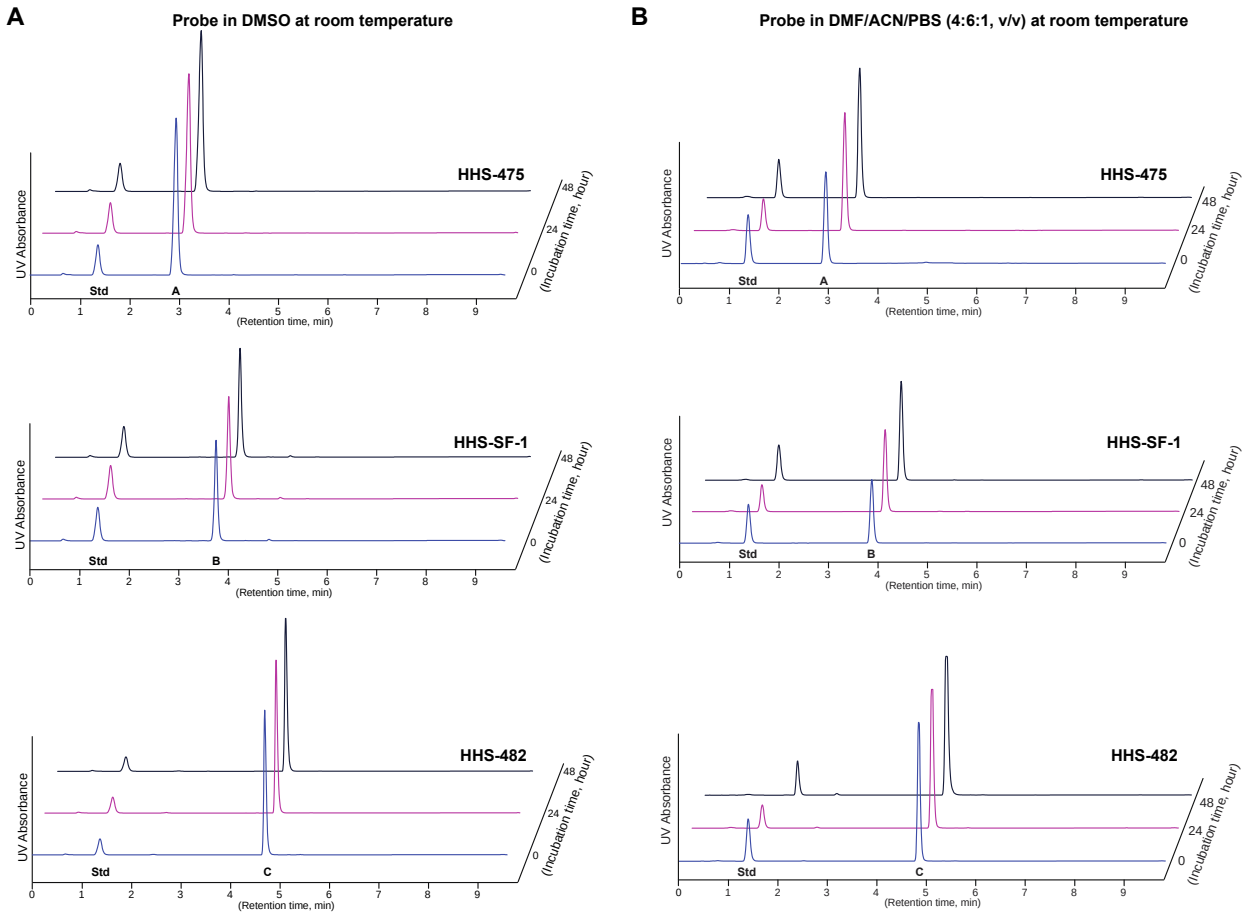


Figure 2.23 Stability of sulfonyl probes in DMSO and aqueous/solvent mixtures. DMSO solutions of HHS-475 (20 mM), HHS-SF-1 (20 mM), and HHS-482 (10 mM) were prepared and HPLC analysis of these probes measured at the indicated time points. Negligible degradation, as judged by reduction of probe signal, was observed after 24- and 48-hours incubation in DMSO or DMF:ACN:PBS (4:6:1, (v/v)) at room temperature. See Supplementary Methods for additional details of the stability assay. DMSO: dimethyl sulfoxide, DMF: dimethylformamide, ACN: acetonitrile, PBS: phosphate-buffered saline. Data shown are representative of three independent experiments (*n*=3).

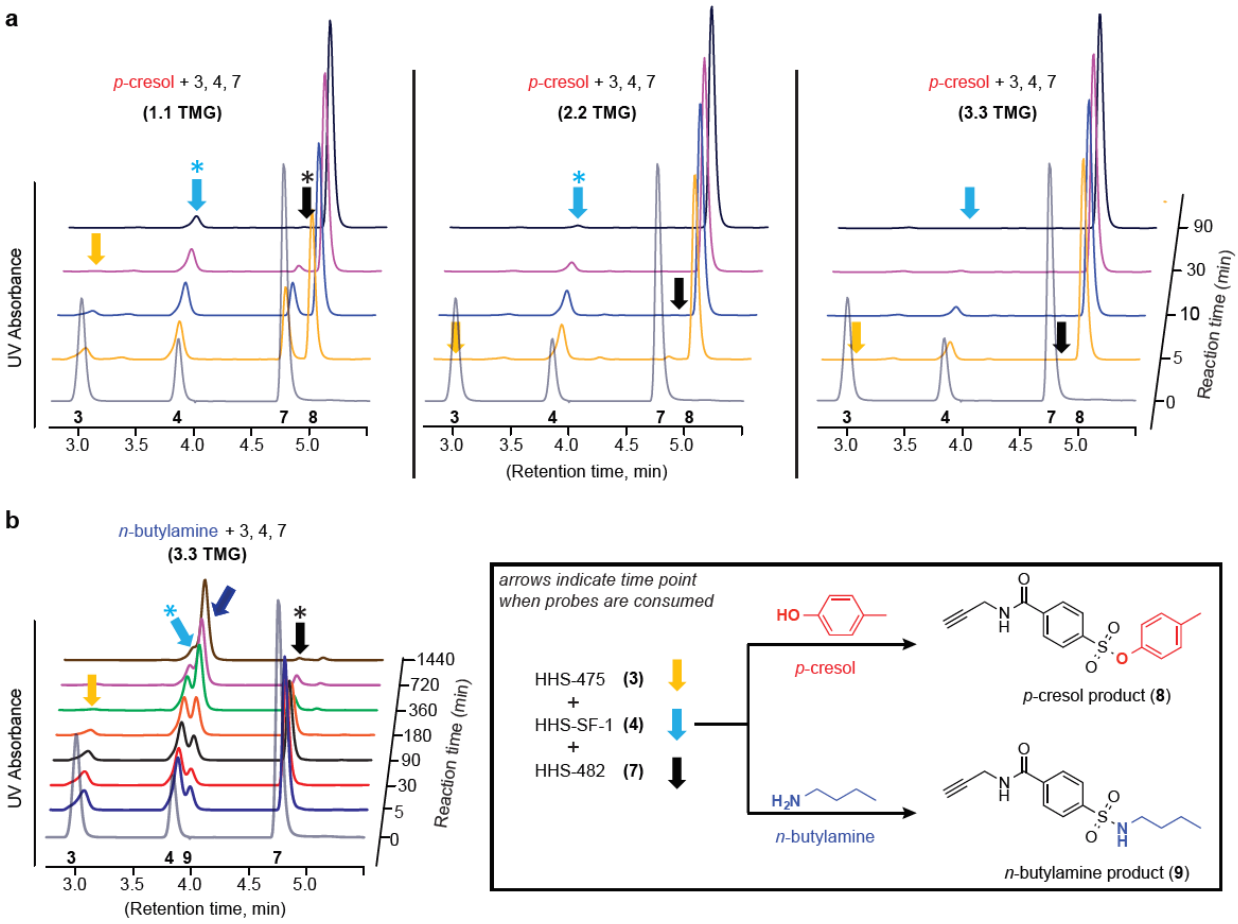


Figure 2.24. Triazole LG enhances phenol reactivity of sulfonyl probes in solution. a) A mixture of HHS-475 (peak 3), HHS-SF-1 (peak 4), and HHS-482 (peak 7) was incubated with *p*-cresol in the presence of increasing amounts of tetramethylguanidine (TMG) base and time dependent covalent reaction monitored by reduction of respective probe signal. Formation of the common *p*-cresol-probe adduct (peak 8) was confirmed by retention time that matched our synthetic standard KY-2-48 (Fig. 2.22). Colored arrows denote the time points when each respective probe was consumed, and the asterisks denote time points corresponding to substantial but not complete probe depletion. b) Reduced reactivity of *n*-butylamine against sulfonyl probes under high TMG conditions (3.3 equivalents). Formation of the *n*-butylamine-probe adduct (peak 9) was validated

by retention time that matched our KY-2-42 synthetic standard (Fig. 2.22). See Methods for additional details. Data shown are representative of three independent experiments (n=3).

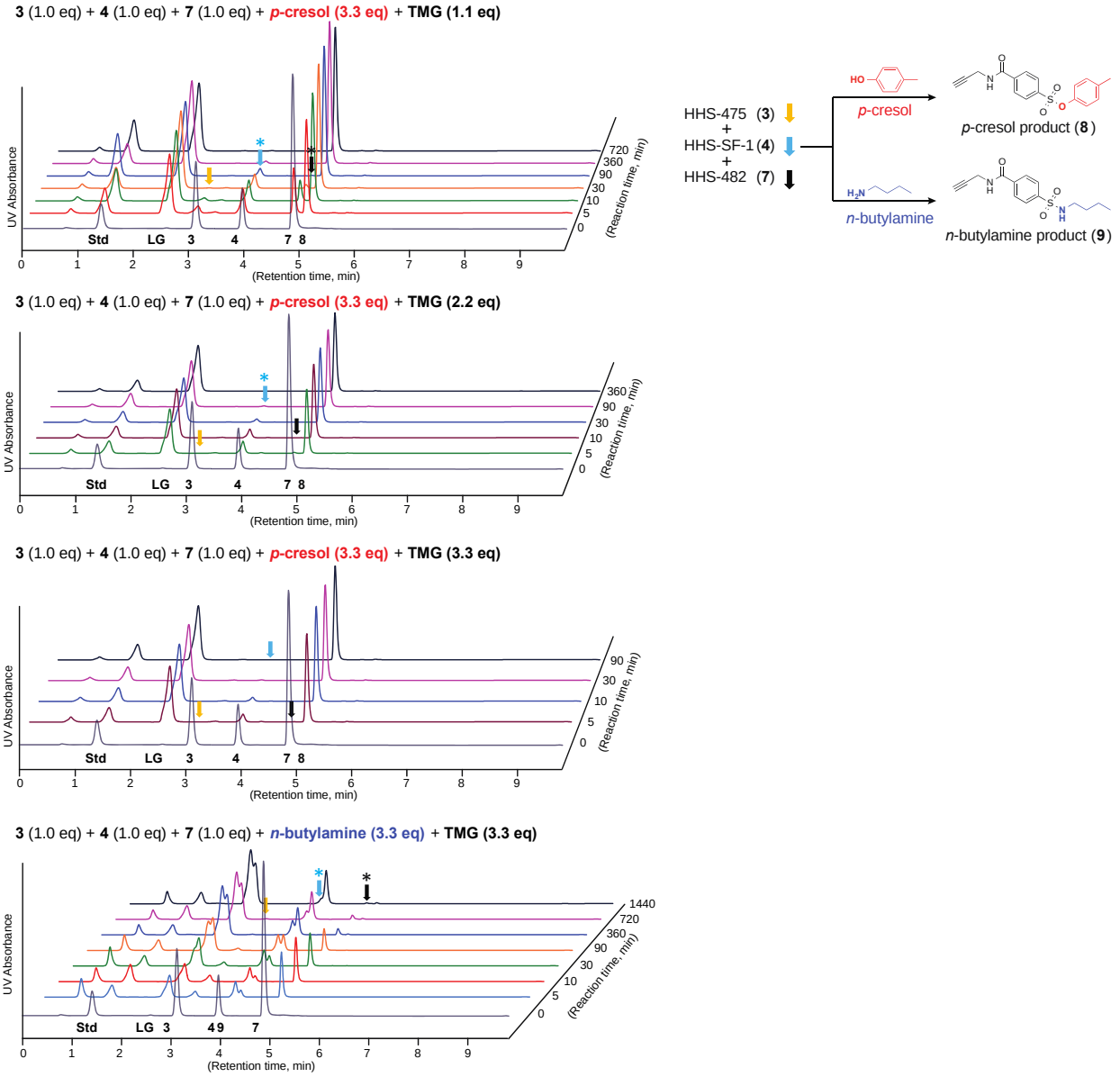


Figure 2.25 Comparison of SuTEx and SuFEx reactivity against nucleophiles in solution. Time-dependent reactions between *p*-cresol or *n*-butylamine with a mixture of HHS-475 (peak 3), HHS-SF-1 (peak 4), and HHS-482 (peak 7) under increasing amounts of TMG base. Formation of the corresponding *p*-cresol-probe (peak 8) or *n*-butylamine-probe products (peak 9) was confirmed by retention times that matched the synthetic standards KY-2-48 and KY-2-42, respectively (Fig. 2.22). Std: internal standard, Caffeine, LG: leaving group of HHS-482 (3-(4-methoxy phenyl)-

1,2,4-triazole) from covalent reaction, TMG: tetramethylguanidine. Data shown are representative of three independent experiments (*n*=3).

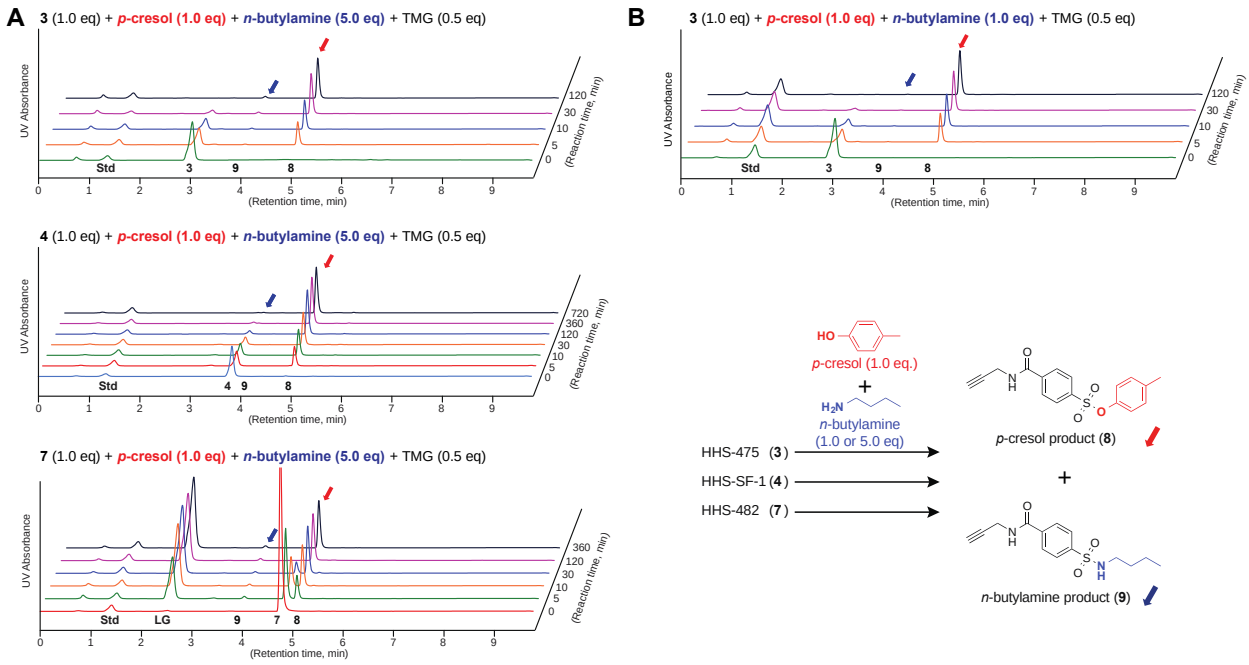


Figure 2.26. Chemoselectivity of sulfonyl probes against *p*-cresol and *n*-butylamine in solution. Reaction of individual sulfonyl probes against a mixture of *n*-butylamine and *p*-cresol mixture. Condition A shows the traces of a mixture of *n*-butylamine and *p*-cresol of 5:1 ratio under catalytic TMG (0.5 equivalents). Condition B entails the reaction of the sulfonyl probes with a mixture of equivalent amounts of *n*-butylamine and *p*-cresol under catalytic amount of TMG (0.5 equivalents). Red arrows show formation of *p*-cresol-probe product (peak 8) and blue arrows show formation of *n*-butylamine-probe product (peak 9) under respective reaction conditions. Std: internal standard, caffeine, LG: leaving group of HHS-482 (3-(4-methoxy phenyl)-1,2,4-triazole) from covalent reaction, TMG: tetramethylguanidine. Data shown are representative of three independent experiments (n=3).

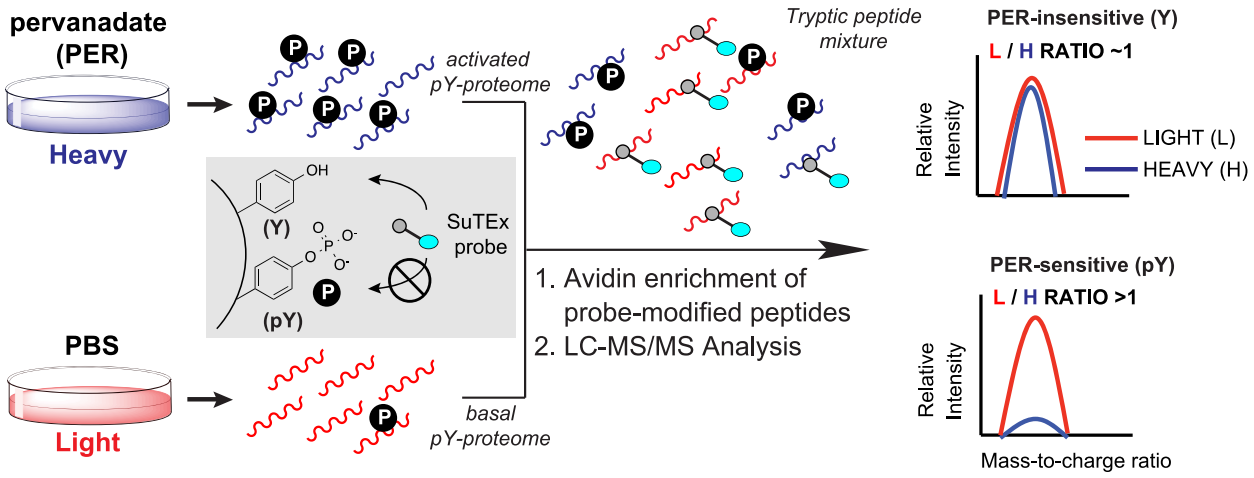


Figure 2.27. Schematic of a SuTEx platform for global tyrosine phosphoproteomic studies. Activation of tyrosine phosphorylation (pY) using a general tyrosine phosphatase inhibitors (pervanadate) will reduce availability of tyrosines (Y) for SuTEx probe labeling, which can be readout by quantitative chemical proteomics (SILAC).

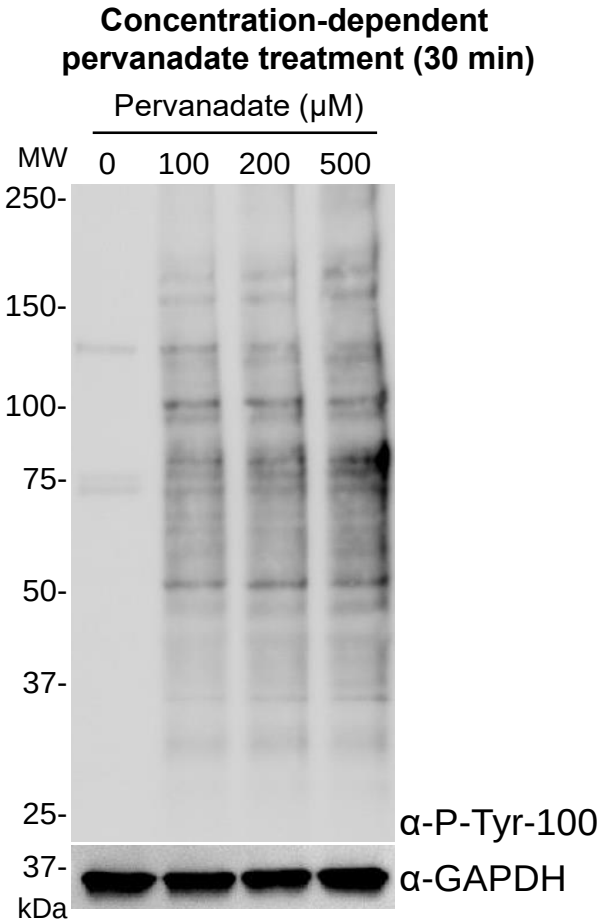


Figure 2.28. Concentration-dependent activation of global tyrosine phosphorylation. A549 cells were treated with vehicle (PBS) or pervanadate at the indicated concentrations for 30 min followed by cell lysis in PBS + protease and phosphatase inhibitors. Activation of global tyrosine phosphorylation was assessed by western blot analysis with a phospho-tyrosine monoclonal antibody (P-Tyr-100). Equivalent protein loading was confirmed using an antibody against GAPDH. Data shown are representative of two experiments (*n*=2 biologically independent experiments).

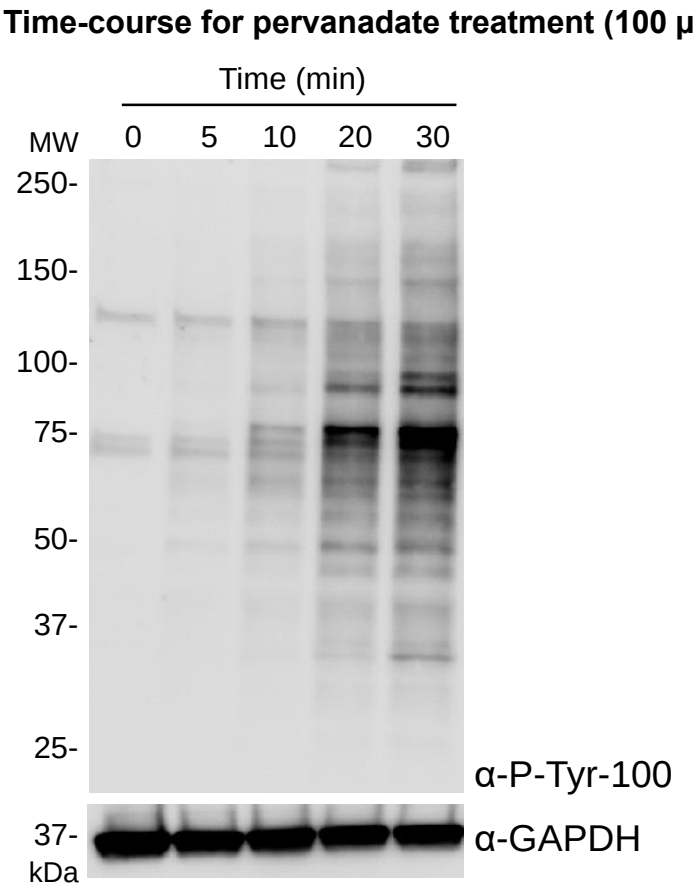


Figure 2.29. Time-dependent activation of global tyrosine phosphorylation. A549 cells were treated with vehicle (PBS) or pervanadate (100 µM) and lysed in PBS + protease and phosphatase inhibitors at the indicated time points. Global tyrosine phosphorylation activation measured by western blot analysis using a phospho-tyrosine monoclonal antibody (P-Tyr-100). Equivalent protein loading was confirmed using an antibody against GAPDH. Data shown are representative of two experiments (*n*=2 biologically independent experiments).

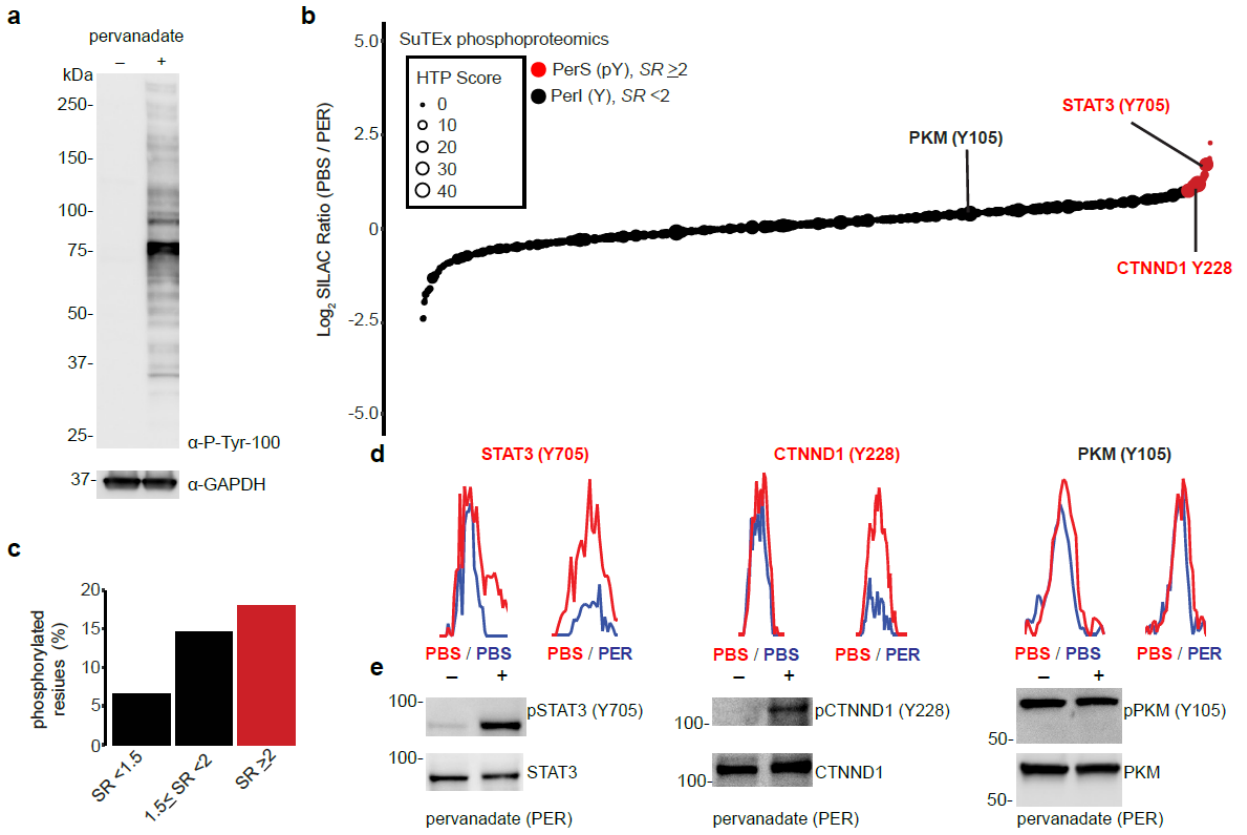


Figure 2.30. Chemical phosphotyrosine-proteomics by SuTEx. a) Western blot analysis confirming activation of global tyrosine phosphorylation (detected via a phospho-tyrosine monoclonal antibody, P-Tyr-100) with pervanadate treatment conditions of A549 cells (100 µM, 30 min) used for chemical proteomic studies. b) Plot of HHS-475-modified tyrosine sites (represented by individual circles) as a function of SILAC ratios (*SR*, light (PBS)/heavy (pervanadate or PER)). Size of circles reflect the HTP score (PhosphoSitePlus). Tyrosine sites were further segregated into pervanadate-insensitive (PerI) and -sensitive (PerS) groups based on SR <2 or ≥2, respectively. Soluble proteomes from pervanadate activated-A549 cells were labeled with HHS-475 (100 µM) for 30 min at 37 °C. c) Bar plot showing trend towards increased number of phosphotyrosine annotations (HTP ≥10) on tyrosine sites with enhanced pervanadate sensitivity. Validation that blockade of HHS-475 labeling (d) of individual tyrosine sites on STAT3 (Y705),

CTNND1 (Y228), and PKM (Y105) coincides with increased phosphorylation at respective sites with pervanadate activation (e). Equivalent protein loading was confirmed by western blot analysis of non-phosphorylated protein counterparts. See Methods for additional details of pervanadate activation and phosphotyrosine western blot procedures. All data shown are representative of two experiments (*n*=2 biologically independent experiments).

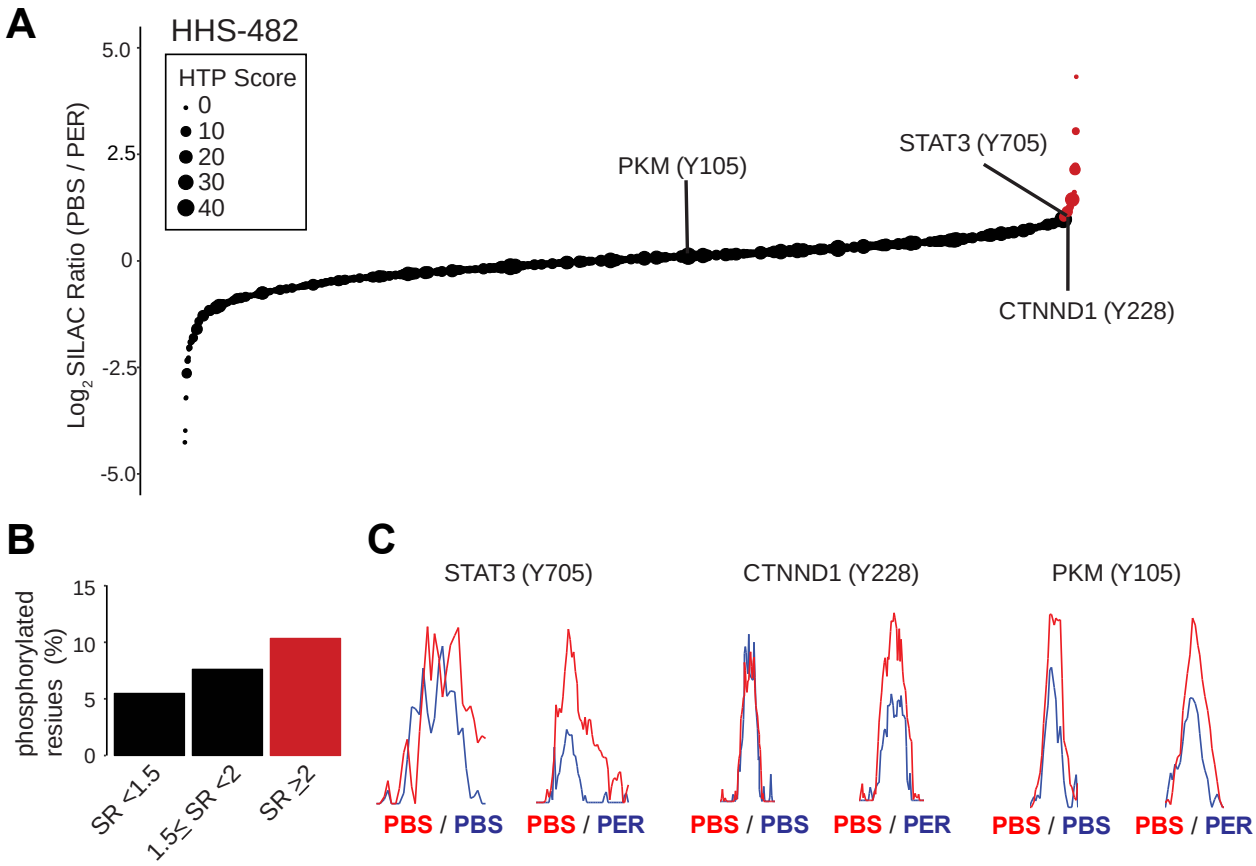


Figure 2.31. Chemical phosphotyrosine-proteomics by HHS-482. (A) Plot of HHS-482-modified tyrosine sites (represented by individual circles) as a function of SILAC ratios (*SR*, light (PBS)/heavy (pervanadate or PER)). Size of circles reflect the number of phosphotyrosine high throughput annotations on PhosphoSitePlus (HTP score). Tyrosine sites were further segregated into pervanadate-insensitive (black circles) and -sensitive (red circles) groups based on *SR* <2 or ≥2, respectively. Soluble proteomes from pervanadate activated-A549 cells were labeled with HHS-482 (100 μM) for 30 min at 37 °C. (B) Bar plot showing trend towards increased phosphotyrosine annotation (HTP ≥10) in tyrosine sites with enhanced pervanadate sensitivity. (C) Validation that blockade of HHS-482 labeling of individual tyrosine sites on STAT3 (Y705), CTNND1 (Y228), and PKM (Y105) coincides with increased phosphorylation at respective sites with pervanadate activation (see Fig 7F). See Table S1 for *SR* values of tyrosines sites detected by chemical proteomics. Data shown are representative of two experiments (*n*=2 biologically independent experiments).

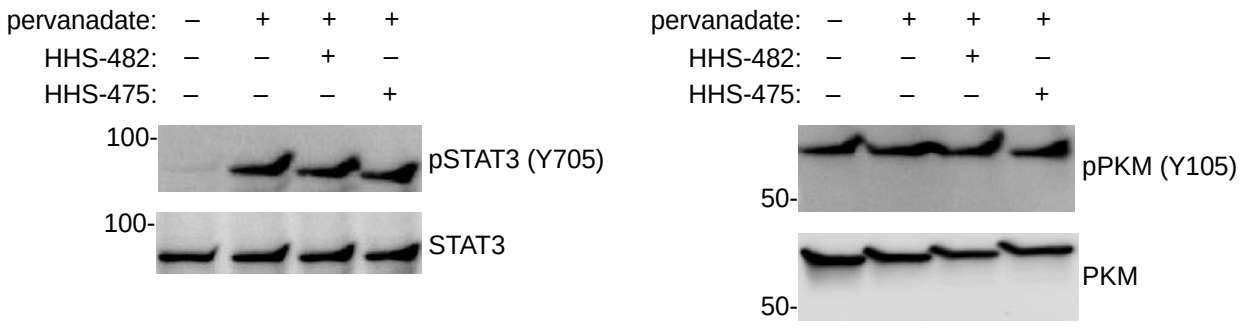


Figure 2.32. SuTEx probe treatment of proteomes from pervanadate-treated cells does not displace phospho-tyrosines. A549 cells were treated with vehicle (PBS) or pervanadate (100 μM, 30 min) followed by lysis in PBS + protease and phosphatase inhibitors. Proteomes from pervanadate-treated cells were treated with either HHS-482 or HHS-475 (100 μM of SuTEx probe) for 30 min at 37 ˚C followed by western blot analysis of individual tyrosine sites on STAT3 (Y705) and PKM (Y105). Treatment with either HHS-482 or HHS-475 did not affect phosphotyrosine signals indicating that SuTEx probes do not non-specifically displace phosphates from tyrosines. See Methods for additional details of western blot analyses. Data shown are representative of two experiments (*n*=2 biologically independent experiments).

ACKNOWLEDGMENTS

We thank all members of the Hsu Lab and colleagues at the University of Virginia for helpful discussions and review of the manuscript. We thank M. Ross for his assistance with mass spectrometry experiments and data analysis. We thank S. Campbell for assistance with molecular biology experiments. We thank D. Dickie for assistance with small molecule crystallography studies. We thank R. Rumana for assistance with cell culture studies. This work was supported by the University of Virginia (start-up funds to K.-L.H.), University of Virginia Cancer Center (A.H.L. and K.-L.H), National Institutes of Health Grants GM801868 (T.B.W.), GM007055 (J.W.B.), CA009109 (A.L.B.), and DA043571 (K.-L.H.), the Schiff Foundation (K.-L.H.), the Wagner Fellowship (A.L.B), the National Science Foundation Graduate Research Fellowship (Grant No. 2018255830 to R.L.M.), and the U.S. Department of Defense (W81XWH-17-1-0487 to K.-L.H.).

2.6 Methods

HPLC assay for profiling solution reactivity and stability of sulfonyl probes

The following reagents were prepared and kept at 0 °C prior to use: 0.1 M solution of caffeine in acetonitrile (ACN), 1.0 M solution of n-butylamine, *p*-cresol, tetramethylguanidine (TMG), acetic acid in ACN, and 10 mM solution of the probes in a mixture of DMF-ACN (v/v=10:90) are made.

(i) *p*-Cresol reactivity against a probe mixture: A solution of *p*-cresol (16.5 µmol, 3.3 eq) was premixed with 1.1, 2.2, or 3.3 eq of TMG. To initiate the reaction, the *p*-cresol/TMG solution was added to a sulfonyl probe mixture of HHS-475/HHS-482/HHS-SF-1 (500 µL, 5 µmol, 1.0 eq each) and the reaction was kept at 0 °C. The reaction progress was monitored by taking out a 50.0 µL aliquot of the reaction mixture at various time points followed by addition of a 10 µL quenching solution of acetic acid (0.5 M final, 5.0 µmol) and the internal caffeine standard (0.05 M final, 0.5 µmol). Sample (1.0 µL) was injected and analyzed by reverse-phase HPLC on a Shimadzu 1100 Series spectrometer with UV detection at 254 nm. Reaction progress was evaluated by monitoring consumption of sulfonyl probes because all probes generate a shared *p*-cresol and *n*-butylamine product. Chromatographic separation was performed using a Phenomenex Kinetex C18 column (2.6 µm, 50 mm x 4.6 mm). Mobile phases A and B were composed of H_2O (with 0.1% AcOH) and CH_3CN (with 0.1% AcOH), respectively. Using a constant flow rate of 0.8 mL/min, the gradient was as follows: 0-0.5 min, 15% B; 0.5-6.5 min 15-85% B (linear gradient); 6.5-7 min 85-100% B (linear gradient); 7- 8.5 min 100% B; 8.5-9 min 100-15% B (linear gradient); 9-9.8 min 15% B.

(ii) *n*-Butylamine reactivity against a probe mixture: Reactivity of sulfonyl probes against *n*-butylamine (3.3 eq) was performed as described above except the amount of TMG was fixed at 3.3 eq.

(iii) Probe reactivity against a *p*-cresol/*n*-butylamine mixture: A solution of *n*-Butylamine (50.0 µL, 50.0 µmol, 5.0 eq), *p*-cresol (10.0 µL, 10.0 µmol, 1.0 eq), and TMG (5.0 µL, 5 µmol, 0.5 eq) were prepared. Probe reaction was initiated by addition of this solution to HHS-475, HHS-482, or HHS-SF-1 (10 µmol, 1.0 eq) at 0 °C. Reaction progress was monitored as described above. A control experiment was also performed where equal amounts of *n*-butylamine (1.0 eq) and *p*-cresol (1.0 eq) were mixed.

(iv) Probe stability studies: Each probe was dissolved in DMSO or a solution of DMF:ACN:PBS (4:6:1 (v/v)) at the following concentrations: 20 mM of HHS-475, 20 mM HHS-SF-1, and 10 mM

of HHS-482 in a final volume of 50 µL. The internal caffeine standard (0.5 µmol) was spiked into each probe sample. Probe stability was monitored at room temperature by taking 1.0 µL of sample at three time points (0, 24, and 48 hours) and analyzing probe degradation by HPLC as described above.

Cell culture

Cell lines were cultured at 37 ˚C with 5% CO_2 with manufacturer recommended media supplemented with 10% fetal bovine serum (FBS, U.S. Source, Omega Scientific) and 1% L-glutamine (Fisher Scientific): HEK293T: DMEM; DM93, A549, Jurkat, H82: RPMI. Cells were harvested for experimental use when they reached ~90% confluency. The media was aspirated, cells washed with cold PBS (2X) and scraped from plates. The cells were pelleted by centrifugation at 400 × g for 5 min, snap-frozen using liquid nitrogen and stored at -80 ˚C until further use.

SILAC cell culture

SILAC HEK293T cells were cultured at 37 ˚C with 5% CO_2 in either 'light' or 'heavy' media consisting of DMEM (Fisher Scientific) supplemented with 10% dialyzed FBS (Omega Scientific), 1% L-glutamine (Fisher Scientific), penicillin/streptomycin, and isotopically-labeled amino acids. Light media was supplemented with 100 µg/mL L-arginine and 100 µg/mL L-lysine. Heavy media was supplemented with 100 µg/mL [$^{13}C_6{}^{15}N_4$]L-arginine and 100 µg/mL [$^{13}C_6{}^{15}N_2$]L-lysine. The cells were grown for 6 passages before use in proteomics experiments. Cells were washed with PBS (2X), harvested, snap-frozen using liquid nitrogen and stored at -80 ˚C until further use.

Transient Transfection

Recombinant protein production by transient transfection of HEK293T cells was performed as previously described[51]. The following plasmid constructs (human proteins) were purchased from GenScript: pcDNA3.1-GSTP1-FLAG, pcDNA3.1-DPP3-FLAG, pcDNA3.1-PGAM1-FLAG, pcDNA3.1-EDC3-FLAG. Site-directed mutagenesis of wild-type constructs was used to generate mutant plasmids: pcDNA3.1-GSTP1 (Y8F)-FLAG, pcDNA3.1-DPP3 (Y417F)-FLAG, pcDNA3.1-PGAM1 (Y92F)-FLAG, pcDNA3.1-EDC3 (Y475F)-FLAG.

Pervanadate Activation

Pervanadate (100 mM) was prepared as previously described[41] by mixing 100 µL of sodium orthovanadate (100 mM Na_3VO_4, New England BioLabs #P0758S) with 1 µL of hydrogen peroxide (H_2O_2, 30% v/v in water) on ice. The mixture was incubated on ice for 15 min followed by immediate addition to cells (1:1000, 100 µM final) and incubation for 30 min at 37 ˚C with 5% CO_2 for general inhibition of protein tyrosine phosphatases. After pervanadate treatment, cells were washed twice with cold PBS followed by harvest. Cell pellets were resuspended in PBS supplemented with protease and phosphatase inhibitor mini tablets (Thermo Scientific #A32959) and then lysed by sonication (3 x 1 sec pulse, 20% amplitude). For CTNND1 western blot studies, cell pellets were lysed in NP40 Cell Lysis Buffer (Invitrogen #FNN0021) supplemented with protease/phosphatase inhibitor tablets. Cell lysates were separated via centrifugation at 100,000 x *g* for 45 min at 4 ˚C for western blot or chemical proteomic studies. Note: pervanadate treatments are performed on live cells but SuTEx probe labeling occurs in proteomes *in vitro*.

Western blot analysis

Western blot analysis of recombinant protein expression was performed as previously described[51]. For analysis of tyrosine phosphorylation, the protocol used was the same except the nitrocellulose blot was blocked with 3% BSA instead of 5% milk in TBS-T. The following antibodies were purchased from Cell Signaling Technology (CST) for phosphotyrosine studies: Phospho-tyrosine (pY): P-Tyr-100 biotinylated, CST #9417S; pPKM: Phospho-PKM (Y105) Rabbit Ab, CST #3827S; PKM: PKM Rabbit Ab, CST #3198S; pSTAT3: Phospho-STAT3 (Y705) Rabbit mAb, CST #9145S; STAT3: STAT3 Mouse mAb, CST #9139S; pCTNND1: Phospho-Catenin δ-1 (Tyr228) Rabbit Ab, CST #2911; CTNND1: Catenin δ-1 Rabbit Ab, CST #4989; GAPDH: GAPDH Rabbit mAb, CST #2118S. The following secondary antibodies were used for fluorescence detection: Goat Anti-Rabbit IgG DyLight 550 Conjugated, Thermo Scientific, #84541; Goat Anti-Mouse IgG DyLight 650 Conjugated, Invitrogen, #84545; Streptavidin DyLight 550 Conjugated, Thermo Scientific, #84542.

Gel-based chemical proteomic assay

Cell pellets were lysed in PBS by sonication and fractionated (100,000 × *g*, 45 min, 4 ˚C) to generate soluble and membrane fractions. Protein concentrations were determined using the Bio-Rad DC protein assay and adjusted to 1 mg/mL in PBS. Proteome samples (49 µL aliquots) were

treated with sulfonyl-triazole or -fluoride probes at the indicated concentrations (1 µL, 50x stock in DMSO) for 1 hr at room temperature. Probe-labeled samples were conjugated by copper-catalyzed azide-alkyne cycloaddition (CuAAC) to rhodamine-azide (1 µL of 1.25 mM stock; final concentration of 25 µM) using tris(2-carboxyethyl)phosphine (TCEP; 1 µL of fresh 50 mM stock in water; final concentration of 1 mM), tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA, 3 µL of a 1.7 mM 4:1 t-butanol/DMSO stock, final concentration of 100 µM), and copper sulfate ($CuSO_4$, 1 µL of 50 mM stock, final concentration of 1 mM). Samples were reacted for 1 hr at room temperature, quenched with 17 µL of 4X SDS-PAGE loading buffer and beta-mercaptoethanol (βME), and quenched samples (30 µL) analyzed by SDS-PAGE gel and in-gel fluorescence scanning.

Live cell evaluation of sulfonyl-triazole probes

Cells grown to ~90% confluency in 10 cm plates were treated with DMSO vehicle or sulfonyl-triazole probe (10 µL of 1000X DMSO stock) in serum-free media for the indicated concentrations and times at 37 ˚C with 5% CO_2. After treatment, cells were washed with cold PBS twice before harvesting and preparation for gel-based chemical proteomic evaluation as described above. For LC-MS studies, protein concentrations were normalized to 2.3 mg/mL and 432 µL (for 1 mg final protein amount) were used for sample preparation as detailed below.

Preparation of proteomes for LC-MS/MS analysis

Proteomes were diluted to 2.3 mg/ mL in PBS and sample aliquots (432 µL) were treated with sulfonyl-triazole or -fluoride probes at the indicated concentrations (5 µL, 100X stock in DMSO), mixed gently and incubated for 1 h at room temperature. Probe-modified proteomes were subjected to CuAAC conjugation to desthiobiotin-PEG3-azide (10 µL of 10 mM stock in DMSO; final concentration of 200 µM) using TCEP (10 µL of fresh 50 mM stock in water; 1 mM final concentration), TBTA ligand (33 µL of a 1.7 mM 4:1 t-butanol/DMSO stock, 100 µM final concentration), and $CuSO_4$ (10 µL of 50 mM stock, 1 mM final concentration). Samples were mixed by vortexing and then incubated for 1 h at room temperature. Excess reagents were removed by chloroform-methanol extraction as previously described[51]. Protein pellets were re-suspended in 500 µL of 6M urea/25 mM ammonium bicarbonate followed by DTT reduction and IAA alkylation as previously described[51]. Excess reagents were removed by chloroform/methanol extraction as

described above, and the protein pellet was re-suspended in 500 µL of 25 mM ammonium bicarbonate and then digested to peptides using trypsin/Lys-C (7.5 µg in 15 µL of ammonium bicarbonate, sequencing grade from Promega) was added to the mixture and incubated for 3 hrs at 37 ˚C. Probe-modified peptides were enriched by avidin affinity chromatography, eluted, and prepared for LC-MS analysis as previously described[51].

Preparation of SILAC proteomes for LC-MS/MS analysis

Heavy and light proteomes (432 µL of each) were diluted to 2.3 mg/mL in PBS. For 10:1 comparisons, heavy and light proteomes were treated with 250 µM and 25 µM of HHS465, respectively (5 µL, 100x stock in DMSO). In a control 1:1 comparison experiment both heavy and light proteome were treated with 25 µM of HHS465. Samples were mixed gently and incubated for 1 h at room temperature. Light and heavy samples were separately conjugated to desthiobiotin-PEG3-azide as described above. Light and heavy samples were mixed during the chloroform-methanol extraction step. Probe-modified peptides were prepared for LC-MS/MS analysis as described above.

LC-MS/MS analysis of samples

Nano-electrospray ionization-liquid chromatography-mass spectrometry analyses (LC-MS/MS) were performed using an Ultimate 3000 RSLC nanoSystem-Orbitrap Q Exactive Plus mass spectrometer (Thermo Scientific) as previously described[51] except LC conditions were modified to use the following gradient (A: 0.1% formic acid/H_2O; B: 80% ACN, 0.1% formic acid in H_2O): 0-1:48 min 1% B, 400 nL/min; 1:48 – 2:00 min 1% B, 300 nL/min; 2-90 min 16% B; 90-146 25% B; 146-147 min 95% B; 147-153 min 95% B; 153-154 min1% B; 154.0-154.1 min 1% B, 400 nL/min; 154.1-180 min 1% B, 400 nL/min. A top 10 data-dependent acquisition MS method was used.

LC-MS/MS data analysis

Identification of peptides and proteins from tandem mass spectrometry analyses was accomplished using the Byonic™ software package (Protein Metrics Inc.[31]). Data were searched against a modified human protein database (UniProt human protein database, angiotensin I and vasoactive intestinal peptide standards; 40,660 proteins) with the following parameters: up to 3 missed

cleavages to account for a lysine probe modification, 10 ppm precursor mass tolerance, 20 ppm fragment mass tolerance, too high (narrow) "precursor isotope off by x", precursor and charge assignment computed from MS1, maximum of 1 precursor per MS2, 0.01 smoothing width, 1% protein false discovery rate, variable (common) methionine oxidation (+15.9949 Da) and fixed cysteine carbamidomethylation (+57.021464 Da). Sulfonyl-probe modifications of tyrosine, lysine, and other amino acids were included as a variable (common) modification of +635.27374 Da. Search results were imported into R and filtered for fully tryptic peptides (except N- and C-terminally modified), a Byonic score of $\geq$300 (unless otherwise specified), and a precursor mass error between -5 ppm and +5 ppm. A Byonic score of 300 was applied for a more inclusive initial evaluation of the search results and thereby consider more possible probe-modified sites. We manually verified the MS1 and MS2 spectra corresponding to the highest-scoring tyrosine- and lysine (internal or non-C terminal)-modified sequences (~50-100 peptides). The next most frequently matched and high-scored probe-modified amino acid residues were C-terminal lysines or arginines, which were determined to be false positive matches based on manual analysis of MS2 spectra (top ~50 highest Byonic scored-matches). These findings are consistent with the observation from previous studies with other probes[11, 51] that trypsin does not cleave after a modified lysine or arginine. Distinct peptides containing probe-modified amino acid residues (termed sites) were determined by identifying all unique razor protein and site combinations across all of the proteomes tested.

Analysis and comparison of sulfonyl probe modified amino acid sites

To compare amino acid residues modified by sulfonyl probes, protein and peptide identifications were accomplished as described above with variable (common) modification of +635.27374 Da on the following amino acid residues: cysteine, aspartic acid, glutamic acid, histidine, lysine, methionine, asparagine, glutamine, arginine, serine, threonine, tryptophan, and tyrosine. For these amino acid comparisons, carbamidomethylation (+57.021464 Da) of cysteines was searched as a variable/common modification to allow for the potential of probe modification on cysteines. Comparisons of probe-modified sites across all probes and cell lines tested were performed using the R package ggplot2 (https://ggplot2.tidyverse.org/). Venn diagrams for comparisons were generated using the VennDiagram R package[52]. For amino acid comparisons, a Byonic score cutoff

of 600 was used to minimize false positive identifications of modified residues, which were confirmed by manual evaluation to be incorrect assignments.

Domain enrichment analysis

Probe-modified sites were compared to ProRule domain annotations (available on PROSITE, release 20.85[53], http://prosite.expasy.org/) using the annotated human UniProt proteome (https://www.uniprot.org/) as a database for identifying amino acid sequences that match ProRule domains. A probe-modified site that is within a ProRule domain is considered a "hit" and is counted as enrichment of a domain by the sulfonyl probe. Several sites within the same ProRule domain annotation are a considered a single hit. If a site had several annotations each one was considered a hit; for example, a modified site within the proton acceptor region of a kinase domain would be annotated as a hit for ProRules PRU10027 and PRU00159, respectively. The database count is determined by the number of non-overlapping occurrences of the domain such that calmodulin would account for 4 EF-hand domains (PRU00448). We find the probability of the domains P(D) in the reference UniProt human database to determine how frequently they exist in nature:

$$P(D) = n(D)/N$$

Where n(D) is the number of domain occurrences in the database and N is the total number of domains in the reference database. The p-values were calculated using a binomial test previously reported for GO statistical overrepresentation test[54].

$$Pvalue = \sum \binom{K}{k} P(D)^k (1 - P(D))^{K-k}$$

Binomial test

Where K is number of domain annotation hits in the experimental data (sulfonyl probe). The p-value was then corrected for a 1% false discovery rate (FDR) using Benjamini-Hochberg correction for multiple hypothesis testing. From these statistical analyses, ProRule domains that show statistically significant overrepresentation (Q value < 0.01) are used to generate bar graphs and pie charts shown in figures. Note that a -log (Q value) is used so that positive values are shown

for simplicity. In order to verify that the binomial approximation to hypergeometric probability we ensured sum of all n(D) was less than 5 percent of N and verified that using a hypergeometric test did not alter the enriched domains. The enriched domains were grouped according to their function into four categories; nucleotide binding, enzyme, protein-protein interaction and undefined based on gene ontology molecular function annotation of the respective ProRule domain. Pie charts and bar graphs were generated using the ggplot2 package in R.

Classification of protein domains

The distinction between protein-protein interaction (PPI) and nucleotide binding domains was determined by whether the interacting partner of the domain is annotated as a peptide or a nucleotide sequence. The SH2 domain (PRU00191) which interacts with proteins featuring phosphorylated tyrosines is classified as a PPI domain, and a Homeobox DNA-binding domain (PRU00108) is classified as a nucleotide binding domain. An enzyme domain is the protein subunit that has been shown to catalyze the conversion of a substrate to a product. The Ribonuclease H domain (PRU00408) functions as an endonuclease which will interact with RNA but is classified as an enzyme domain because of its nuclease activity. We applied gene ontology (GO) molecular function annotations associated with the ProRule domains that inherit the annotation for catalytic activity (GO:0003824) to determine if proteins belong to the enzyme domain group. For example, the term Ribonuclease H domain (PRU00408) has the GO annotation for endonuclease activity (GO:0004519) which has catalytic activity (GO:0003824) in its ancestor chart and is therefore classified as an enzyme.

DrugBank analysis

Proteins labeled by sulfonyl probes in live cells were compared against protein targets of FDA approved and all drugs in the DrugBank databases[33] (version 5.1.1).

Phosphosite Plus analysis

Probe-labeled sites were searched for in the PhosphoSitePlus database[32] either unfiltered or filtered by a high-throughput reference score of 10 or greater where specified.

Nucleophilicity data analysis (SILAC)

Peptide and protein identification was accomplished using Byonic as previously described above. SILAC samples were searched with added masses for heavy-labeled amino acids (+10.0083 Da for R, +8.0142 Da for K) and converted into mzXML (from raw data file) and mzid (exported from Byonic) format for export into Skyline-daily[55] to determine SILAC ratios (*SR*) of light/heavy peptides as previously described[51]. SILAC ratios from peptides with the same probe-modified site were averaged. The SILAC ratios were then plotted using the ggplot2 package in R[56]. Nucleophilicity was defined as follows: hyper-reactive, $SR \leq 2$; mild reactivity, $2 < SR \leq 5$; low reactivity, $SR > 5$.

GSTP1 biochemical substrate assay

Recombinant GSTP1-HEK293T soluble cell proteomes were diluted to 1 mg/ml in assay buffer (100 mM NaH_2PO_4, pH 7.0). GSH stock solution (250 mM in water) was diluted to 4 mM in assay buffer and 25 µL of diluted GSH solution was added to each sample. A substrate stock solution of 75 mM 1-bromo-2,4-dinitrobenzene (DBNB) in ethanol was diluted to 2 mM in assay buffer. Samples (25 µL) were aliquoted into a 96 well plate and spun briefly via centrifuge. 50 µl of 2 mM BDNB was added to each well and the reaction was monitored in kinetic mode by measuring absorbance at 340 nm for 10 min on a BMG Labtech CLARIOstar plate reader.

DPP3 biochemical substrate assay

Substrate assays were performed on recombinant DPP3-HEK293T soluble proteomes diluted to 1 mg/mL in assay. DPP3 sample (10 µL) was diluted to 85 µL with assay buffer and transferred to a black 96-well plate. A stock solution of DPP3 substrate (Arg-Arg β-naphthylamide trihydrochloride, 0.5 mM; Sigma-Aldrich) was diluted to 100 µM in assay buffer. Substrate solution (5 µL) was added to each sample. Samples were mixed briefly by shaking and reaction monitored in kinetic mode by measuring fluorescence at 450 nm for 10 min on a BMG Labtech CLARIOstar plate reader.

Data Availability

All data produced or analyzed for this study are included in the published article (and its supplementary information files) or are available from the corresponding author on reasonable request. Crystallographic data for small molecules has been deposited in the Cambridge

Crystallographic Data Centre and have been assigned the following deposition numbers HHS-465 (CCDC 1954297), HHS-475 (CCDC 1954298), HHS-483 (CCDC 1954299).

Code Availability

All code is available upon reasonable request from the corresponding author.

REFERENCES

1. Cravatt BF, Wright AT, Kozarich JW. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu Rev Biochem* 2008, 77: 383-414.

2. Sadaghiani AM, Verhelst SH, Bogyo M. Tagging and detection strategies for activity-based proteomics. *Curr Opin Chem Biol* 2007, 11(1): 20-28.

3. Niphakis MJ, Cravatt BF. Enzyme inhibitor discovery by activity-based protein profiling. *Annu Rev Biochem* 2014, 83: 341-377.

4. Bachovchin DA, Cravatt BF. The pharmacological landscape and therapeutic potential of serine hydrolases. *Nat Rev Drug Discov* 2012, 11(1): 52-68.

5. Deu E, Verdoes M, Bogyo M. New approaches for dissecting protease functions to improve probe development and drug discovery. *Nat Struct Mol Biol* 2012, 19(1): 9-16.

6. Patricelli MP, Szardenings AK, Liyanage M, Nomanbhoy TK, Wu M, Weissig H, *et al.* Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* 2007, 46(2): 350-358.

7. Kumar S, Zhou B, Liang F, Wang WQ, Huang Z, Zhang ZY. Activity-based probes for protein tyrosine phosphatases. *Proc Natl Acad Sci U S A* 2004, 101(21): 7943-7948.

8. Vocadlo DJ, Bertozzi CR. A strategy for functional proteomic analysis of glycosidase activity from cell lysates. *Angew Chem Int Ed Engl* 2004, 43(40): 5338-5342.

9. Liu Y, Patricelli MP, Cravatt BF. Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci U S A* 1999, 96(26): 14694-14699.

10. Weerapana E, Wang C, Simon GM, Richter F, Khare S, Dillon MB, *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* 2010, 468(7325): 790-795.

11. Hacker SM, Backus KM, Lazear MR, Forli S, Correia BE, Cravatt BF. Global profiling of lysine reactivity and ligandability in the human proteome. *Nat Chem* 2017, 9(12): 1181-1190.

12. Lin S, Yang X, Jia S, Weeks AM, Hornsby M, Lee PS, *et al.* Redox-based reagents for chemoselective methionine bioconjugation. *Science* 2017, 355(6325): 597-602.

13. Matthews ML, He L, Horning BD, Olson EJ, Correia BE, Yates JR, 3rd, *et al.* Chemoproteomic profiling and discovery of protein electrophiles in human cells. *Nat Chem* 2017, 9(3): 234-243.

14. Parker CG, Galmozzi A, Wang Y, Correia BE, Sasaki K, Joslyn CM, *et al.* Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* 2017, 168(3): 527-541 e529.

15. Narayanan A, Jones LH. Sulfonyl fluorides as privileged warheads in chemical biology. *Chem Sci* 2015, 6(5): 2650-2659.

16. Gao B, Zhang L, Zheng Q, Zhou F, Klivansky LM, Lu J, *et al.* Bifluoride-catalysed sulfur(VI) fluoride exchange reaction for the synthesis of polysulfates and polysulfonates. *Nat Chem* 2017, 9(11): 1083-1088.

17. Dong J, Sharpless KB, Kwisnek L, Oakdale JS, Fokin VV. SuFEx-based synthesis of polysulfates. *Angew Chem Int Ed Engl* 2014, 53(36): 9466-9470.

18. Fahrney DE, Gold AM. Sulfonyl Fluorides as Inhibitors of Esterases. I. Rates of Reaction with Acetylcholinesterase, α-Chymotrypsin, and Trypsin. *Journal of the American Chemical Society* 1963, 85(7): 997-1000.

19. Shannon DA, Gu C, McLaughlin CJ, Kaiser M, van der Hoorn RA, Weerapana E. Sulfonyl fluoride analogues as activity-based probes for serine proteases. *Chembiochem* 2012, 13(16): 2327-2330.

20. Gu C, Shannon DA, Colby T, Wang Z, Shabab M, Kumari S, *et al.* Chemical proteomics with sulfonyl fluoride probes reveals selective labeling of functional tyrosines in glutathione transferases. *Chem Biol* 2013, 20(4): 541-548.

21. Zhao Q, Ouyang X, Wan X, Gajiwala KS, Kath JC, Jones LH, *et al.* Broad-Spectrum Kinase Profiling in Live Cells with Lysine-Targeted Sulfonyl Fluoride Probes. *J Am Chem Soc* 2017, 139(2): 680-685.

22. Yang B, Wu H, Schnier PD, Liu Y, Liu J, Wang N, *et al.* Proximity-enhanced SuFEx chemical cross-linker for specific and multitargeting cross-linking mass spectrometry. *Proc Natl Acad Sci U S A* 2018, 115(44): 11162-11167.

23. Yang X, Michiels TJM, de Jong C, Soethoudt M, Dekker N, Gordon E, *et al.* An Affinity-Based Probe for the Human Adenosine A2A Receptor. *J Med Chem* 2018, 61(17): 7892-7901.

24. Dong J, Krasnova L, Finn MG, Sharpless KB. Sulfur(VI) fluoride exchange (SuFEx): another good reaction for click chemistry. *Angew Chem Int Ed Engl* 2014, 53(36): 9430-9448.

25. Chen W, Dong J, Plate L, Mortenson DE, Brighty GJ, Li S, *et al.* Arylfluorosulfates Inactivate Intracellular Lipid Binding Protein(s) through Chemoselective SuFEx Reaction with a Binding Site Tyr Residue. *J Am Chem Soc* 2016, 138(23): 7353-7364.

26. Mortenson DE, Brighty GJ, Plate L, Bare G, Chen W, Li S, *et al.* "Inverse Drug Discovery" Strategy To Identify Proteins That Are Targeted by Latent Electrophiles As Exemplified by Aryl Fluorosulfates. *J Am Chem Soc* 2018, 140(1): 200-210.

27. Fadeyi OO, Hoth LR, Choi C, Feng X, Gopalsamy A, Hett EC, *et al.* Covalent Enzyme Inhibition through Fluorosulfate Modification of a Noncatalytic Serine Residue. *ACS Chem Biol* 2017, 12(8): 2015-2020.

28. Liu Z, Li J, Li S, Li G, Sharpless KB, Wu P. SuFEx Click Chemistry Enabled Late-Stage Drug Functionalization. *J Am Chem Soc* 2018, 140(8): 2919-2925.

29. Adibekian A, Martin BR, Wang C, Hsu KL, Bachovchin DA, Niessen S, *et al.* Click-generated triazole ureas as ultrapotent in vivo-active serine hydrolase inhibitors. *Nat Chem Biol* 2011, 7(7): 469-478.

30. Ahn K, Boehm M, Brown MF, Calloway J, Che Y, Chen J, *et al.* Discovery of a Selective Covalent Inhibitor of Lysophospholipase-like 1 (LYPLAL1) as a Tool to Evaluate the Role of this Serine Hydrolase in Metabolism. *ACS Chem Biol* 2016, 11(9): 2529-2540.

31. Bern M, Kil YJ, Becker C. Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics* 2012, Chapter 13: Unit13 20.

32. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* 2015, 43(Database issue): D512-520.

33. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018, 46(D1): D1074-D1082.

34. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* 2018, 19(5): 327-341.

35. Yaffe MB. Phosphotyrosine-binding domains in signal transduction. *Nat Rev Mol Cell Biol* 2002, 3(3): 177-186.

36. Shin M, Franks CE, Hsu KL. Isoform-selective activity-based profiling of ERK signaling. *Chem Sci* 2018, 9(9): 2419-2431.

37. Choi EJ, Jung D, Kim JS, Lee Y, Kim BM. Chemoselective Tyrosine Bioconjugation through Sulfate Click Reaction. *Chemistry* 2018, 24(43): 10948-10952.

38. Shannon DA, Banerjee R, Webster ER, Bak DW, Wang C, Weerapana E. Investigating the proteome reactivity and selectivity of aryl halides. *J Am Chem Soc* 2014, 136(9): 3330-3333.

39. Humphrey SJ, Yang G, Yang P, Fazakerley DJ, Stockli J, Yang JY, *et al.* Dynamic adipocyte phosphoproteome reveals that Akt directly regulates mTORC2. *Cell Metab* 2013, 17(6): 1009-1020.

40. Lundby A, Secher A, Lage K, Nordsborg NB, Dmytriyev A, Lundby C, *et al.* Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. *Nat Commun* 2012, 3: 876.

41. Hilger M, Bonaldi T, Gnad F, Mann M. Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics. *Mol Cell Proteomics* 2009, 8(8): 1908-1920.

42. Song G, Chen L, Zhang B, Song Q, Yu Y, Moore C, *et al.* Proteome-wide Tyrosine Phosphorylation Analysis Reveals Dysregulated Signaling Pathways in Ovarian Tumors. *Mol Cell Proteomics* 2019, 18(3): 448-460.

43. Song L, Turkson J, Karras JG, Jove R, Haura EB. Activation of Stat3 by receptor tyrosine kinases and cytokines regulates survival in human non-small cell carcinoma cells. *Oncogene* 2003, 22(27): 4150-4165.

44. Hong JY, Oh IH, McCrea PD. Phosphorylation and isoform use in p120-catenin during development and tumorigenesis. *Biochim Biophys Acta* 2016, 1863(1): 102-114.

45. Hitosugi T, Kang S, Vander Heiden MG, Chung TW, Elf S, Lythgoe K, *et al.* Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth. *Sci Signal* 2009, 2(97): ra73.

46. Weerapana E, Simon GM, Cravatt BF. Disparate proteome reactivity profiles of carbon electrophiles. *Nat Chem Biol* 2008, 4(7): 405-407.

47. Manley JL, Krainer AR. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev* 2010, 24(11): 1073-1074.

48. Hargous Y, Hautbergue GM, Tintaru AM, Skrisovska L, Golovanov AP, Stevenin J, *et al.* Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J* 2006, 25(21): 5126-5137.

49. Harris TK, Turner GJ. Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life* 2002, 53(2): 85-98.

50. Decker CJ, Parker R. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb Perspect Biol* 2012, 4(9): a012286.

51. Franks CE, Campbell ST, Purow BW, Harris TE, Hsu KL. The Ligand Binding Landscape of Diacylglycerol Kinases. *Cell Chem Biol* 2017, 24(7): 870-880 e875.

52. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 2011, 12: 35.

53. Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, *et al.* New and continuing developments at PROSITE. *Nucleic Acids Res* 2013, 41(Database issue): D344-347.

54. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* 2013, 8(8): 1551-1566.

55. Bereman MS, Beri J, Sharma V, Nathe C, Eckels J, MacLean B, *et al.* An Automated Pipeline to Monitor System Performance in Liquid Chromatography-Tandem Mass Spectrometry Proteomic Experiments. *J Proteome Res* 2016, 15(12): 4763-4769.

56. Wickham H, SpringerLink (Online service), SpringerLINK ebooks - Mathematics and Statistics (2016). ggplot2 Elegant Graphics for Data Analysis. 2nd ed. [S.l.]: Springer International Publishing; 2016.

Chapter 3. Development of tyrosine ligands to perturb protein function

3.1 Introduction

After years of reservations, the use of covalent small molecules in drug discovery is recently facing resurgence. Although associated with potential toxicities arising from promiscuous reactions, haptenization, and idiosyncratic drug reactions[1], they exhibit improved potency, prolonged duration (leading to less frequent dosing), beneficial pharmacokinetic profile, increased ability to evade mutations, and ability to target proteins in shallow binding sites previously deemed "undruggable"[2-3]. This appealing risk-benefit profile has shifted industry sentiment toward the use of covalent small molecule compounds. There are currently approximately 50 FDA-approved covalent inhibitors divided into two categories: mechanism-based inhibitors, which form covalent bonds with an enzyme active-site catalytic residue, and targeted covalent inhibitors, which bind to bystander or non-catalytic residues[4].

The recent advancement of chemoproteomic technologies has led to discoveries of electrophilic reactive warheads targeting lysine[5-6] and cysteine[7-8] residues that have enormously expanded our understanding of biology, in combination with identification of novel protein targets and their associated modalities[1]. Still, of the ~20,000 human proteins, only ~11% have been liganded and only 1.4-4% have a quality chemical probe[9]. To expand the scope of liganded proteins, we need to explore residues beyond cysteine and lysine.

We recently developed chemical probes using sulfur-triazole exchange (SuTEx) chemistry that enabled the study of >10,000 tyrosine residues in ~3700 proteins in lysates and live cells[10]. We discovered that a large fraction (~77%) of the proteins enriched by the SuTEx probes were not in the Drugbank database. As a result, we developed a library of SuTEx electrophile fragments and screened them in lysates and live cells[11]. The fragments liganded 305 tyrosine sites on 213

distinct proteins across >1500 probe-modified unique tyrosine sites. We demonstrated that liganding sites in non-catalytic tyrosine and phosphotyrosine sites can disrupt protein function and that SuTEx, as compared to SuFEx, can dramatically enhance the potency of the sulfur electrophile while maintaining reasonable specificity across the proteome.

Encouraged by these results, we sought to optimize the SuTEx fragments to improve selectivity and potency in an endogenous system. The fragment library was designed based on the previously reported 1,2,4-triazole probe (HHS-475)[10] scaffold because it showed the highest enrichment of tyrosine sites in live cells among all probes tested[10]. As proof of principle, we synthesized six SuTEx fragments and chose to modify the adduct group as previous results demonstrated that adduct modifications had a greater effect on reactivity compared to leaving group modifications[11]. The fragments exhibited high selectivity, liganding 97 unique tyrosine sites on 95 proteins, corresponding to 6% and 10% of quantified probe-modified tyrosine sites and proteins, respectively. More importantly, the fragments displayed great potency in live cells against two target proteins, prostaglandin reductase 2 (PTGR2, IC_{50} = 1 μM) and fumarylacetoacetase (FAH, IC_{50} = 2 μM)

3.2 Results and discussion

3.2.1 SuTEx fragment design and synthesis

The fragment library was synthesized by first coupling amine to 4-(chlorosulfonyl) benzoyl chloride, followed by coupling with the 1,2,4-triazole (Fig 1). The molecular weight of the fragment library ranged from 292 to 396 Da. Fragments were selected based on similarity to motifs found in natural and synthetic clinically approved drugs.

3.2.2 Chemical proteomic profiling of SuTEx fragments in DM93 live cells

Cognizant of some covalent fragments showing high activity in lysates but different profiles in live cells[12], we performed all of our experiments in live cells to provide the greater complexity of a native biological system. First, we set out to elucidate the protein targets of our fragment library using chemical proteomics. Briefly, DM93 melanoma cells were cultured using the stable isotopic labeling by amino acids in cell culture (SILAC)[13] method, in which cells cultured in "light" media (supplemented with [$^{12}C_6{}^{14}N_2$] L-lysine and [$^{12}C_6{}^{14}N_4$] L-arginine) were treated with dimethyl sulfoxide (DMSO) vehicle and cells cultured in "heavy" media (supplemented with [$^{13}C_6{}^{15}N_2$] L-lysine and [$^{13}C_6{}^{15}N_4$] L-arginine) were treated with SuTEx fragment (25 μM, 2 hours). All cells then were treated with HHS-475 alkyne probe (100 μM, 2 hours) and harvested. The probe-modified proteins from the soluble fraction were conjugated to a desthiobiotin-azide tag by copper-catalyzed azide-alkyne cycloaddition (CuAAC)[14], and the light and heavy proteomes combined. The proteomes were digested with trypsin protease to produce probe-modified peptides, followed by enrichment and purification by avidin affinity chromatography. Probe-modified peptides then were eluted and analyzed by high-resolution liquid chromatography-mass spectrometry (LC-MS/MS). Resultant data were analyzed as previously described[10] (Fig 2A).

SuTEx fragments were screened across independent biological replicates (n=2-3), and target peptides and proteins were determined by the following metrics: identified in more than one biological replicate, satisfied quality control peptide-MS2 match confidence criteria of ≥300 Byonic™ score[15], and ≤ 5 ppm mass accuracy. The ratio of light-peptide (DMSO treated) to heavy-peptide (SuTEx fragment treated) precursor ion abundances (peak areas), referred to as SILAC ratio (SR), was used to identify sites competed by the SuTEx fragments. As expected, the fragment library showed negligible interaction with amino acid residues other than tyrosine, so we focused

our efforts on investigating the modified tyrosine sites. We chose to evaluate tyrosine sites that showed ≥ 50% (SR ≥ 2) reduced enrichment by HHS-475 after pretreatment with the fragment library. The SuTEx fragments exhibited high selectivity, liganding 97 unique tyrosine sites on 95 proteins, representing 6% and 10% of total quantified tyrosines and proteins, respectively (Fig. 2B&C). The liganded tyrosines were enriched for functional domains involved in RNA binding (RNA recognition motif (RRM) and SUZ domain), nucleotide exchange (DHR-2 domain), and regulation of DNA-templated transcription (MAD homology domain 2 (MH2)) (Fig. 2E). Additionally, ~50 % of the liganded sites (SR ≥ 2) are not in the Drugbank database[16] (Fig. 2D), highlighting the potential important use of the SuTEx fragments in drug discovery

To validate LC-MS/MS results by competitive gel-based assay, we chose three functionally diverse proteins: fumarylacetoacetase (FAH), abhydrolase domain containing 10 (ABHD10) and prostaglandin reductase 2 (PTGR2) because of their reported roles in human diseases

3.2.3 ABPP profiling of FAH

FAH catalyzes the hydrolysis of fumarylacetoacetate to fumarate and acetoacetate in the degradation of tyrosine and phenylalanine [17]. LC-MS/MS analysis showed that HHS-0401 engaged Y244 of FAH (Fig 2B and Fig 3C), which was previously reported to be a hyperreactive tyrosine site[10]. HHS-0601 was identified as inactive against FAH and, therefore, used as a control compound. We confirmed these results by gel-based ABPP. HEK293T cells that expressed recombinant FAH were treated with SuTEx fragments (25 μM, 2 hours) followed by treatment with HHS-475 (100 μM, 2 hours) and then cells were harvested. The probe-labeled proteins were reacted with rhodamine azide and visualized by in-gel fluorescence scanning. We found that HHS-0401, but not HHS-0601, blocked labeling of FAH by HHS-475 in a concentration-dependent manner (IC_{50} = 1.8 μM, Fig 3A and B). Deficiency of FAH leads to the accumulation of

fumarylacetoacetate causing hepatorenal toxicity, which can lead to the fatal condition tyrosinemia type 1[18]. The development of a FAH inhibitor is important to obtain a better understanding of tyrosinemia type 1 and to determine how its effects can be alleviated.

3.2.4 ABPP profiling of ABHD10

ABHD10 is a putative mitochondrial localized protein belonging to the metabolic serine hydrolase superfamily and has been implicated in the deglucurodination of acyl glucuronide[19]. Recently Cao et al. reported PRDX5[20] as the first endogenous ABHD10 substrate. They discovered that ABHD10 modulates mitochondrial antioxidant function through the depalmitoylation of C100 of PRXD5. Encouraged by their findings, we set out to investigate the effect of liganding tyrosine sites on ABHD10. Because ABHD10 is a serine hydrolase, FP-rhodamine probe was used to evaluate ABHD10 activity in live cells and mouse brain tissue. LC-MS/MS studies showed that HHS-0301, 0401, and 0701 binds to Y215 on ABHD10 (Fig 2B and Fig 4B). To validate these results by gel-based ABPP, live DM93 cells were treated with varying concentrations of SuTEx fragments for 2 hours, and then the soluble fractions were treated with FP-rhodamine (0.5 hours, 37 °C). Additionally, because ABHD10 is abundantly expressed in the brain, we treated soluble mouse brain lysates with SuTEx fragments followed by FP-rhodamine (0.5 hours, 37 °C). Proteins were separated by SDS-PAGE and visualized by in-gel fluorescence scanning. HHS-0301, 0401 and 0701 in DM93 cells and mouse brain lysates inhibited FP-rhodamine labeling in a concentration-dependent manner (IC_{50} ~5 μM, Fig 4A).

Although SuTEx fragments are not as potent as aza-β-lactam ABL303 ($IC_{50} \approx 30$ nM)[21], they inhibit ABHD10 activity through a different mechanism. The two mechanisms may be key to identifying different ABHD10 endogenous substrates, which will be a focus of our future work.

3.2.5 ABPP profiling of PTGR2

PTGR2 catalyzes the NADPH-dependent reduction of 15-keto-PGE2, an endogenous ligand for the nuclear receptor PPARγ, to 13,14-dihydro-15-keto-PGE2[22]. PTGR2 has been found to be expressed in pancreatic cancer tissues but absent in normal pancreatic tissue. Knockdown of PTGR2 was found to reduce tumor growth and induce apoptosis through ROS-mediated signaling involving ERK 1/2 and caspase 3 activities[23]. To the best of our knowledge, the only reported PTGR2 inhibitors are reversible[24] or exhibit weak *in vitro* activity[25]. LC-MS/MS studies identified HHS-0701 as a potent PTGR2 Y100 inhibitor, while HHS-0101 exhibited no activity and was used as a negative control (Fig 2B and Fig 5B). Gel-based ABPP confirmed that HHS0701, but not HHS0101, blocked labeling of PTGR2 by HHS475 in a concentration-dependent manner (IC_{50} = 1 μM, Fig 5A).

3.2.6 Liganding PTGR2 Y100 affects protein function

Next, we sought to investigate the effect of liganding tyrosine sites on protein function. We chose PTGR2 because of the aforementioned role in human disease and the availability of a published, commercial PTGR2 activity assay, the 15-Keto PGE_2 substrate assay[25-26]. Briefly, HEK293T cells transiently transfected with PTGR2 were treated with DMSO (vehicle control), HHS-0701, and negative control HHS-0101 for 2 hours. Cells were processed to lysates, and the soluble fractions were incubated with EDTA (1 mM), DTT (1 mM), NADPH (1 mM), and 15-Keto PGE_2 (20 μM) for 0.5 hours at 37°C. We quantified the amount of unconsumed substrate by adding NaOH followed by absorbance measurement at 500 nm (15-keto-PGE2 forms a labile red chromophore in alkaline solution)[25]. We observed concentration-dependent inhibition of PTGR2 activity by HHS-0701 (IC_{50} = 1.5 μM, Figure 5D), with a significant effect observed at 25 μM. No effect was observed with the negative control HHS-0101. Our observations support the PTGR2-15-keto-

PGE2 binding mode analysis reported by Wu et al. that the carboxylic head of the 15-keto-PGE_2 α chain is hydrogen bound to the hydroxyl group of Y100[25].

Collectively, these results highlight the importance of the Y100 site on enzyme activity. Considering that the role of the PTGR2 substrate 15-keto PGE2 is to function as an endogenous ligand of PPARγ, which regulates adipocyte differentiation and lipid homeostasis[25], a potent, covalent, and *in situ* active PTGR2 ligand is a significant step toward developing a novel therapeutic agent. Future studies will focus on optimizing SuTEx fragments to improve the potency and selectivity for live cell and *in vivo* studies.

3.3 Conclusion

We have demonstrated the utility of SuTEx platform to develop active live cell, covalent small molecule fragments exhibiting high selectivity to disrupt functional tyrosines. Using PTGR2 as an example, we showed that targeting tyrosine sites in live cells can affect the enzymatic function of the protein. Of note, the PTGR2 inhibitor HHS-0701 and the PTGR2 reversible probe **8** reported by Parker et al.[24]. have the phenylpiperidine moiety in common, which supports the concept of converting reversible inhibitors to covalent reagents by introducing an electrophilic warhead[27].

The gel-based ABPP results of ABHD10, FAH, and PTGR2 inhibition with the PTGR2 enzymatic activity assay results validate our chemical proteomics results. In future studies, we will seek to understand the biological effects of inhibition of other liganded vital proteins, such as acetyl-CoA acetyltransferase 2 (ACAT2) and dedicator of cytokinesis protein 5 (DOCK5). ACAT2 is involved in regulating lipid metabolism by catalyzing the synthesis of acetoacetyl-CoA from two acetyl-CoA molecules, which is later converted to cholesterol. Mouse model studies showed reduced atherosclerosis in ACAT2 deficient models[28]. Dock5, a guanine nucleotide

exchange factor for the small GTPase Rac, has been implicated in bone resorption by osteoclasts[29], and inhibition of DOCK5 has been shown to reduce bone degradation in mice[30]. Our LC-MS/MS studies identified HHS-0201 and HHS-0601 as potent ACAT2 and DOCK5 ligands, respectively, potentially providing chemical tools to study these key proteins.

The small molecule screening results highlight the potential benefits of expanding the SuTEx fragment library to target the largely untapped human proteome. To do that, we plan to expand our current ~150 1,2,4 sulfonyl triazoles library and explore other heterocyclic moieties, such as imidazole, pyrazole, and tetrazole, as the core of the leaving group. Moreover, our current fragment library and probes lack 3-dimensional character, and I think it will be worthwhile exploring introducing stereogenic character to investigate the difference in activity between diastereomers. Preliminary results indicate that the imidazole, pyrazole, and tetrazole moieties might be less reactive than the triazoles, however, that might improve selectivity. To compensate for the reduced reactivity, we can take advantage of the advances in targeted protein degradation and proteolysis-targeting chimeras (PROTACs) technology to reduce the abundance of the targeted protein[31]. Protein degradation technology could turn weakly binding ligands to effective inhibitors, enabling the use of low doses, resulting in improved selectivity, and subsequently fewer off-target effects.

Overall, we believe SuTEx can serve as a valuable platform to reveal new biological insights and can be suitable warheads for drugs because of their combined reactivity, selectivity and tunability. Currently there are about 50 FDA-approved covalent drugs, with 14 of them approved in the last 10 years. Among the newly approved drugs, 9 of them have α, β-unsaturated carbonyl as their warhead (target cysteine)[32]. Most recently, BTK inhibitor zanubrutinib and EGFR inhibitor dacomitinib were approved to treat adults with mantle cell lymphoma and

metastatic non-small cell lung cancer, respectively. Apart from FDA-approved covalent drugs, there are many more inhibitors in late-stage clinical development. For example, AMG510 (Amgen) and MRTX-849 (Mirati Therapeutics) targeting RAS protein, which for decades was considered "undruggable" because their smooth surface did not offer any obvious pockets to target with a drug[33]. The fact that about 77 % of the proteins labeled by SuTEx probes and ~50 % of the proteins liganded by SuTEx fragments were absent from the DrugBank database, highlight the important role SuTEx can play in drug discovery. Future work will focus on screening SuTEx fragments against proteins implicated in diseases with high unmet need.

3.4 Figures

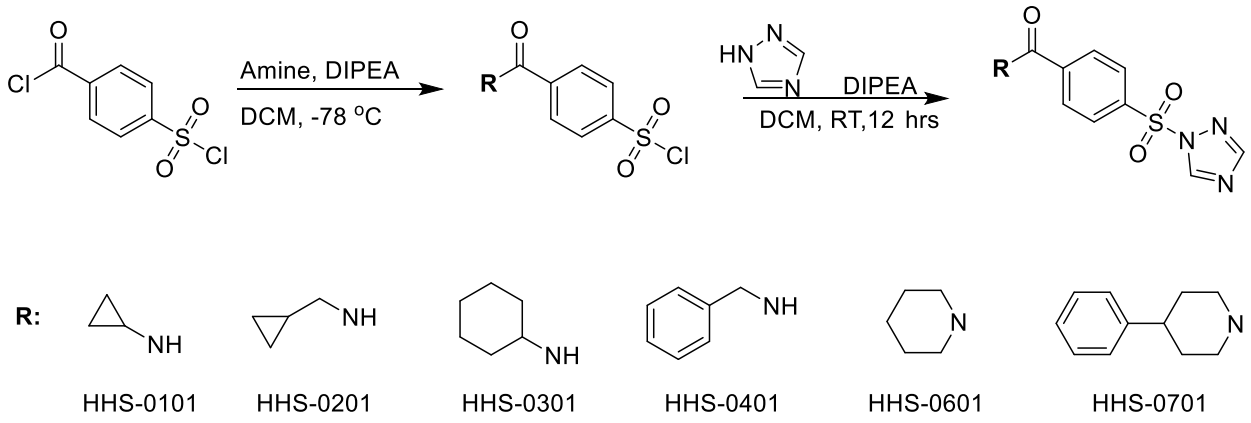


Figure 1. SuTEx fragment library general synthesis scheme

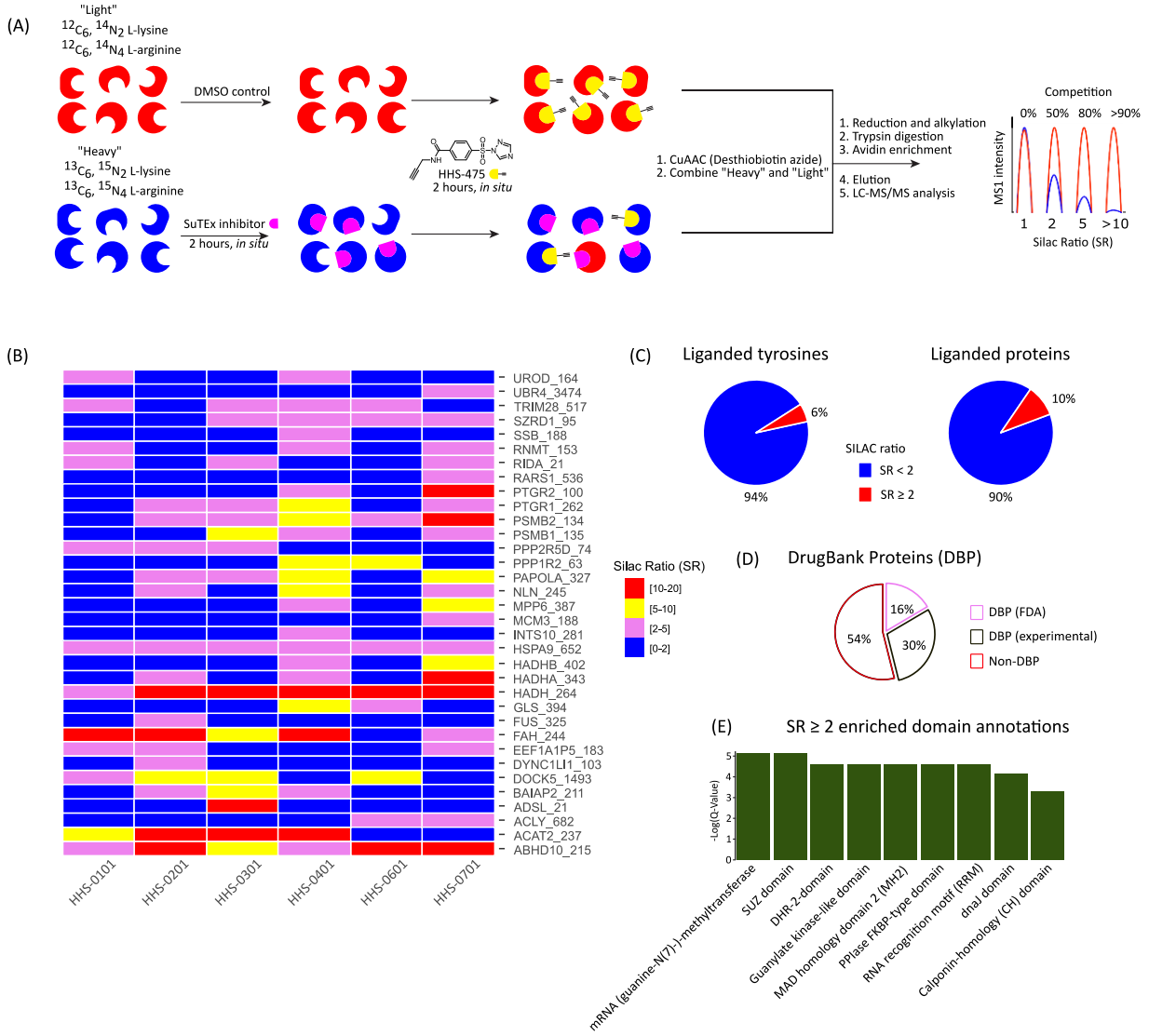


Figure 2. Activity-based protein profiling (ABPP) of SuTEx fragments in DM93 live cells. (A) Experimental overview of quantitative chemical proteomic profiling of SuTEx fragments reactivity and selectivity. (B) Heatmap representation of all tyrosine sites ≥66% (SR ≥3) competed by at least one of the SuTEx fragments. The SILAC ratio (SR) is a quantification of the area under the curve of MS1 extracted ion chromatograms (EIC) from HHS-475 labeled peptides in DMSO control (light, red) versus fragment-treated (heavy, blue) proteomes. (C) Distribution of liganded and non-liganded tyrosine sites and proteins. (D) Distribution of liganded proteins (SR > 2) found

in DrugBank (DBP group) compared with proteins that did not match a DrugBank entry (Non-DBP) (E) Distribution of protein domain groups targeted by SuTEx fragments with SR ≥2 and meet Q < 0.05 after Benjamini−Hochberg correction of a two-sided binomial test criteria.

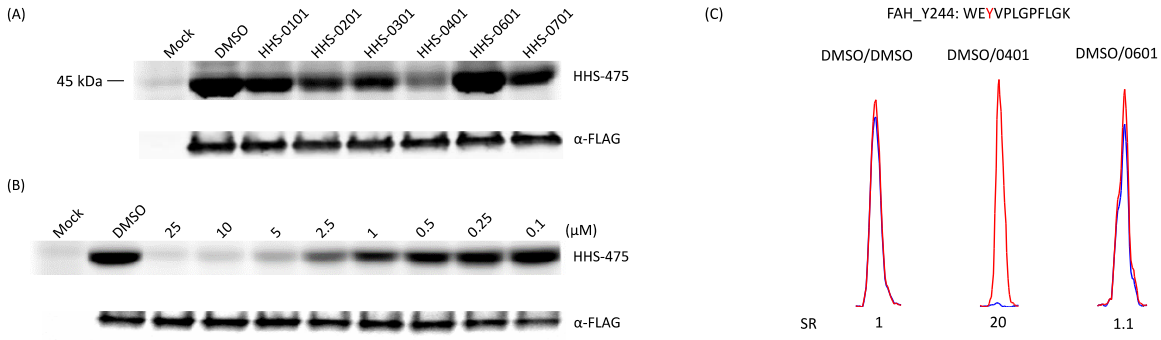


Figure 3. Chemical proteomic analysis of SuTEx fragment library profiling with FAH. (A) Gel-based analysis of recombinant FAH treated with vehicle or 25 μM of SuTEx fragment, followed by treatment with 100 μM of HHS-475, validated HHS-0401 as FAH lead inhibitor (B) FAH dose-dependent inhibition by HHS-0401 (IC_{50}=2 μM) (C) EIC representation of LC-MS/MS analysis showing that FAH Y244 is inhibited by HHS-0401 and not HHS-0601.

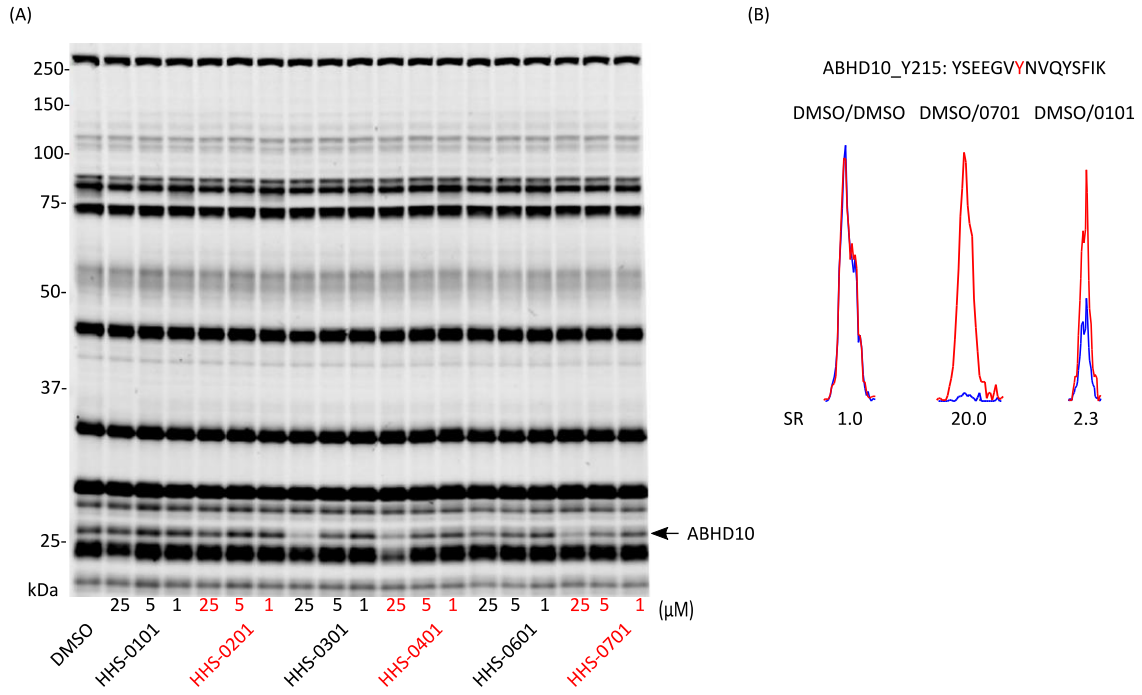


Figure 4. ABPP profiling of SuTEx fragments against serine hydrolases. (A) Competitive ABPP gel results of soluble mouse brain treated with indicated concentrations of SuTEx fragments, followed by treatment with 50 μM FP-rhodamine validated HHS-0301,0401 and 0701 as ABHD10 inhibitors (IC_{50} =5 μM). (B) EIC representation of ABHD10 Y215 site blockade by SuTEx fragments.

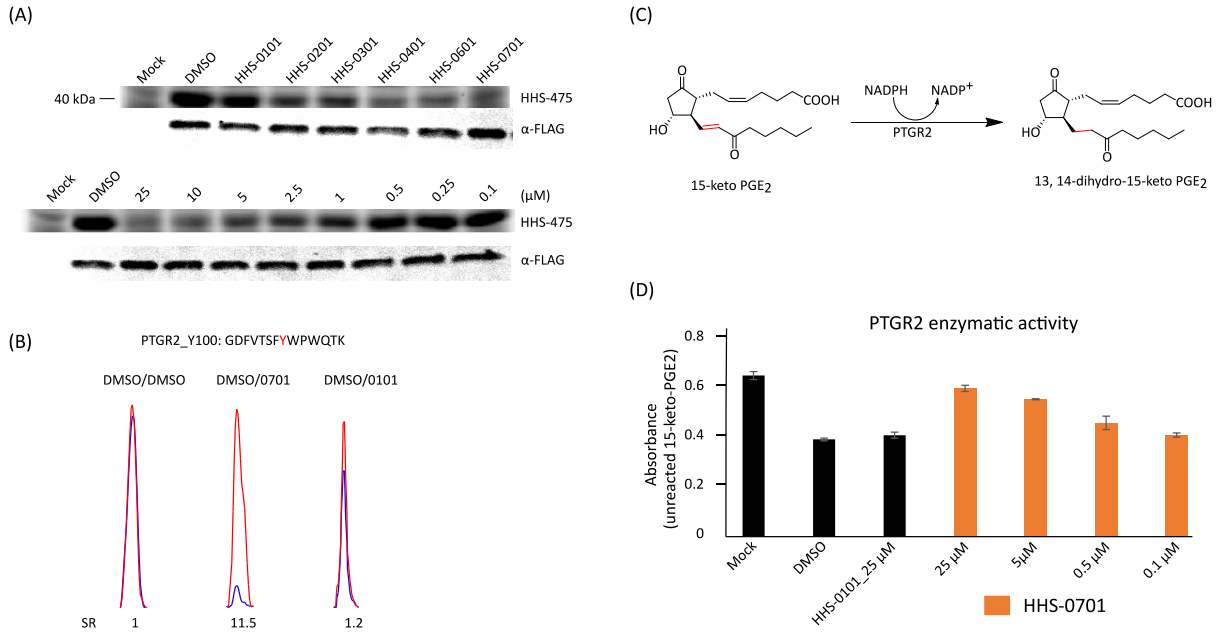


Figure 5. Effect of treating recombinant PTGR2 cells with SuTEx fragments on PTGR2 enzymatic activity. (A) Gel-based analysis of recombinant PTGR2 expressed in HEK293Tcells. Live cells treated with vehicle or 25 μM of SuTEx fragment followed by treatment with 100 μM of HHS-475 validates HHS-0701 as PTGR2 lead inhibitor, and HHS-0101 as negative control. (B) PTGR2 dose-dependent inhibition by HHS-0701 (IC_{50} = 1 μM) (C) Scheme showing PTGR2 NADPH mediated role in reducing 15-keto PGE2 to 13-14 dihydro 15-keto PGE2. (D) PTGR2 Y100 inhibitor, HHS-0701, but not inactive control, HHS-0101, reduced PTGR2 reductase activity as measured using 15-keto substrate assay (IC_{50} = 1.5 μM). All data shown are representative of n=2-3 biologically independent experiments.

3.5 Methods

Cell culture. Cell lines were grown at 37 ˚C with 5% CO_2. DM93 cells were grown in RPMI medium supplemented 10% fetal bovine serum (Omega Scientific), penicillin, streptomycin and L-glutamine (Fisher Scientific). HEK293T cells used for transfection were grown in DMEM medium supplemented with 10 % fetal bovine serum and glutamine.

SILAC cell culture. SILAC DM93 cells were grown at 37 ˚C with 5% CO_2 in either "light" or "heavy" media supplemented with 10% dialyzed fetal bovine serum (Omega Scientific), penicillin, streptomycin, L-glutamine (Fisher Scientific) and isotopically labelled L-lysine and L-arginine amino acids. "Light" media was supplemented with 100 µg/mL [$^{12}C_6{}^{14}N_2$] L-lysine and 100 µg/mL [$^{12}C_6{}^{14}N_4$] L-arginine. "Heavy" media was supplemented with 100 µg/mL [$^{13}C_6{}^{15}N_2$] L-lysine and 100 µg/mL [$^{13}C_6{}^{15}N_4$] L-arginine. The cells were grown for six passages to incorporate the amino acids before being used for proteomics experiments.

Sample preparation for LC-MS/MS analysis

All cells were grown in 10 cm plates to ~90% confluence. The media was replaced with serum-free media supplemented with only penicillin and streptomycin. Light cells were treated with DMSO vehicle while the heavy cells were treated with SuTEx fragments (10 µL of 1,000x DMSO stock) and incubated for 2 hours. Thereafter, the cells were treated with SuTEx alkyne probe (HHS475, 40 µL of 250x DMSO stock) and incubated for 2 more hours. The media was aspirated and cells washed with cold PBS twice and harvested. The cells were lysed in PBS by sonication and fractionated (100,000g, 45 min, 4 ˚C). Separately, the soluble fractions of heavy and light cells were diluted to 2.3 mg/mL in PBS and 432 µL were used for analysis (1 mg total protein). The

probe-modified proteomes were conjugated to desthiobiotin-PEG3-azide (10 µl of 10 mM stock in DMSO; 200 µM final concentration) using TCEP (10 µl of fresh 50 mM stock in water, 1 mM final concentration), TBTA ligand (33 µl of a 1.7 mM 4:1 t-butanol/DMSO stock, 100 µM final concentration) and CuSO4 (10 µl of 50 mM stock, 1 mM final concentration). Samples were vortexed to mix and incubated for 1 hour at room temperature. The heavy and light proteomes were mixed 1:1 and the subsequent steps, including LC-MS/MS data analysis, were performed as previously described[10].

Heatmap generation. All tyrosine sites ≥66% (SR ≥3) competed by at least one of the SuTEx fragments were identified, and using VLOOKUP matched the SILAC ratios to the treatment conditions. The heatmap was generated in RStudio (v.1.2.5033) using the ggplot2 algorithm.

Plasmid Construction and Amplification

Plasmid constructs (human protein) were purchased from GenScript. pcDNA 3.1-PTGR2-FLAG and pcDNA 3.1-FAH-FLAG. Plasmids were amplified through transforming XL-1 Blue *E. coli* through electroporation with 10 ng of plasmid. Transformed bacteria was grown in terrific broth (TB, 1mL) for 45 minutes at 37°C. An aliquot of this bacteria was plated onto agar bacterial growth plates that had been made with carbenicillin (100µg/mL), and plates were incubated at 37°C for 14-18 hours. Plates were stored at 4°C and wrapped in parafilm for no longer than one month for use. Single bacterial colonies were scraped from the agar plates and grown in Falcon tubes with terrific broth (5mL) supplemented with carbenicillin (100µg/mL) at 37°C for 14-18 hours while shaking. Plasmid was extracted and purified using QIAPrep Spin Miniprep Kit protocols and materials and stored at -80°C until needed.

ABPP gel-based profiling of SuTEx against recombinant FAH and PTGR2

HEK293T cells at ~30 % confluency were transfected with FAH or PTGR2 plasmid DNA for 48 hrs. The media was replaced with serum free media, and the cells were treated with SuTEx fragments at indicated concentrations and incubated for 2 hours. Thereafter, the cells were treated with 100 µM of SuTEx alkyne probe (HHS475, 40 µL of 250x DMSO stock) and incubated for 2 more hours. The cells were washed with cold PBS twice and harvested. The cells were lysed in PBS by sonication and fractionated (100,000g, 45 min, 4 ˚C). The soluble fraction was diluted to 1 mg/mL in PBS and 49 µL was used for analysis. The probe-modified proteomes were conjugated to Rhodamine-azide (1 µl of 1.25 mM stock in DMSO) using TCEP (1 µl of fresh 50 mM stock in water), TBTA ligand (3 µl of a 1.7 mM 4:1 t-butanol/DMSO stock,) and CuSO4 (1 µl of 50 mM stock) and incubated for 1 hour at room temperature. The reaction was quenched with 17 µL of 4X SDS-PAGE loading buffer + βME and vortexed to mix. The samples were analyzed by SDS-PAGE and imaged by in-gel fluorescence scanning.

Western Blotting

Western blot analysis of recombinant protein expression of FAH and PTGR2 was performed as previously described[34]. Cell lysates were separated via centrifugation at 100,000xg for 45 min at 4°C. Proteins were separated by SDS-PAGE (7.5% polyacrylamide, TGX Stain-Free Mini Gel) at 150 V for 45 min. Gel transfers were performed using the Bio-Rad Trans-Blot Turbo RTA Midi Nitrocellulose Transfer Kit with a Bio-Rad Trans-Blot Turbo Transfer System (25V, 10 min). The nitrocellulose blot was incubated in blocking solution (40 mL, 3% bovine serum albumin (BSA) in TBS-T (1.5 M NaCl, 0.25 M Tris pH 7.4 in ddH_2O with 0.1% (v/v%) Tween^{TM} 20)) for 1 h at 25°C with gentle shaking. The blot was transferred immediately to primary antibody solution (See

manufacturer specifications for primary antibody concentrations; 40 mL, 3% bovine serum albumin (BSA) in TBS-T) and incubated overnight at 4°C with gentle shaking. The blot was rinsed 5 times for 5 min in TBS-T, transferred immediately into secondary antibody solution (1:10,000 anti-species DyLight 550 in TBS-T), and incubated for 1 h at 25°C with gentle shaking. The blot was rinsed 5 times for 5 min in TBS-T, transferred into ddH_2O, and imaged by in-blot fluorescence scanning on a ChemiDoc MP Imaging System.

PTGR2 enzymatic assay. PTGR2 enzymatic assay was performed based on chromogenic method[25-26]. HEK293T cells expressing PTGR2 were treated with DMSO vehicle or indicated concentrations of SuTEx fragments for 2 hours. The cells were washed with cold PBS twice and harvested. The proteomes were lysed in PBS by sonication and fractionated (100,000 x g, 45 min, 4 ˚C). 0.5 mg of soluble fraction was incubated for 30 min at 37 °C with 1 mM of EDTA, DTT, NADPH and 20 µM 15-Keto PGE_2 in 0.1 M Tris-HCl (PH 7.5) (final volume 230 µL) and the substrate consumed was determined by adding 20 µL 20 N NaOH, mixing and measuring absorbance at 500 nm after 5 minutes.

References

1. Gehringer, M.; Laufer, S. A., Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *Journal of Medicinal Chemistry* 2019, *62* (12), 5673-5724.
2. Counihan, J. L.; Ford, B.; Nomura, D. K., Mapping proteome-wide interactions of reactive chemicals using chemoproteomic platforms. *Curr Opin Chem Biol* 2016, *30*, 68-76.
3. Tuley, A.; Fast, W., The Taxonomy of Covalent Inhibitors. *Biochemistry* 2018, *57* (24), 3326-3337.
4. Bum-Erdene, K.; Liu, D.; Gonzalez-Gutierrez, G.; Ghozayel, M. K.; Xu, D.; Meroueh, S. O., Small-molecule covalent bond formation at tyrosine creates a binding site and inhibits activation of Ral GTPases. *Proceedings of the National Academy of Sciences* 2020, *117* (13), 7131-7139.
5. Zhang, T.; Hatcher, J. M.; Teng, M.; Gray, N. S.; Kostic, M., Recent Advances in Selective and Irreversible Covalent Ligand Development and Validation. *Cell Chemical Biology* 2019, *26* (11), 1486-1500.
6. Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F., Global profiling of lysine reactivity and ligandability in the human proteome. *Nature chemistry* 2017, *9* (12), 1181-1190.
7. Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F., Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* 2010, *468* (7325), 790-5.
8. Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; González-Páez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J.; Wolan, D. W.; Cravatt, B. F., Proteome-wide covalent ligand discovery in native biological systems. *Nature* 2016, *534* (7608), 570-574.
9. Antolin, A. A.; Tym, J. E.; Komianou, A.; Collins, I.; Workman, P.; Al-Lazikani, B., Objective, Quantitative, Data-Driven Assessment of Chemical Probes. *Cell Chem Biol* 2018, *25* (2), 194-205.e5.
10. Hahm, H. S.; Toroitich, E. K.; Borne, A. L.; Brulet, J. W.; Libby, A. H.; Yuan, K.; Ware, T. B.; McCloud, R. L.; Ciancone, A. M.; Hsu, K.-L., Global targeting of functional tyrosines using sulfur-triazole exchange chemistry. *Nature Chemical Biology* 2020, *16* (2), 150-159.
11. Brulet, J. W.; Borne, A. L.; Yuan, K.; Libby, A. H.; Hsu, K.-L., Liganding Functional Tyrosine Sites on Proteins Using Sulfur–Triazole Exchange Chemistry. *Journal of the American Chemical Society* 2020.
12. Lee, H. Y.; Suciu, R. M.; Horning, B. D.; Vinogradova, E. V.; Ulanovskaya, O. A.; Cravatt, B. F., Covalent inhibitors of nicotinamide N-methyltransferase (NNMT) provide evidence for target engagement challenges in situ. *Bioorganic & medicinal chemistry letters* 2018, *28* (16), 2682-2687.
13. Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics* 2002, *1* (5), 376-386.
14. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A stepwise huisgen cycloaddition process: copper (I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angewandte Chemie* 2002, *114* (14), 2708-2711.
15. Bern, M.; Kil, Y. J.; Becker, C., Byonic: advanced peptide and protein identification software. *Current protocols in bioinformatics* 2012, *Chapter 13*, Unit13.20.
16. Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M., DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* 2018, *46* (D1), D1074-d1082.
17. Lindblad, B.; Lindstedt, S.; Steen, G., On the enzymic defects in hereditary tyrosinemia. *Proc Natl Acad Sci U S A* 1977, *74* (10), 4641-4645.
18. Nyhan, W. L.; Haas, R., Inborn Errors of Amino Acid Metabolism. In *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease*, Elsevier: 2015; pp 627-632.

19. Iwamura, A.; Fukami, T.; Higuchi, R.; Nakajima, M.; Yokoi, T., Human alpha/beta hydrolase domain containing 10 (ABHD10) is responsible enzyme for deglucuronidation of mycophenolic acid acyl-glucuronide in liver. *The Journal of biological chemistry* 2012, *287* (12), 9240-9.
20. Cao, Y.; Qiu, T.; Kathayat, R. S.; Azizi, S.-A.; Thorne, A. K.; Ahn, D.; Fukata, Y.; Fukata, M.; Rice, P. A.; Dickinson, B. C., ABHD10 is an S-depalmitoylase affecting redox homeostasis through peroxiredoxin-5. *Nature Chemical Biology* 2019, *15* (12), 1232-1240.
21. Zuhl, A. M.; Mohr, J. T.; Bachovchin, D. A.; Niessen, S.; Hsu, K.-L.; Berlin, J. M.; Dochnahl, M.; López-Alberca, M. P.; Fu, G. C.; Cravatt, B. F., Competitive Activity-Based Protein Profiling Identifies Aza-β-Lactams as a Versatile Chemotype for Serine Hydrolase Inhibition. *Journal of the American Chemical Society* 2012, *134* (11), 5068-5071.
22. Chou, W. L.; Chuang, L. M.; Chou, C. C.; Wang, A. H.; Lawson, J. A.; FitzGerald, G. A.; Chang, Z. F., Identification of a novel prostaglandin reductase reveals the involvement of prostaglandin E2 catabolism in regulation of peroxisome proliferator-activated receptor gamma activation. *The Journal of biological chemistry* 2007, *282* (25), 18162-72.
23. Chang, E. Y.; Chang, Y. C.; Shun, C. T.; Tien, Y. W.; Tsai, S. H.; Hee, S. W.; Chen, I. J.; Chuang, L. M., Inhibition of Prostaglandin Reductase 2, a Putative Oncogene Overexpressed in Human Pancreatic Adenocarcinoma, Induces Oxidative Stress-Mediated Cell Death Involving xCT and CTH Gene Expressions through 15-Keto-PGE2. *PloS one* 2016, *11* (1), e0147390.
24. Parker, C. G.; Galmozzi, A.; Wang, Y.; Correia, B. E.; Sasaki, K.; Joslyn, C. M.; Kim, A. S.; Cavallaro, C. L.; Lawrence, R. M.; Johnson, S. R.; Narvaiza, I.; Saez, E.; Cravatt, B. F., Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* 2017, *168* (3), 527-541 e29.
25. Wu, Y. H.; Ko, T. P.; Guo, R. T.; Hu, S. M.; Chuang, L. M.; Wang, A. H., Structural basis for catalytic and inhibitory mechanisms of human prostaglandin reductase PTGR2. *Structure (London, England : 1993)* 2008, *16* (11), 1714-23.
26. Hansen, H. S., Purification and assay of 15-ketoprostaglandin delta 13-reductase from bovine lung. *Methods in enzymology* 1982, *86*, 156-63.
27. Ghosh, A. K.; Samanta, I.; Mondal, A.; Liu, W. R., Covalent Inhibition in Drug Discovery. *ChemMedChem* 2019, *14* (9), 889-906.
28. Bell III, T. A.; Brown, J. M.; Graham, M. J.; Lemonidis, K. M.; Crooke, R. M.; Rudel, L. L., Liver-Specific Inhibition of Acyl-Coenzyme A: Cholesterol Acyltransferase 2 With Antisense Oligonucleotides Limits Atherosclerosis Development in Apolipoprotein B100–Only Low-Density Lipoprotein Receptor−/− Mice. *Arteriosclerosis, thrombosis, and vascular biology* 2006, *26* (8), 1814-1820.
29. Vives, V.; Laurin, M.; Cres, G.; Larrousse, P.; Morichaud, Z.; Noel, D.; Côté, J. F.; Blangy, A., The Rac1 exchange factor Dock5 is essential for bone resorption by osteoclasts. *Journal of Bone and Mineral Research* 2011, *26* (5), 1099-1110.
30. Vives, V.; Cres, G.; Richard, C.; Busson, M.; Ferrandez, Y.; Planson, A.-G.; Zeghouf, M.; Cherfils, J.; Malaval, L.; Blangy, A., Pharmacological inhibition of Dock5 prevents osteolysis by affecting osteoclast podosome organization while preserving bone formation. *Nature communications* 2015, *6* (1), 1-9.
31. Zou, Y.; Ma, D.; Wang, Y., The PROTAC technology in drug development. *Cell biochemistry and function* 2019, *37* (1), 21-30.
32. Sutanto, F.; Konstantinidou, M.; Dömling, A., Covalent inhibitors: a rational approach to drug discovery. *RSC Medicinal Chemistry* 2020.
33. Khan, I.; Rhett, J. M.; O'Bryan, J. P., Therapeutic targeting of RAS: New hope for drugging the "undruggable". *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 2020, *1867* (2), 118570.
34. Franks, C. E.; Campbell, S. T.; Purow, B. W.; Harris, T. E.; Hsu, K.-L., The ligand binding landscape of diacylglycerol kinases. *Cell chemical biology* 2017, *24* (7), 870-880. e5.

Chapter 4: Conclusion and future directions

We have demonstrated that sulfur triazole exchange chemistry is a powerful platform to investigate tyrosine reactivity, function, and post-translational modification states in lysates and live cells. Replacing the fluorine with triazole leaving group not only increased the coverage of modified sites by ~4 fold, but also improved tyrosine chemoselectivity (after manual annotation to eliminate false lysine annotations, the tyrosine-to-lysine ratio for SuFEx was ~1, and ≥3 for SuTEx probes). Furthermore, tuning the triazole by adding a phenyl group improved the coverage of modified sites and tyrosine chemoselectivity. In total, the SuTEx probes modified >10,000 distinct tyrosine sites in ~3700 proteins in live cells and lysates. Strikingly, ~70 % of the proteins labeled by the probes were not in the DrugBank database and included proteins in RNA-recognition motifs and domains mediating protein-protein interactions, which have been historically challenging to target with small molecules. Noteworthy, each probe tested modified at least 112 tyrosine sites with HHS-475, -481 and -483 each modifying ~400 unique tyrosine sites. As part of future work, to expand the depth and breadth of proteome coverage we plan to further diversify the probes to include heterocyclic moieties such as pyridine, pyran, and purine, which are involved in hydrogen bonding interactions and might reveal novel targets.

Considering the low frequency of tyrosine phosphorylation (< 2 %) compared to phosphoserines and phosphothreonines, it was remarkable that SuTEx probes captured ~30 % of annotated phosphotyrosine sites in live cells. We hypothesize that the SuTEx platform can be used to identify unknown phosphotyrosine sites. For example, our chemical proteomics experiment identified Y81 in ERK1 and Y82 in SHMT1 as pervanadate-sensitive sites but they are not annotated as phosphotyrosine sites. While we do not have enough data to be certain that pervanadate sensitivity is directly correlated to a site being a phosphotyrosine site, the majority of

probe-modified sites in pervanadate-sensitive group were annotated phosphotyrosine sites (HTP>10 in PhosphoSitePlus). Additionally, for the three proteins (STAT3, CTNND1, and PKM) for which we found phosphoantibodies, the chemical phosphoproteomics results matched the western blot analysis and published work. Considering the important role of phosphorylation and that the current phosphoproteomics tools such as immobilized metal affinity chromatography (IMAC) are more suited for the more abundant phosphoserine and phosphothreonine, SuTEx can bridge the gap. To further validate the platform, we can work with research groups or companies to generate phospho (pY) antibodies so that we can confirm the phosphoproteomics results.

To expand the utility of SuTEx platform, we designed small molecule fragments based on the tested probes with the goal of targeting proteins absent in DrugBank database. Remarkably, ~50 % of liganded proteins were not in the DrugBank database. Further, the small-fragment library showed high selectivity and potency in live cells. Using PTGR2, as an example, we demonstrated that liganding tyrosine sites on proteins affects protein function. Our observation that akin to lysine, cysteine, and methionine residues, tyrosines also possess hyperreactive sites, gave us the confidence to treat live cells with a low dose (25μM) of the fragments in our library. Our initial studies identified Y244 of FAH as a hyperreactive site, and our follow-up results showed that the Y244 ligand HHS-0401 inhibited FAH with IC_{50}= 2 μM. Additionally, my colleague observation that SuTEx fragments preferentially reacted with the Y8 hyperreactive site among several tyrosines on GSTP1 strongly suggests that hyperreactive sites are potential targets for drug discovery, and should be further exploited not only for tyrosine but also other residues such as methionine and lysine.

One of the drawbacks with hyperreactive residues is that because of their heightened reactivity, it can nonspecifically react with even moderately electrophilic fragments, leading to

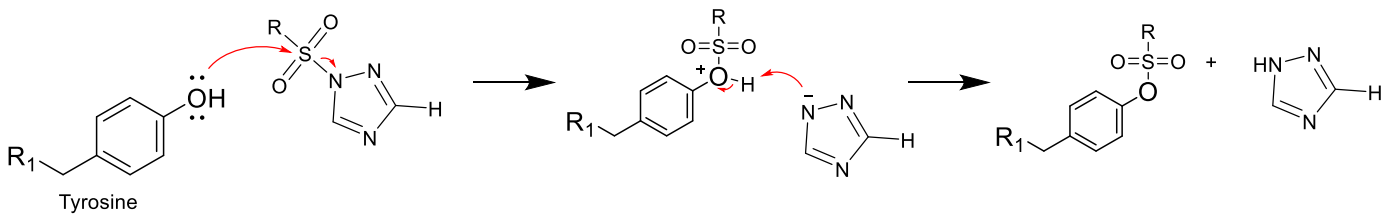
them being common “off-targets”. In our SuTEx fragments screening, out of the six compounds screened, hyperreactive FAH_Y244 and HADH_264 highly reacted (SR > 5) with four and five of the compounds respectively. Therefore, while hyperreactivity can provide opportunity to screen proteomes with low concentrations of small molecule, caution should be taken that the proteins targeted may not result in adverse events, as observed with FAH.

The SuTEx alkyne probes, e.g. HHS-475 and HHS-482, have been valuable in elucidating the binding sites of the small molecule fragments. However, we miss some sites liganded by small molecule fragments by using a structurally-different alkyne probe, in our case HHS-475. As observed with HHS-481, HHS-482, and HHS- 483 which differs by having H, F, or -OMe substituents yet each label about 400 different tyrosine sites. To address this challenge, the small molecule fragments can be designed with the reporter affinity handles such as alkyne and azide appended on them eliminating the need for a second reporter probe. In addition to alkyne and azide groups, Sharpless and coworkers recently reported an efficient strategy of converting primary amines to azides. The proposed strategy, in addition to enabling the accurate mapping of small molecule binding sites, will allow the cells to be treated with lower doses of the inhibitors and still achieve similar efficacy by treating them for a longer time because they react covalently. Currently, we treat the cells with the inhibitors for 2 hours, then add the reporter alkyne probe for 2 more hours and harvest the cells. Having the inhibitors appended with the reporter handle, we can treat the cells for a longer time (e.g., 4 hours) and with lower doses, which might result in improved selectivity and fewer off-target effects.

Lastly, we are striving to understand what drives SuTEx chemoselectivity for tyrosine residues. Although lysine residues are twice as abundant and have a lower pKa than tyrosine residues, the SuTEx probe preferentially binds to tyrosine residues by ∼4 fold. In the solution-

based assays, while it took less than 5 minutes for p-cresol reaction with HHS-475 and HHS-482 to complete with 3.3 equivalent of TMG, it took 6 hours for n-butylamine reaction with HHS-475, and more than 24 hours with HHS-482. While we currently cannot explain with certainty the difference in reactivity between SuTEx probes, we have two hypotheses for the probe's chemoselectivity. First, although tyrosines are protonated at physiological pH, they are weak nucleophiles and can therefore attack the SuTEx probes. We postulate the leaving triazole ion can then deprotonate the tyrosine-sulfonyl complex, stabilizing the complex (Fig. 1a). Using the solution-based assay we tested this hypothesis by reacting p-cresol with HHS-475 in the absence of the TMG base which we had used earlier to deprotonate p-cresol. We found that in the absence of TMG the reaction was not complete even after 12 hours (Fig 1b).

In contrast, when 1.1 equivalent of TMG was added HHS-475 was consumed after 30 minutes, suggesting the phenol group has to be deprotonated for it to attack SuTEx probes. These results were the premise of our second hypothesis that SuTEx reactivity is largely influenced by the protein microenvironment with the reaction first initiated by the positively charged residues such as arginine promoting the deprotonation of phenol by lowering their pKa. This hypothesis is supported by the fact that while most proteins have several tyrosine residues, SuTEx probes label only a few of them, and in the case of proteins such as DPP3 and NUDT12 only a single tyrosine site is modified. Further work is being undertaken to identify if there are any common features surrounding the modified sites that are possibly driving the reactions.



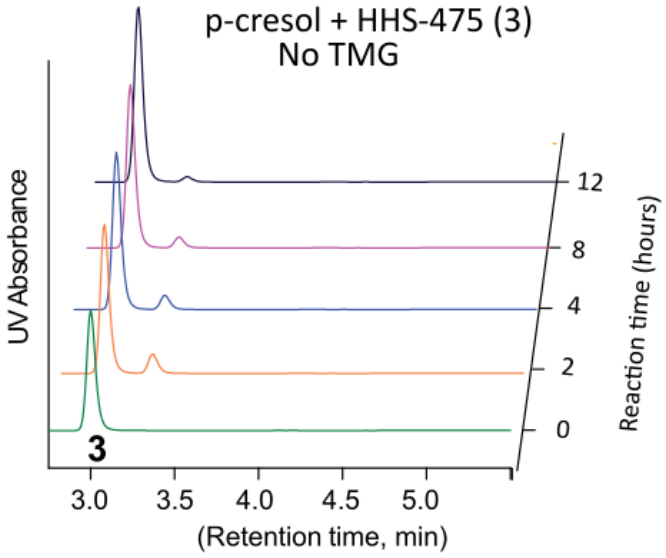


Fig. 1: (A) Proposed SuTEx reaction mechanism. Because tyrosine is a weak nucleophile, we hypothesized it can attack SuTEx probes without being deprotonated (B) Solution-based assay of p-cresol reaction with SuTEx probe (HHS-475) in the absence of base (TMG). After 12 hours the reaction was not complete suggesting that the phenol has to be deprotonated to accelerate the reaction.

Appendix

SuTEx Probe synthesis

General Information

All chemicals used were reagent grade and used as supplied, except where noted. *N*,*N*-Dimethylformamide (DMF), dichloromethane (DCM), toluene and tetrahydrofuran (THF) were used without any further purification steps. Analytical thin layer chromatography (TLC) was performed on Merck silica gel 60 F254 plates (0.25 mm). Flash column chromatography was carried out using forced flow of the indicated solvent on Silica Gel 60 (230-400 mesh) purchased from Fisher Scientific. Compounds were visualized by UV-irradiation and iodine chamber. Analytical HPLC chromatograms were recorded on a Shimadzu 1100 Series spectrometer. ^{1}H and ^{13}C NMR spectra were recorded on a Varian Inova 500 (500 MHz), 600 (600MHz), or Bruker Avance III 800 (800MHz) spectrometers in $CDCl_3$, Acetone-d6, or DMSO-d6 with chemical shifts referenced to internal standards (CD_3OD: 3.31 ppm ^{1}H, 49.03ppm ^{13}C; $(CD_3)_2CO$: 2.05 ppm ^{1}H, 29.84 and 206.26 ppm ^{13}C; $CDCl_3$: 7.26 ppm ^{1}H, 77.16 ppm ^{13}C) unless stated otherwise. Splitting patterns are indicated as s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; br, broad singlet for 1H-NMR data. NMR chemical shifts (δ) are reported in ppm and coupling constants (*J*) are reported in Hz. High resolution mass spectral (HRMS) data were obtained by an Agilent 6545B LC/Q-TOF (Agilent Technologies, Santa Clara, CA, USA).

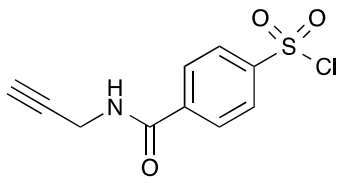
Chemical Suppliers:

The chemicals below were purchased from:

Fisher Scientific: 1H-1,2,4-Triazole, *N,N*-Diisopropylethylamine, Propargylamine, *N,N*-Dimethylformamide dimethyl acetal

Combi-Blocks: 4-(chlorosulfonyl)benzoyl chloride

2.1 Probe Synthesis



4-(Prop-2-yn-1-ylcarbamoyl)benzenesulfonyl chloride, S1

To a solution of 4-(chlorosulfonyl)benzoic acid (1.8 g, 8.2 mmol, 1.0 eq.) in DCM (41 mL) was added 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI) (1.7 g, 8.98 mmol, 1.1 eq.), and propargylamine (624 μL, 8.2 mmol, 1.0 eq.) at 0 °C and the reaction mixture was stirred for 1 h. The reaction was quenched with 1 M aqueous HCl, diluted with DCM, and the organic layer was separated. The aqueous layer was extracted with DCM. The organic layers were combined, washed with saturated aqueous $NaHCO_3$, dried over $MgSO_4$ and concentrated *in vacuo*. The crude product was purified by silica gel column chromatography (hexane:ethyl acetate:DCM = 7:2:1 to 7:3:1, v/v/v) to give **S1** (1.11 g, 53%). **^{1}H NMR** (600 MHz, $CDCl_3$) 8.13 – 8.11 (m, 2H), 8.03 – 8.01 (m, 2H), 6.51 (s, 1H), 4.28 (dd, *J* = 5.2, 2.6 Hz, 2H), 2.32 (t, *J* = 2.6 Hz, 1H). **^{13}C NMR** (150 MHz, $CDCl_3$) δ 165.02, 146.73, 139.98, 128.62, 127.55, 78.76, 72.72, 30.30. ESI-TOF (HRMS) m/z $[M+H]^+$ calculated for $C_{10}H_9ClNO_3S$ 257.9986, found 257.9990

General Protocol to synthesize HHS-465 and HHS-475 (Figure S1)

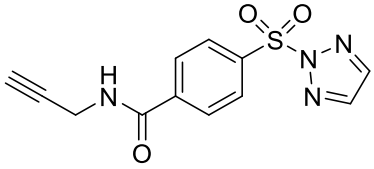
To a solution of compound **S1** (0.1 g, 0.39 mmol, 1.0 eq.) in anhydrous DCM (1.9 mL, 0.2 M) was added the corresponding triazole (1.94 mmol, 5.0 eq.) and *N,N*-diisopropylethylamine (DIPEA) (124 μL, 0.78 mmol, 2.0 eq.) at 0 °C. Then the reaction mixture was stirred at room temperature for overnight. The crude product was directly loaded and purified using silica

gel flash column chromatography (acetone/DCM = 5:100 to 10:100) to afford **HHS-465** and **HHS-475**, respectively.

General Protocol to synthesize HHS-481, HHS-482, and HHS-483

To a solution of compound **S1** (0.1 g, 0.39 mmol, 1.0 eq.) in anhydrous DCM (1.9 mL, 0.2 M) was added the corresponding triazole (0.47 mmol, 1.2 eq.) and DIPEA (124 μL, 0.47 mmol, 1.2 eq.) at 0 °C. Then the reaction mixture was stirred at room temperature for overnight. The crude product was directly loaded and purified using silica gel flash column chromatography (acetone/DCM = 3:100 to 7:100) to afford **HHS-481, HHS-482**, and **HHS-483**, respectively.

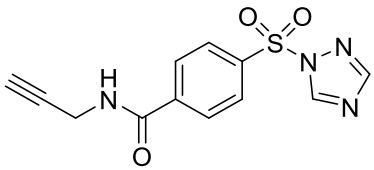
4-((2*H*-1,2,3-Triazol-2-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide, HHS-465



HHS-465

Yield: 35%, **^{1}H NMR** (600 MHz, $(CD_3)_2CO$) δ 8.40 (s, 1H, NH), 8.21 – 8.06 (m, 6H), 4.20 (dd, *J* = 5.5, 2.6 Hz, 2H), 2.69 (d, *J* = 2.3 Hz, 1H). **^{13}C NMR** (150 MHz, $(CD_3)_2CO$) δ 165.34, 141.50, 140.48, 139.06, 129.61, 129.49, 80.80, 72.35, 29.78 ESI-TOF (HRMS) m/z $[M+H]^+$ calculated for $C_{12}H_{11}N_4O_3S$ 291.0546, found 291.0546

4-((1*H*-1,2,4-Triazol-1-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide, HHS-475

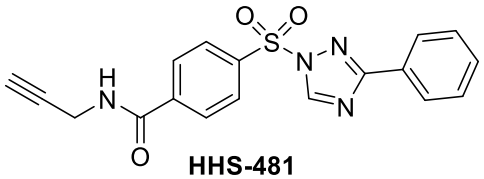


HHS-475

Yield: 66%, **^{1}H NMR** (600 MHz, $(CD_3)_2CO$) δ 9.15 (s, 1H), 8.46 (br, *J* = 5.7 Hz, 1H), 8.25 – 8.22 (m, 2H), 8.21 (s, 1H), 8.20 – 8.17 (m, 2H), 4.22 (dd, *J* = 5.5, 2.6 Hz, 2H), 2.70 (t, *J* = 2.6 Hz, 1H). **^{13}C NMR** (150 MHz, $(CD_3)_2CO$) δ 165.35, 155.46, 146.70, 141.55, 139.00, 129.62, 129.57,

80.72, 72.37, 29.75. ESI-TOF (HRMS) m/z $[M+H]^+$ calculated for $C_{12}H_{11}N_4O_3S$ 291.0546, found 291.0546

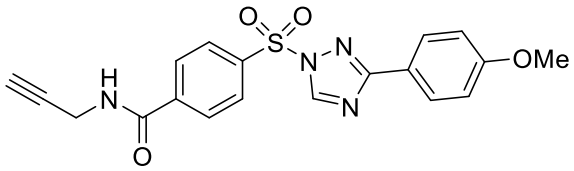
4-((3-Phenyl-1*H*-1,2,4-triazol-1-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide, HHS-481



HHS-481

Yield: 72%, **^{1}H NMR** (600 MHz, $(CD_3)_2CO$) δ 9.18 (s, 1H), 8.38 (br, 1H, NH), 8.31 – 8.28 (m, 2H), 8.21 – 8.16 (m, 2H), 8.10 – 8.04 (m, 2H), 7.53 – 7.44 (m, 3H), 4.20 (dd, *J* = 5.6, 2.6 Hz, 2H), 2.69 (t, *J* = 2.6 Hz, 1H). **^{13}C NMR** (150 MHz, $(CD_3)_2CO$) δ 165.79, 165.33, 147.76, 141.61, 139.20, 131.45, 130.15, 129.67, 129.64, 129.62, 127.61, 80.76, 72.34, 29.72. ESI-TOF (HRMS) m/z $[M+Na]^+$ calculated for $C_{18}H_{14}N_4NaO_3S$ 389.0679, found 389.0681

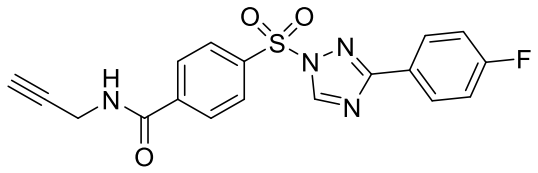
4-((3-(4-Methoxyphenyl)-1*H*-1,2,4-triazol-1-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide, HHS-482



HHS-482

Yield: 82%, **^{1}H NMR** (600 MHz, $(CD_3)_2CO$) δ 9.12 (s, 1H), 8.38 (br, 1H, NH), 8.30 – 8.25 (m, 2H), 8.19 – 8.16 (m, 2H), 8.03 – 7.98 (m, 2H), 7.06 – 7.00 (m, 2H), 4.20 (dd, *J* = 5.6, 2.5 Hz, 2H), 3.85 (s, 3H), 2.69 (t, *J* = 2.6 Hz, 1H). **^{13}C NMR** (150 MHz, $(CD_3)_2CO$) δ 165.79, 165.35, 162.68, 147.65, 141.60, 139.39, 129.67, 129.58, 129.26, 122.63, 115.05, 80.80, 72.33, 55.77, 29.75. ESI-TOF (HRMS) m/z $[M+Na]^+$ calculated for $C_{19}H_{16}N_4NaO_4S$ 419.0784, found 419.0788

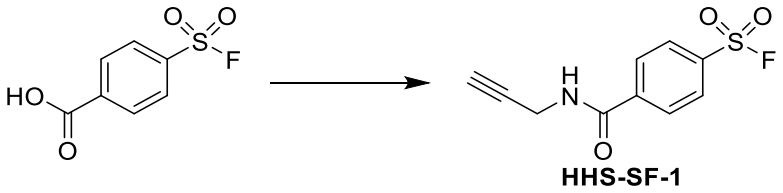
4-((3-(4-Fluorophenyl)-1*H*-1,2,4-triazol-1-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide, HHS-483



HHS-483

Yield: 77%, **^{1}H NMR** (600 MHz, $(CD_3)_2CO$) δ 9.19 (s, 1H), 8.40 (s, 1H), 8.31 – 8.26 (m, 2H), 8.21 – 8.17 (m, 2H), 8.14 – 8.09 (m, 2H), 7.29 – 7.23 (m, 2H), 4.20 (dd, *J* = 5.5, 2.6 Hz, 2H), 2.69 (t, *J* = 2.6 Hz, 1H). **^{13}C NMR** (150 MHz, $(CD_3)_2CO$) δ 165.84, 165.33, 164.92, 164.19, 147.85, 141.62, 139.13, 129.95 (d, *J* = 8.7 Hz), 129.66 (d, *J* = 5.7 Hz), 126.63 (d, *J* = 3.0 Hz), 116.63 (d, *J* = 22.1 Hz), 80.75, 72.36, 29.75. **^{19}F NMR** (564 MHz, $(CD_3)_2CO$) δ -111.48. ESI-TOF (HRMS) m/z $[M+H]^+$ calculated for $C_{18}H_{14}FN_4O_3S$ 385.0765, found 385.0755

4-(Prop-2-yn-1-ylcarbamoyl)benzenesulfonyl fluoride, HHS-SF-1

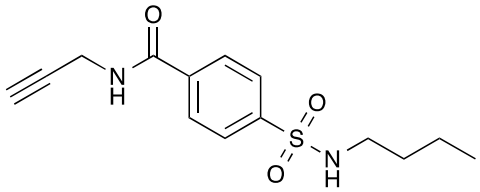


HHS-SF-1

To a solution of 4-(Fluorosulfonyl)benzoic acid (0.2 g, 0.98 mmol, 1.0 eq.) in DCM was added Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI) (206 mg, 1.08 mmol, 1.1 eq.), and propargylamine (68 μL, 0.98 mmol, 1.0 eq.) at 0 ˚C and the reaction mixture was stirred for 1 h. The reaction was quenched with 1 M aqueous HCl, diluted with DCM, and the organic layer was separated. The aqueous layer was extracted with DCM. The organic layers were combined, washed with saturated aqueous $NaHCO_3$, dried over $MgSO_4$ and concentrated *in vacuo*. The crude product was purified by silica gel column chromatography (hexane–ethyl acetate–DCM = 7:2:1 to 7:3:1, v/v/v) to give **HHS-SF-1** (145.5 mg, 62%). **^{1}H NMR** (600 MHz, $CDCl_3$) δ 8.08 (d, J = 8.5 Hz, 2H), 8.03 (d, J = 8.2 Hz, 2H), 6.69 (s, 1H), 4.27 (dd, J = 5.3, 2.5 Hz, 2H), 2.31 (t, J = 2.6 Hz, 1H). **^{13}C NMR** (150 MHz, $CDCl_3$) δ 165.06, 140.40, 135.84 (d, *J* = 25.5 Hz), 128.96, 128.54, 78.77, 72.61, 30.24. **^{19}F NMR** (564 MHz, $(CDCl_3)$ δ +65.91. ESI-TOF (HRMS) m/z $[M+H]^+$ calculated for $C_{10}H_9FNO_3S$ 242.0282, found 242.0281

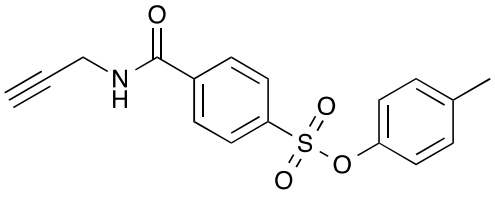
Synthesis of *n*-butylamine-probe and *p*-cresol-probe HPLC standards

4-(N-Butylsulfamoyl)-N-(prop-2-yn-1-yl)benzamide, KY-2-42



To a solution of compound **S1** (0.099 g, 0.38 mmol, 1.0 eq) in anhydrous DCM (3.8 mL, 100 mM) was added the n-butylamine (41.5 µL, 0.43 mmol, 1.1 eq) and DIPEA (73.3µL, 0.42 mmol, 1.1 eq). The reaction was stirred at room temperature for 1.5 hours. The reaction mixture was poured into 1 M HCl aqueous solution (3.8 mL) and extracted with DCM (3.8 mL) 3 times. The organic phase was combined and washed with saturated $NaHCO_3$ (10.0 mL) and brine (10.0 mL). The solution was dried over $MgSO_4$ and evaporated under reduced pressure. The product was purified using silica gel flash column chromatography (ethyl acetate/hexane =1:2 to 1:1) to afford KY-2-42, a white powder. Yield: 86%, **^{1}H NMR** (600 MHz, $(CD_3)_2CO$) δ 8.37 – 8.29 (m, 1H), 8.12 – 8.06 (m, 2H), 7.96 – 7.91 (m, 2H), 6.55 (t, *J* = 6.0 Hz, 1H), 4.22 (dd, *J* = 5.6, 2.5 Hz, 2H), 2.93 (td, *J* = 7.1, 6.1 Hz, 2H), 2.70 (t, *J* = 2.5 Hz, 1H), 1.49 – 1.41 (m, 2H), 1.34 – 1.25 (m, 2H), 0.83 (t, *J* = 7.4 Hz, 3H). **^{13}C NMR** (151 MHz, $(CD_3)_2CO$) δ 166.04, 144.63, 138.52, 128.94, 127.87, 81.11, 72.25, 43.66, 32.42, 29.71, 20.36, 13.87. ESI-TOF (HRMS) m/z $[M+Na]^+$ calculated for $C_{14}H_{18}N_2NaO_3S$ 317.0930 found 317.0930.

***p*-Tolyl 4-(prop-2-yn-1-ylcarbamoyl)benzenesulfonate, KY-2-48**



To a solution of compound **S1** (0.041 g, 0.14 mmol, 1.0 eq.) in anhydrous DCM (1.4 mL, 100 mM) was added the *p*-cresol (15.7 µL, 0.15 mmol, 1.1 eq.), DIPEA (19.1 µL, 0.15 mmol, 1.1 eq) and DMAP (3.4 mg, 0.03 mmol, 0.2 eq). The reaction was stirred at room temperature for 1 hour. The reaction mixture was poured into 1 M HCl aqueous solution (1.4 mL) and extracted with DCM (1.4 mL) 3 times. The organic phase was combined and washed with saturated $NaHCO_3$ (1.4 mL) and brine (1.4 mL). The solution was dried over $MgSO_4$ and evaporated under reduced pressure. The purification was carried out using silica gel flash column chromatography

(ethyl acetate/hexane =1:3 to 1:1) to afford KY-2-48, a colorless oil. Yield: 70%, **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.92 – 7 .88 (m, 2H), 7.88 – 7.84 (m, 2H), 7.07 – 7.03 (m, 2H), 6.84 – 6.80 (m, 2H), 6.55 (t, *J* = 5.3 Hz, 1H), 4.25 (dd, *J* = 5.2, 2.5 Hz, 2H), 2.31 – 2.26 (m, 4H). **^{13}C NMR** (151 MHz, $CDCl_3$) δ 165.31, 147.19, 138.88, 138.23, 137.39, 130.26, 128.87, 127.83, 121.84, 78.79, 72.35, 30.02, 20.84. ESI-TOF (HRMS) m/z $[M+Na]^+$ calculated for $C_{17}H_{15}NNaO_4S$ 352.0614 found 352.0620

3. HPLC METHOD FOR SOLUTION REACTIVITY AND STABILITY PROFILING

The following reagents were prepared and stored at 0 °C prior to use: 0.1 M solution of caffeine in acetonitrile, 1.0 M solution of n-butylamine, *p*-cresol, TMG, acetic acid in acetonitrile, and 10 mM solution of the probes in a mixture of DMF-ACN (v/v=10:90) are made.

3.1 *p*-Cresol reactivity against a probe mixture: A solution of *p*-cresol (16.5 µmol, 3.3 eq) was mixed with 1.1, 2.2, or 3.3 eq of TMG. To initiate the reaction, the *p*-cresol/TMG solution was added to a sulfonyl probe mixture of HHS-475/HHS-482/HHS-SF-1 (500 µL, 5 µmol, 1.0 eq each) and the reaction was kept at 0 °C. The reaction progress was monitored by taking out a 50.0 µL portion of the reaction mixture at various time points followed by addition of a 10 µL quenching solution of acetic acid (0.5 M final, 5.0 µmol) and the internal caffeine standard (0.05 M final, 0.5 µmol). Sample (1.0 µL) was injected and analyzed by reverse-phase HPLC on a Shimadzu 1100 Series spectrometer with UV detection at 254 nm. Reaction progress was evaluated by monitoring consumption of sulfonyl probes because all probes generate a common *p*-cresol and *n*-butylamine product.

Chromatographic separation was performed using a Phenomenex Kinetex C18 column (2.6 µm, 50 mm x 4.6 mm). Mobile phases A and B were composed of H_2O (with 0.1% AcOH) and CH_3CN (with 0.1% AcOH), respectively. Using a constant flow rate of 0.8 mL/min, the mobile phase gradient program was as follows: 0-0.5 min, 15% B; 0.5-6.5 min 15-85% B (linear gradient); 6.5-7 min 85-100% B (linear gradient); 7- 8.5 min 100% B; 8.5-9 min 100-15% B (linear gradient); 9-9.8 min 15% B.

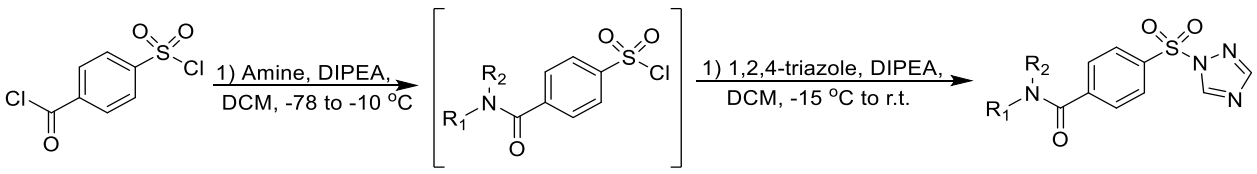
3.2 *n*-Butylamine reactivity against a probe mixture: Reactivity of sulfonyl probes against *n*-butylamine (3.3 eq) was performed as described above except the amount of TMG was fixed at 3.3 eq.

3.3 Probe reactivity against a *p*-cresol/*n*-butylamine mixture: A solution of *n*-butylamine (50.0 µL, 50.0 µmol, 5.0 eq), *p*-cresol (10.0 µL, 10.0 µmol, 1.0 eq), and TMG (5.0 µL, 5 µmol, 0.5 eq) were premixed. Probe reaction was initiated by addition of this solution to HHS-475, HHS-482, or HHS-SF-1 (10 µmol, 1.0 eq) at 0 °C. Reaction progress was monitored as described above. A

control experiment was also performed where equal amounts of *n*-butylamine (1.0 eq) and *p*-cresol (1.0 eq) were mixed.

3.4 Probe stability studies: Each probe was dissolved in DMSO or solution of DMF:ACN:PBS (4:6:1 (v/v)) at the following concentrations: 20 mM of HHS-475, 20 mM HHS-SF-1, and 10 mM of HHS-482 in a final volume of 50 µL. The internal caffeine standard (0.5 µmol) was spiked into each probe sample. Probe stability was monitored at room temperature by taking 1.0 µL of sample at various time points (0, 24, and 48 hours) and analyzing probe degradation by HPLC as described above.

SuTEx fragment synthesis

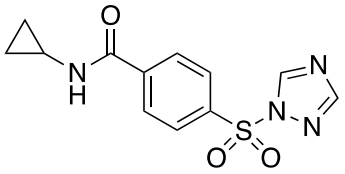


Supplementary Figure 1. Synthesis of Compounds

General protocol to synthesize **HHS-0101, 0201, 0301, 0401, 0601** and **-0701**

To a solution of 4-(chlorosulfonyl) benzoyl chloride (0.1 g, 0.42 mmol, 1.0 eq.) in anhydrous DCM (2.1 mL, 0.2 M) was added the corresponding amine (0.42 mmol, 1.0 eq.) and *N*,*N*-diisopropylethylamine (DIPEA) (80 uL, 0.92 mmol, 1.1 eq.) at -78 ºC. The reaction mixture was slowly warmed up to 0 °C, and then to the reaction mixture was added the 1,2,4-triazole (1.26 mmol, 3.0 eq.) and *N*,*N*-diisopropylethylamine (DIPEA) (80 uL, 0.40 mmol, 1.1 eq.) at -15 ºC. Then the reaction mixture was warmed up to room temperature, and stirred overnight. The crude product was directly loaded and purified using silica gel flash column chromatography (acetone/DCM = 2:100 to 10:100) to afford **HHS-0101, 0201, 0301, 0401, 0601** and **-0701** respectively.

4-((2*H*-1,2,3-triazol-2-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide HHS-0101

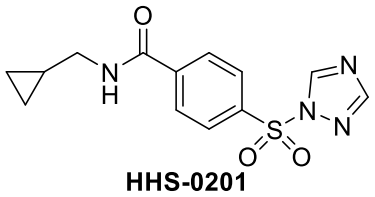


HHS-0101

Yield: 33 mg, 27%, **^{1}H NMR** (600 MHz, MeOH-d_6) δ 9.19 (s, 1H), 8.20 – 8.18 (m, 2H), 8.16 (s, 1H), 8.03 – 8.01 (m, 2H), 2.90 – 2.84 (m, 1H), 0.84 – 0.80 (m, 2H), 0.64 (dtd, *J* = 5.2, 4.6, 2.2

Hz, 2H). **^{13}C NMR** (150 MHz, MeOH-d_6) δ 167.99, 153.93, 145.73, 140.91, 138.01, 128.54, 128.37, 22.76, 5.07. ESI-TOF (HRMS) m/z $[M+H]^+$ calcd for $C_{12}H_{10}N_4O_3S^+$ 293.0703, found 293.0705.

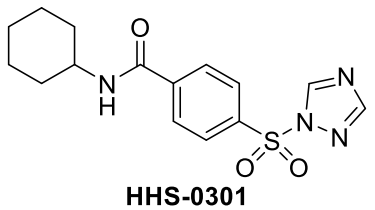
4-((2*H*-1,2,3-triazol-2-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide HHS-0201



HHS-0201

Yield: 56 mg, 44%, **^{1}H NMR** (600 MHz, Acetone-d_6) δ 9.13 (s, 1H), 8.22 – 8.20 (m, 2H), 8.19 (s, 1H), 8.16 – 8.14 (m, 2H), 3.27 (dd, *J* = 7.0, 5.1 Hz, 2H), 1.13 – 1.04 (m, 1H), 0.49 – 0.44 (m, 2H), 0.26 (dt, *J* = 6.3, 4.5 Hz, 2H). **^{13}C NMR** (150 MHz, Acetone-d_6) δ 165.49, 155.52, 146.72, 142.65, 138.72, 129.55, 129.54, 45.29, 11.53, 3.80. ESI-TOF (HRMS) m/z $[M+H]^+$ calcd for $C_{13}H_{15}N_4O_3S^+$ 307.0859, found 307.0859.

4-((2*H*-1,2,3-triazol-2-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide HHS-0301

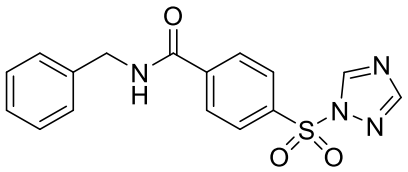


HHS-0301

Yield: 81 mg, 58%, **^{1}H NMR** (600 MHz, Acetone-d_6) δ 9.12 (s, 1H), 8.20 – 8.17 (m, 2H), 8.14 – 8.10 (m, 2H), 7.76 (s, 1H), 3.89 (dtt, *J* = 10.9, 7.6, 4.0 Hz, 1H), 2.00 – 1.92 (m, 2H), 1.80 – 1.74 (m, 2H), 1.69 – 1.61 (m, 1H), 1.44 – 1.29 (m, 4H), 1.25 – 1.13 (m, 1H). **^{13}C NMR** (150 MHz,

Acetone-d_6) δ 164.76, 155.52, 146.71, 142.98, 138.62, 129.58, 129.47, 50.05, 33.42, 26.33, 25.87. ESI-TOF (HRMS) m/z $[M+H]^+$ calcd for $C_{15}H_{19}N_4O_3S+$ 335.1172, found 335.1194.

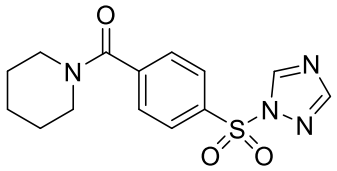
4-((2*H*-1,2,3-triazol-2-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide HHS-0401



HHS-0401

Yield: 101 mg, 71%, **^{1}H NMR** (600 MHz, Acetone-d_6) δ 9.13 (s, 1H), 8.51 (br, 1H), 8.23 – 8.21 (m, 2H), 8.21 – 8.20 (m, 2H), 8.19 (s, 1H), 7.39 – 7.36 (m, 2H), 7.34 – 7.29 (m, 2H), 7.28 – 7.22 (m, 1H), 4.62 (d, *J* = 6.0 Hz, 2H). **^{13}C NMR** (150 MHz, Acetone-d_6) δ 165.56, 155.54, 146.74, 142.31, 139.99, 138.93, 129.66, 129.61, 129.27, 128.51, 127.93, 44.26. ESI-TOF (HRMS) m/z $[M+H]^+$ calcd for $C_{16}H_{15}N_4O_3S^+$ 343.0859, found 343.0859.

4-((2*H*-1,2,3-triazol-2-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide HHS-0601

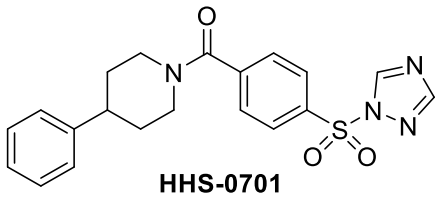


HHS-0601

Yield: 95 mg, 71%, **^{1}H NMR** (600 MHz, Acetone-d_6) δ 9.13 (d, *J* = 2.2 Hz, 1H), 8.19 (dd, *J* = 8.5, 2.2 Hz, 3H), 7.72 (dq, *J* = 8.4, 2.0 Hz, 2H), 3.65 (br, 2H), 3.28 (br, 2H), 1.68 (dd, *J* = 5.6, 2.4 Hz, 2H), 1.62 (br, 2H), 1.51 (br, 2H). **^{13}C NMR** (150 MHz, Acetone-d_6) δ 166.85, 154.57, 145.76,

144.31, 136.22, 128.82, 128.12, 48.06, 42.43, 29.30, 26.13, 25.30, 24.20. ESI-TOF (HRMS) m/z $[M+H]^+$ calcd for $C_{14}H_{17}N_4O_3S^+$ 321.1016, found 321.1015.

4-((2*H*-1,2,3-triazol-2-yl)sulfonyl)-*N*-(prop-2-yn-1-yl)benzamide HHS-0701

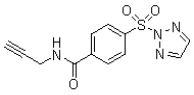


HHS-0701

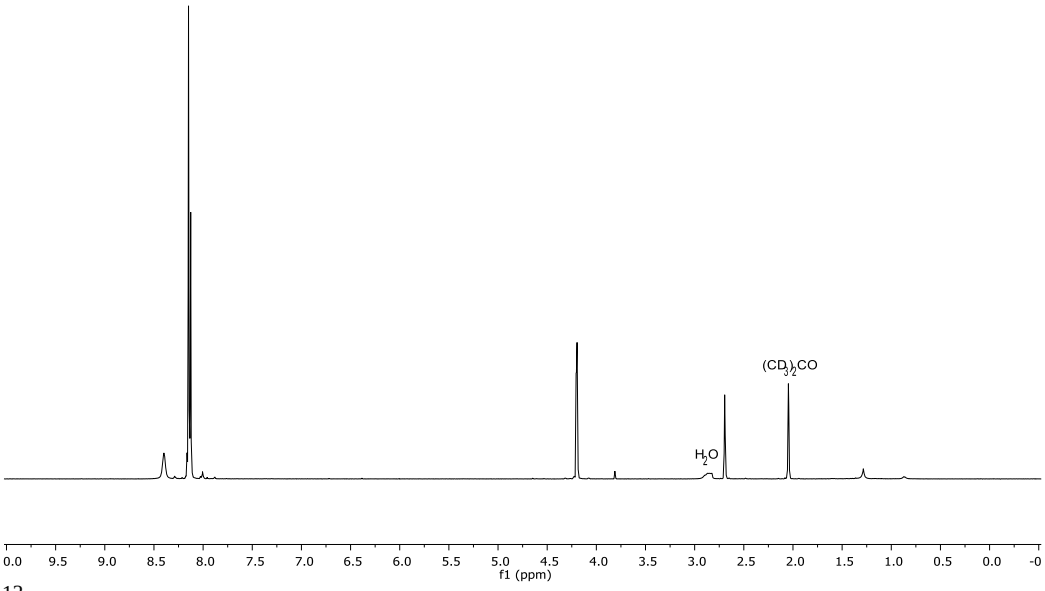
Yield: 121 mg, 73%, **^{1}H NMR** (600 MHz, $CDCl_3$) δ 8.76 (s, 1H), 8.18 – 8.12 (m, 2H), 8.05 (s, 1H), 7.69 – 7.62 (m, 2H), 7.32 (dd, *J* = 8.3, 6.9 Hz, 2H), 7.25 – 7.22 (m, 1H), 7.21 – 7.19 (m, 2H), 4.86 (d, *J* = 13.2 Hz, 1H), 3.66 (d, *J* = 13.4 Hz, 1H), 3.17 (t, *J* = 13.2 Hz, 1H), 2.95 – 2.85 (m, 1H), 2.02 (d, *J* = 13.5 Hz, 1H), 1.84 (d, *J* = 13.3 Hz, 1H), 1.78 (d, *J* = 13.6 Hz, 1H), 1.61 (s, 1H). **^{13}C NMR** (150 MHz, $CDCl_3$) δ 167.54, 154.51, 144.74, 144.53, 143.44, 136.51, 129.09, 128.67, 128.09, 126.73, 126.63, 48.23, 42.91, 42.51, 33.82, 32.72. ESI-TOF (HRMS) m/z $[M+H]^+$ calcd for $C_{20}H_{21}N_4O_3S^+$ 397.1329, found 397.1329.

^{1}H and ^{13}C NMR

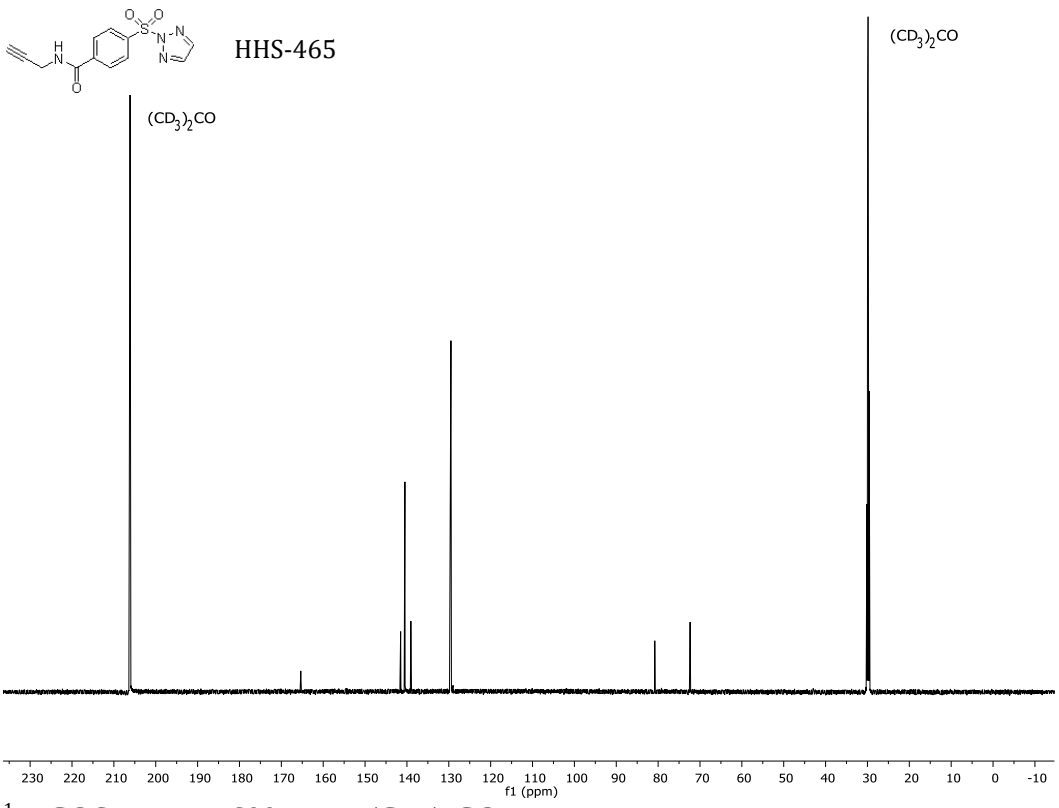
NMR Spectra



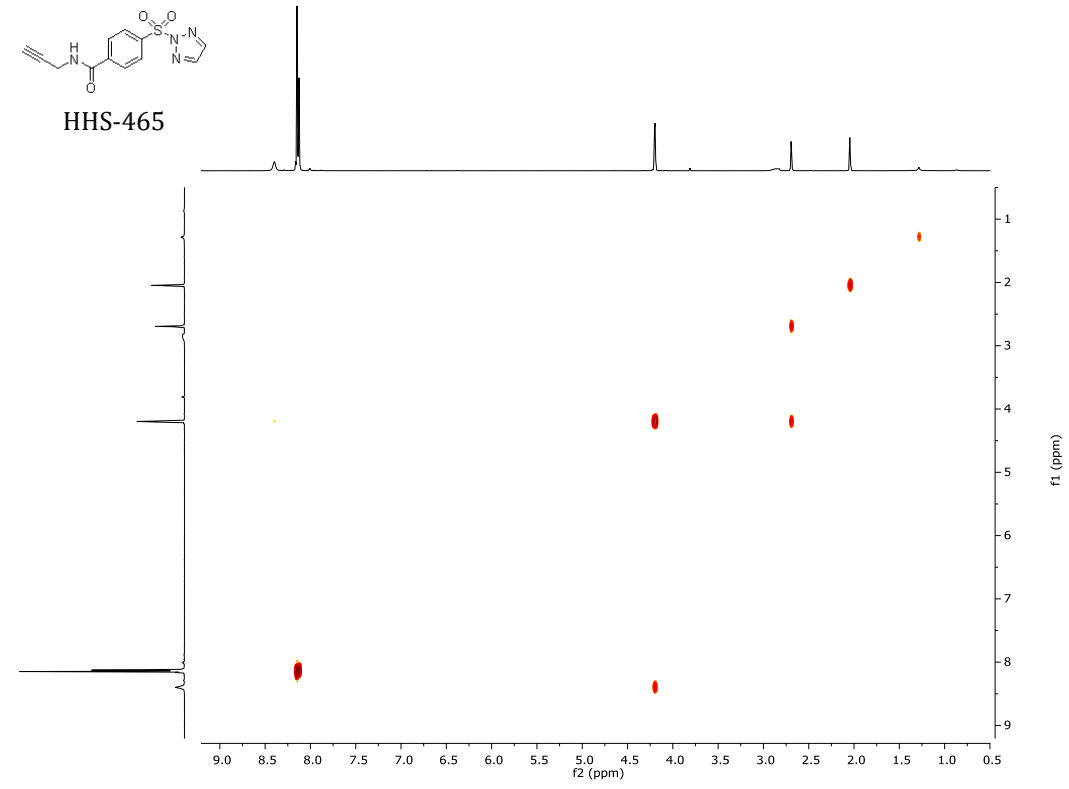
HHS-465



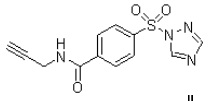
^{13}C NMR, 150 MHz, $(CD_3)_2CO$



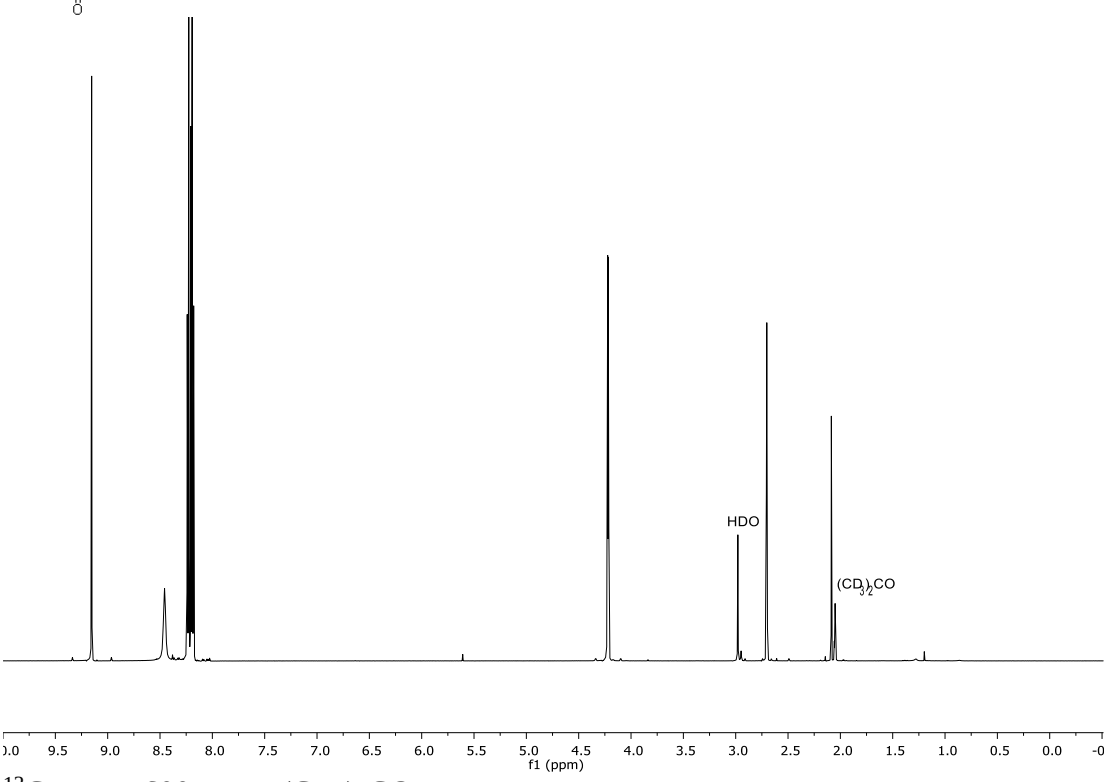
^{1}H-COSY NMR, 600 MHz, $(CD_3)_2CO$



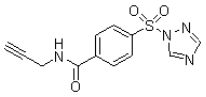
^{1}H NMR, 600 MHz, $(CD_3)_2CO$



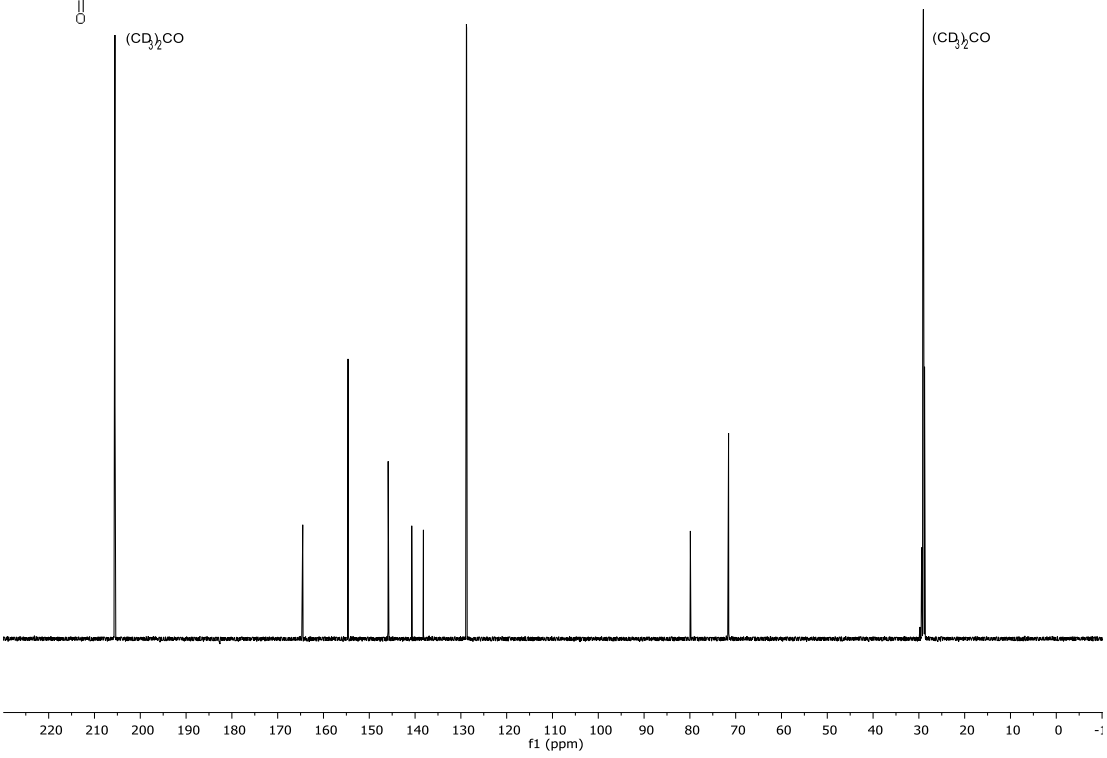
HHS-475



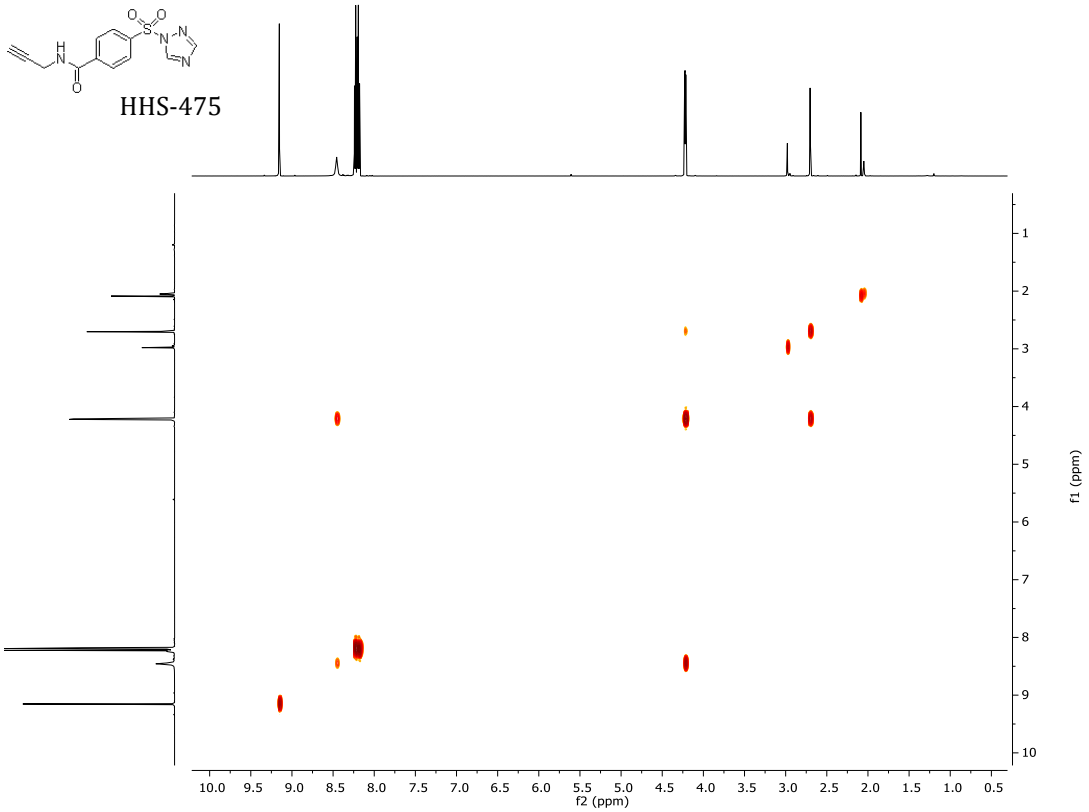
^{13}C NMR, 600 MHz, $(CD_3)_2CO$



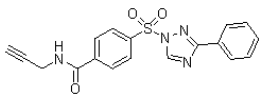
HHS-475



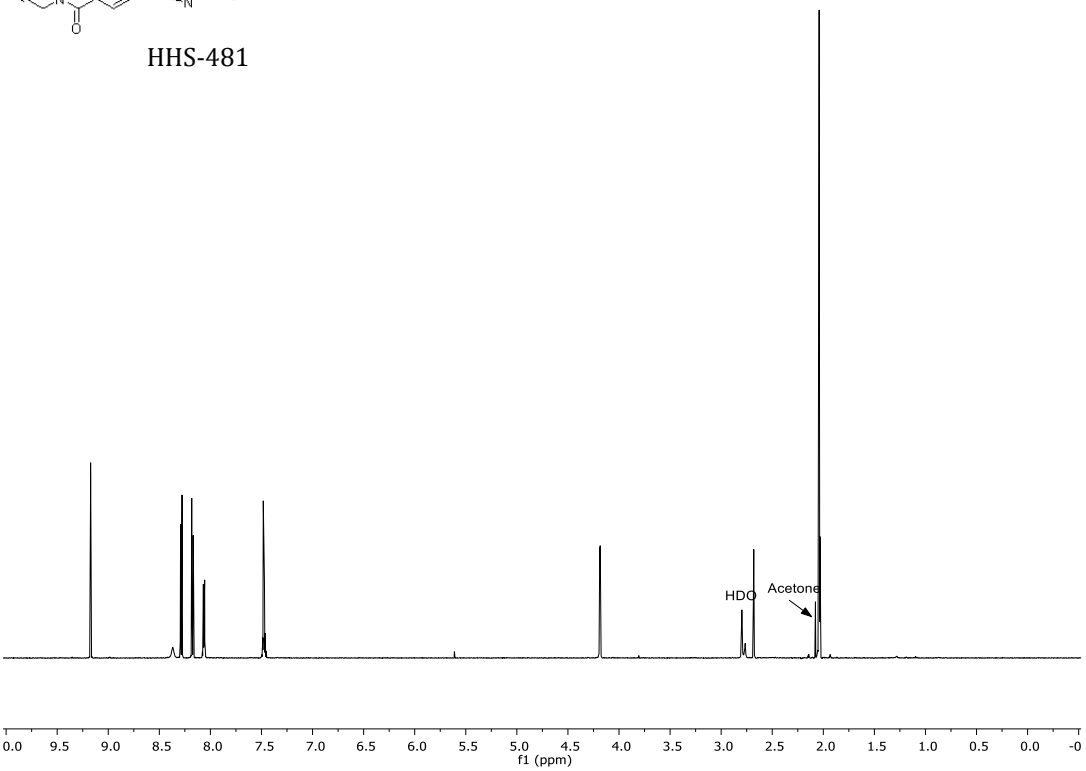
^{1}H-COSY NMR, 600 MHz, $(CD_3)_2CO$



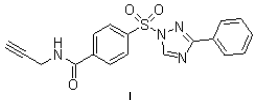
^{1}H NMR, 600 MHz, $(CD_3)_2CO$



HHS-481

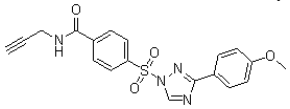


^{13}C NMR, 600 MHz, $(CD_3)_2CO$

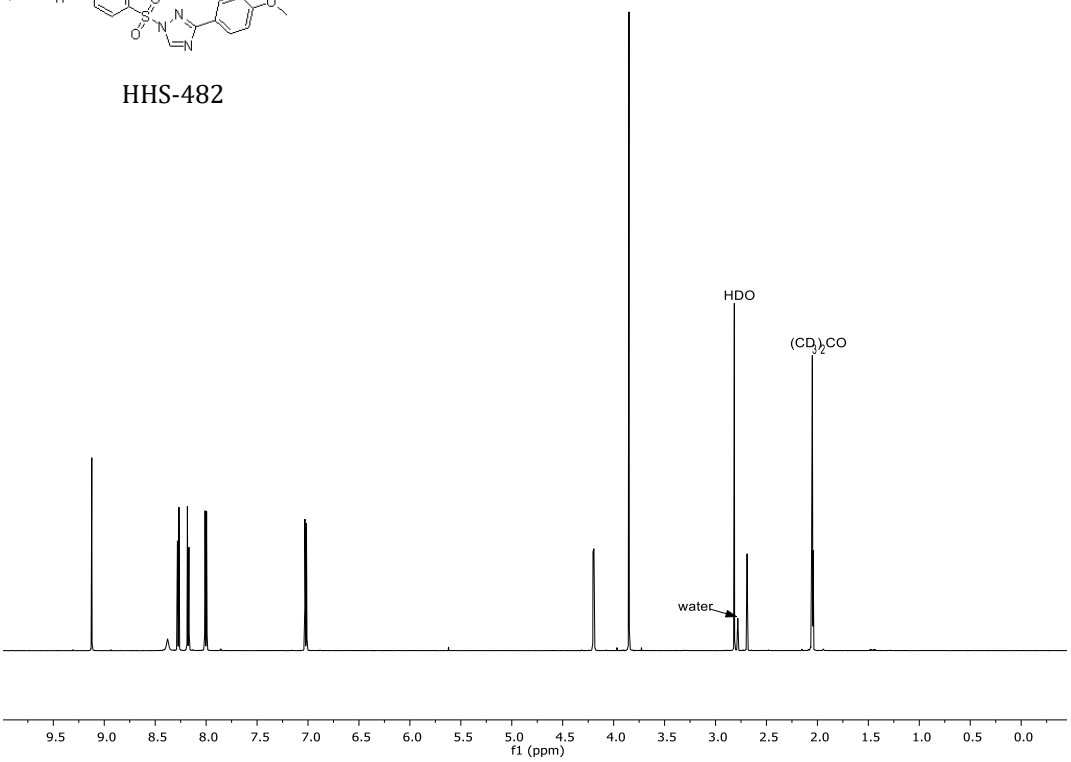


HHS-481

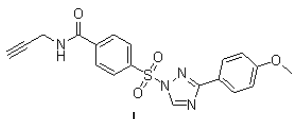
^{1}H NMR, 600 MHz, $(CD_3)_2CO$



HHS-482

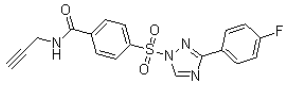


^{13}C NMR, 600 MHz, $(CD_3)_2CO$

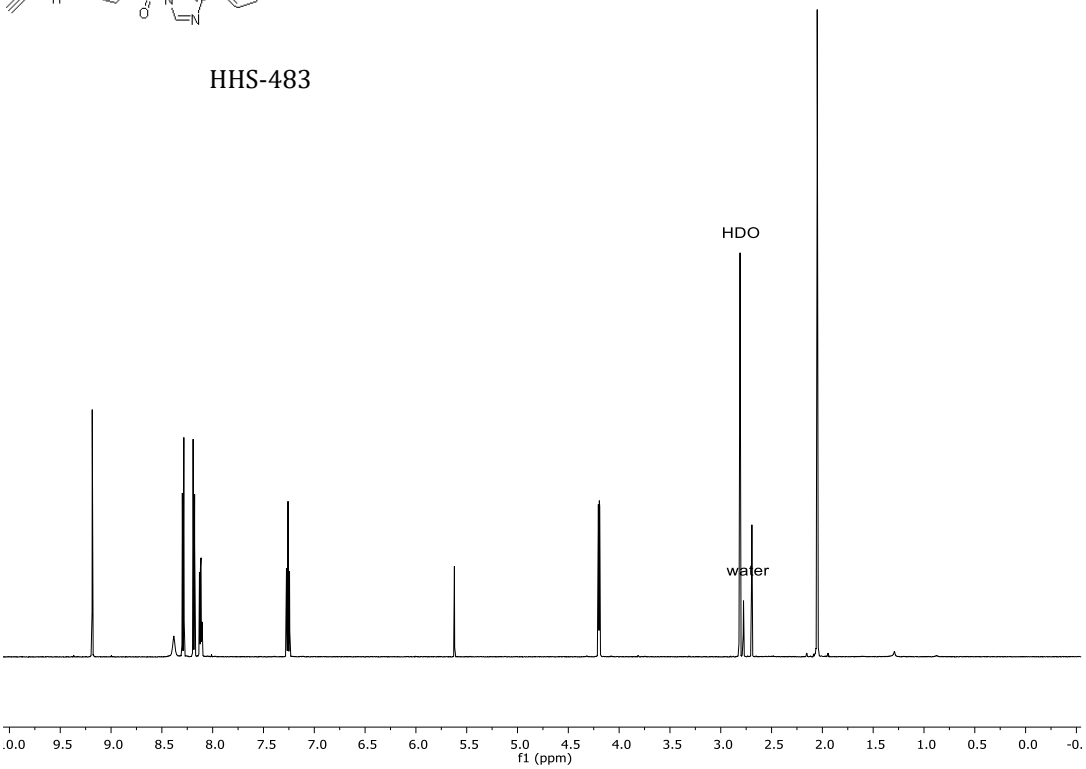


HHS-482

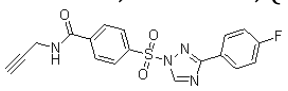
^{1}H NMR, 600 MHz, $(CD_3)_2CO$



HHS-483

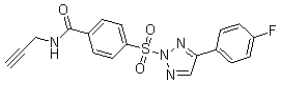


^{13}C NMR, 150 MHz, $(CD_3)_2CO$

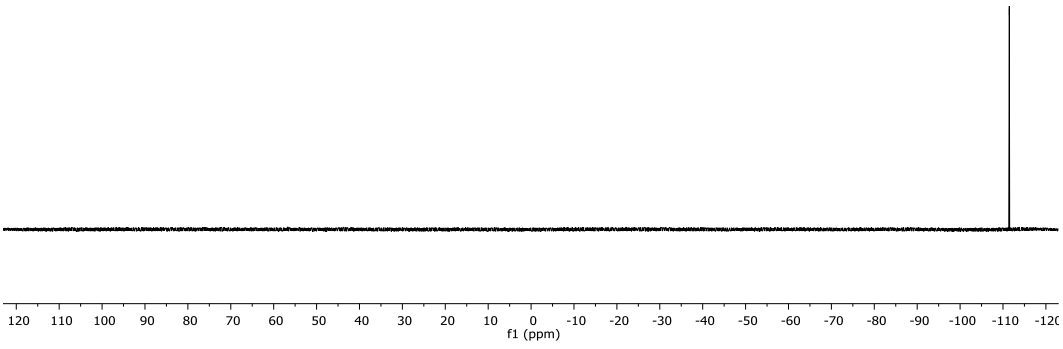


HHS-483

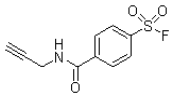
^{19}F NMR, 564 MHz, $(CD_3)_2CO$



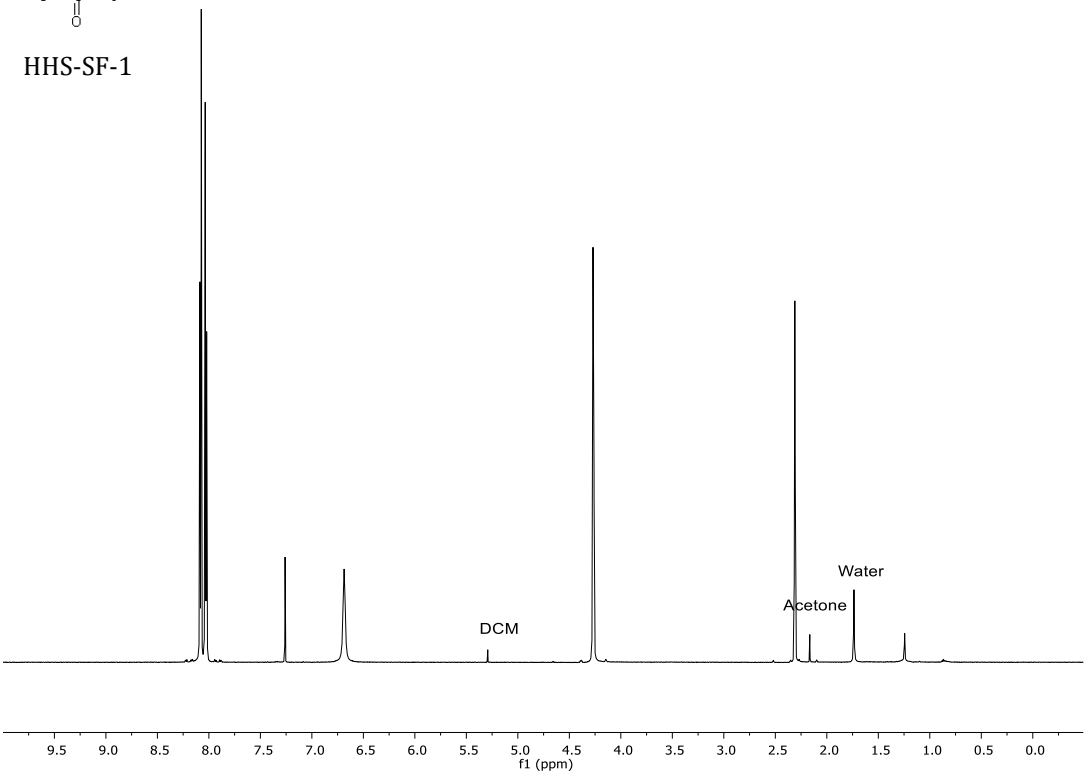
HHS-483



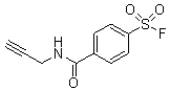
^{1}H NMR, 600 MHz, $(CD_3)_2CO$



HHS-SF-1

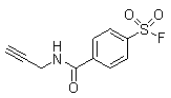


^{13}C NMR, 150 MHz, $(CD_3)_2CO$

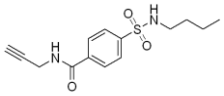


HHS-SF-1

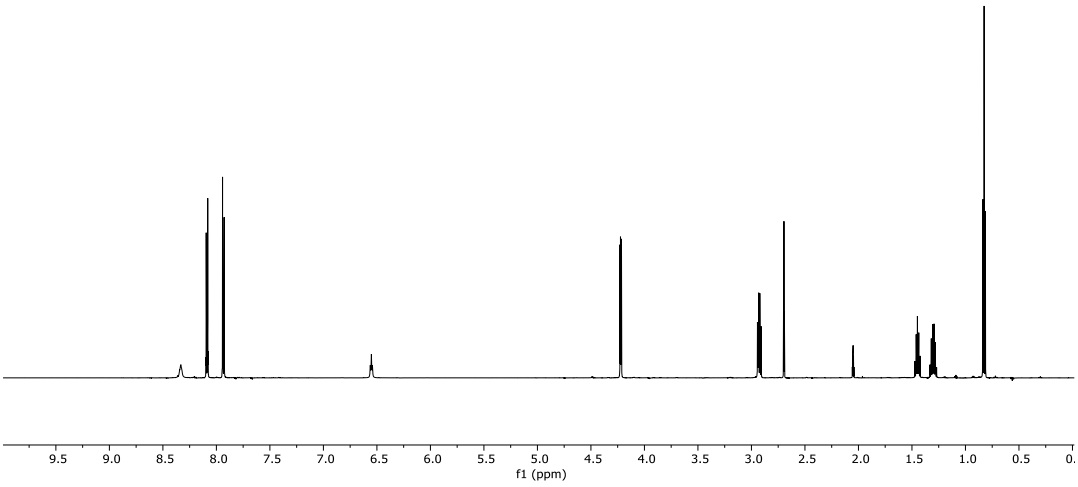
^{19}F NMR, 564 MHz, $CDCl_3$

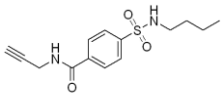


HHS-SF-1

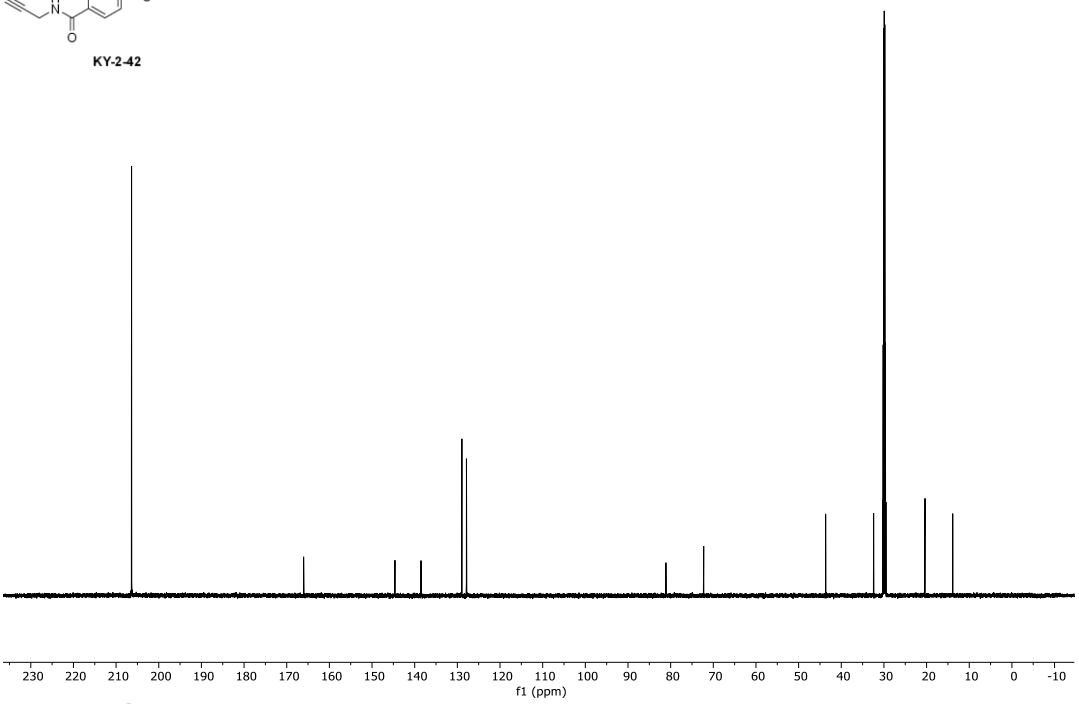


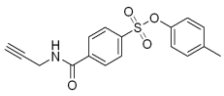
KY-2-42





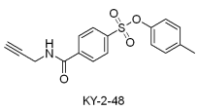
KY-2-42

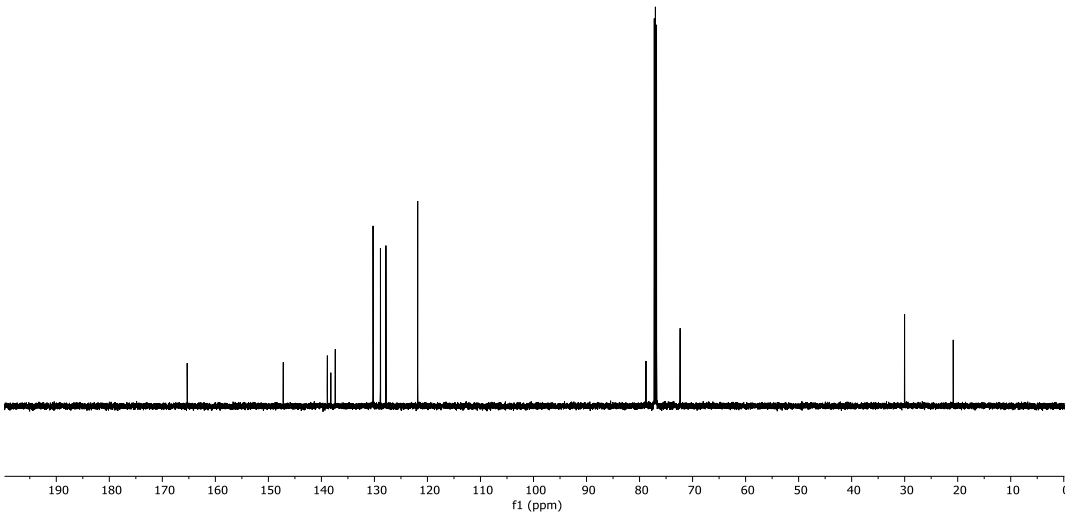

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)



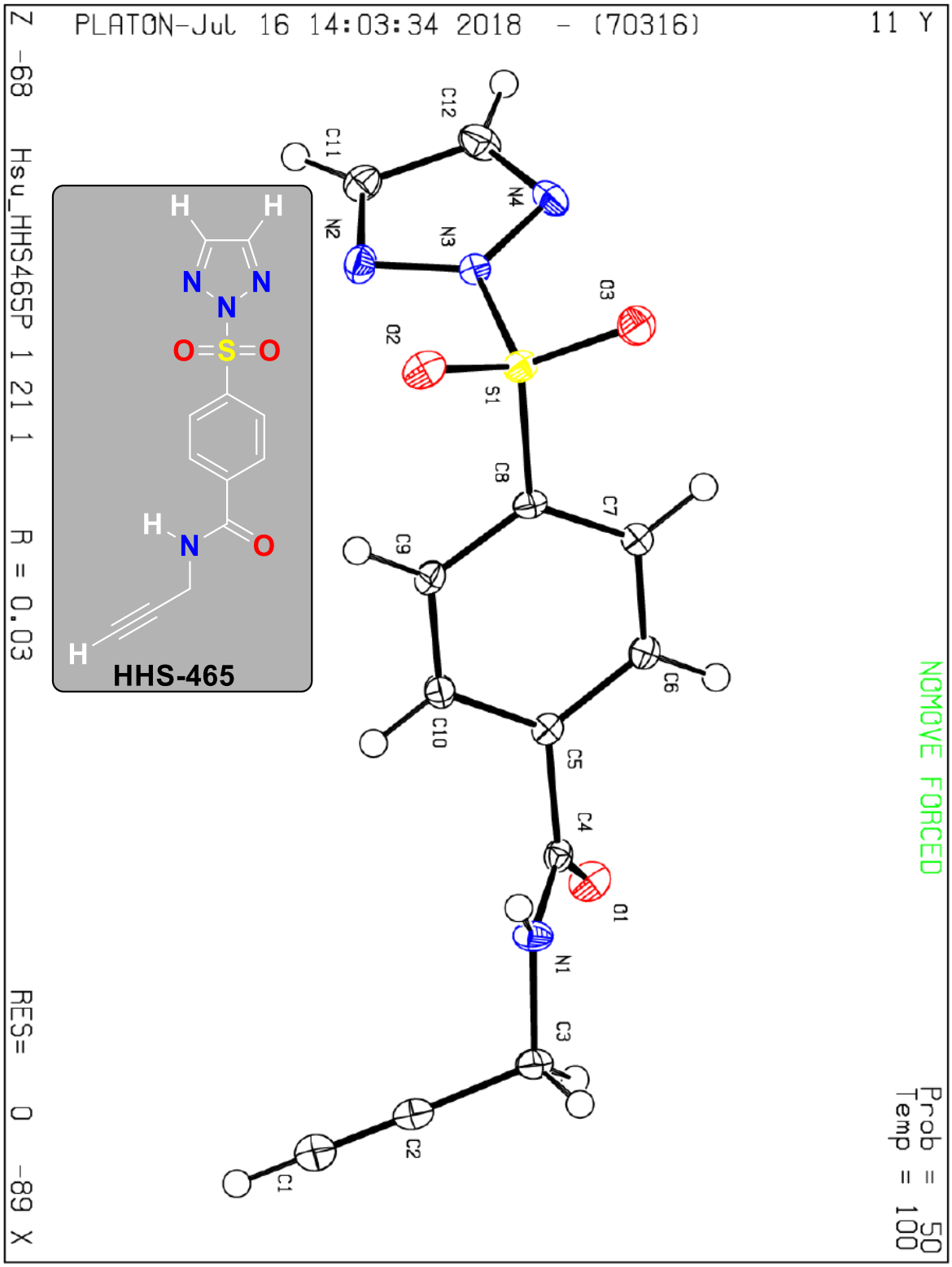
KY-2-48

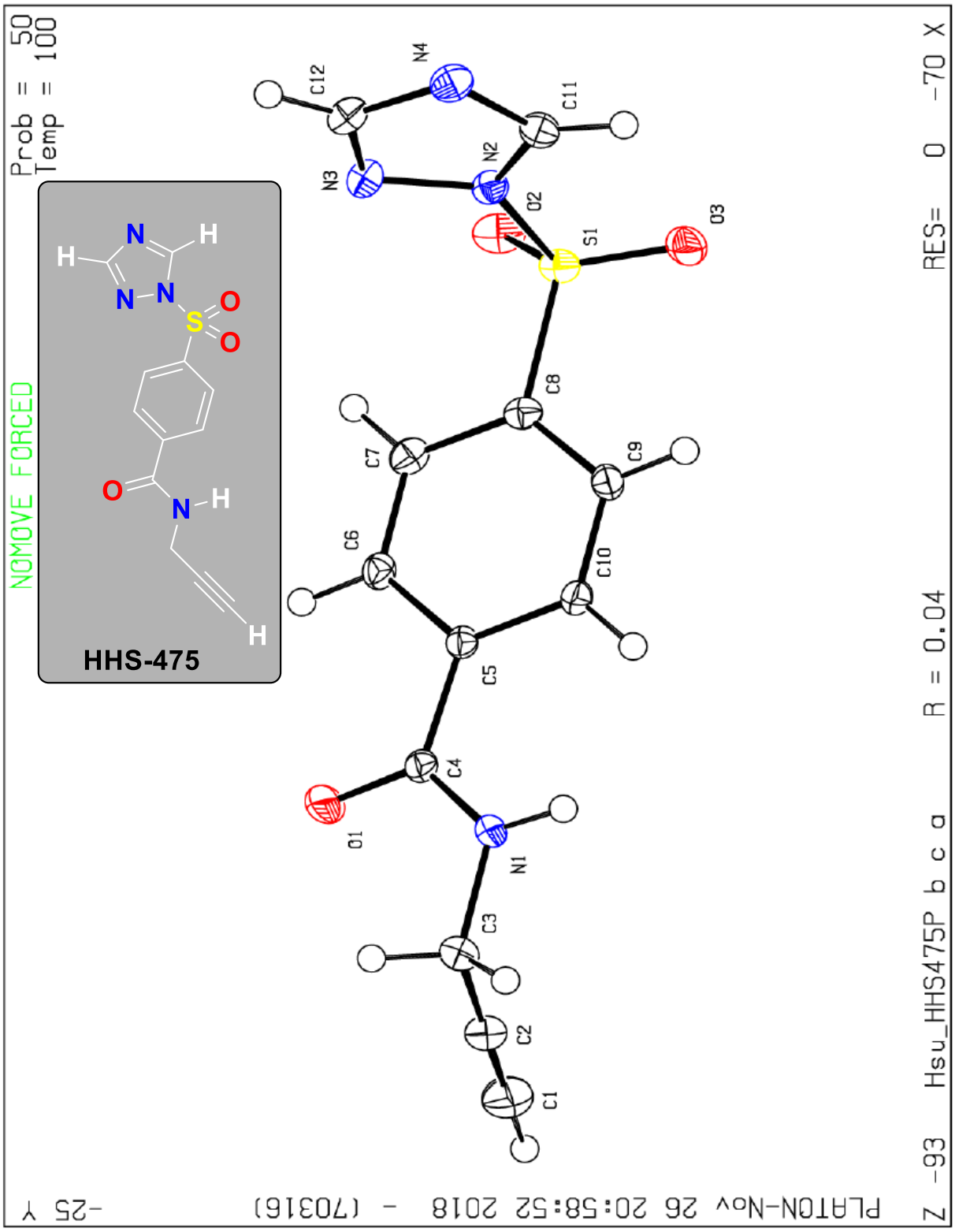
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

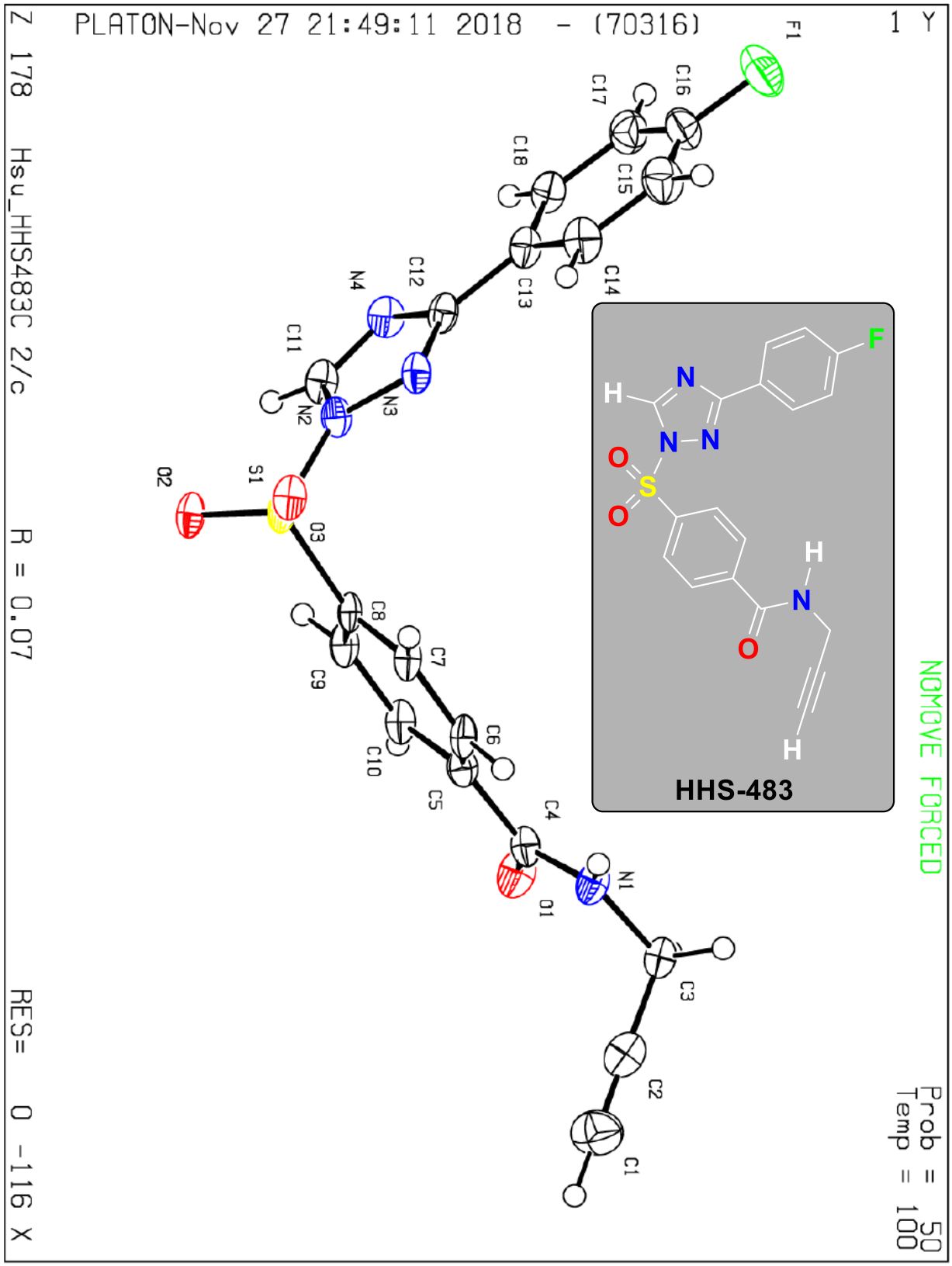




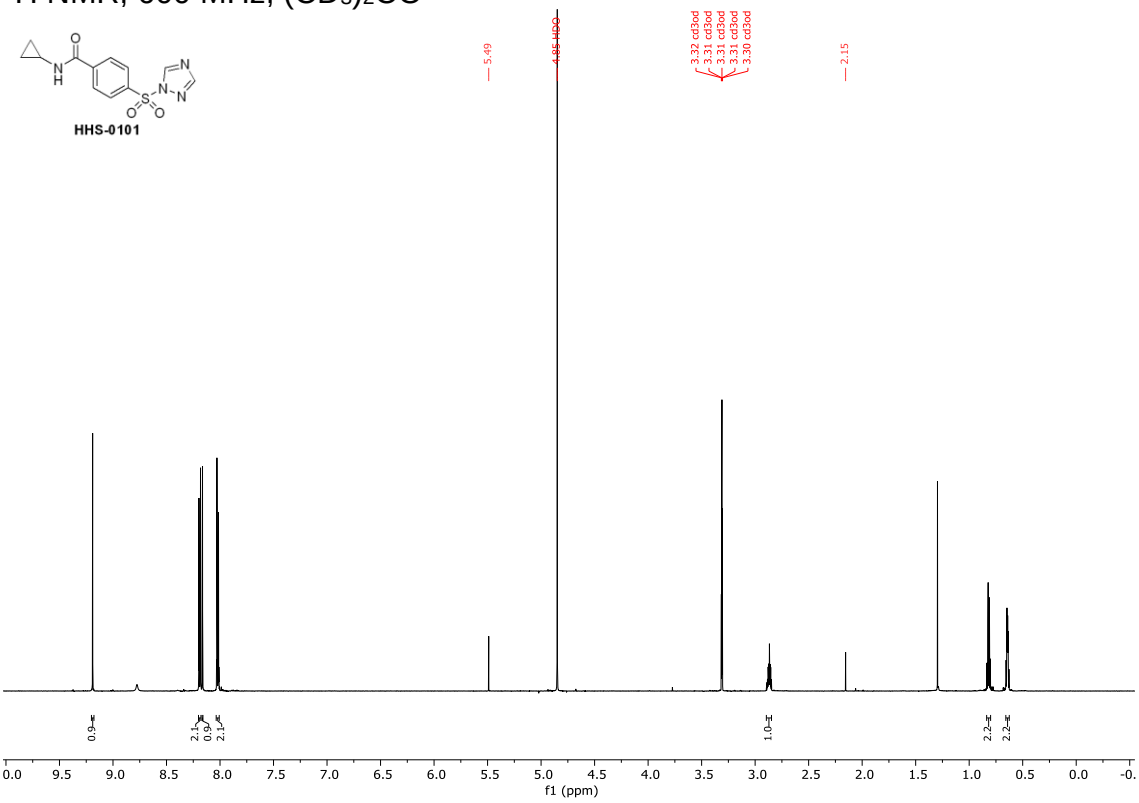
7.2 Crystal Structure of HHS-465


PLATON-Jul 16 14:03:34 2018 - (70316)
11 Y
Z -68 Hsu_HHS465P 1 21 1 R = 0.03
RES= 0 -89 X
NOMOVE FORCED
Prob = 50
Temp = 100
HHS-465
C12
C11
N4
N2
N3
O3
O2
S1
C8
C7
C9
C6
C10
C5
C4
O1
N1
C3
C2
C1

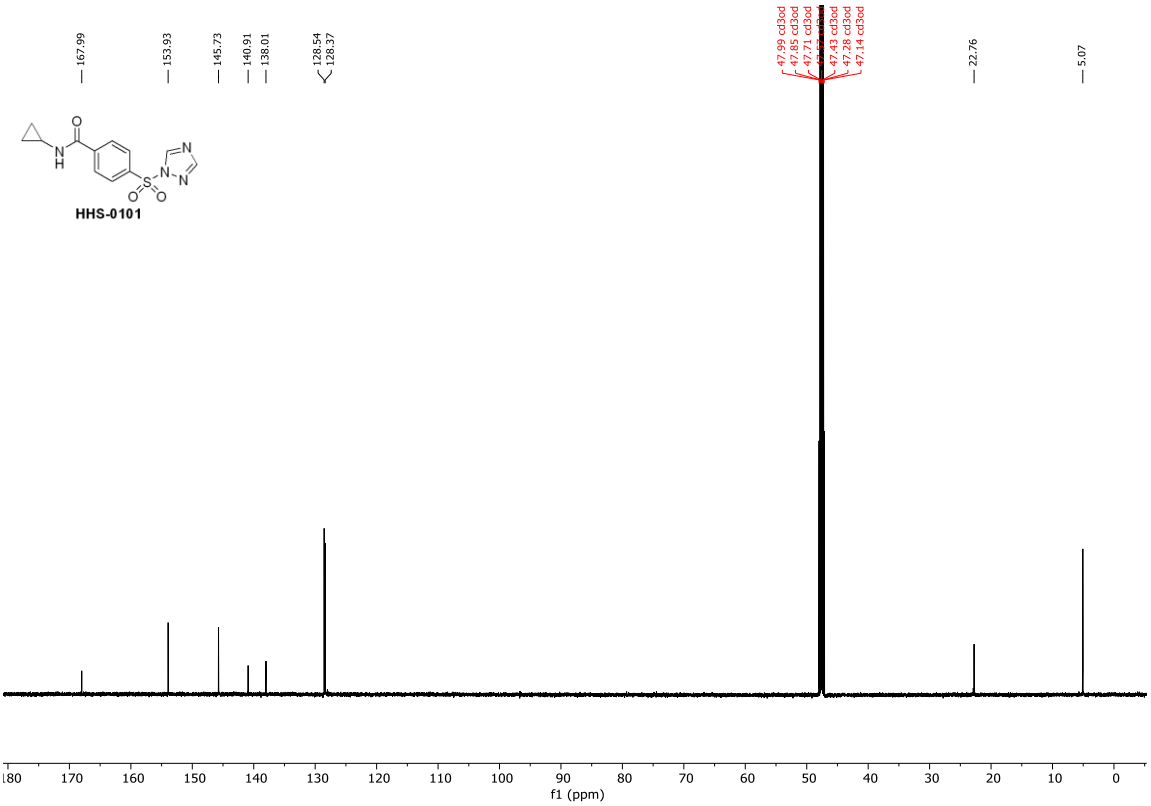

Prob = 50
Temp = 100
NOMOVE FORCED
HHS-475
PLATON-NoV 26 20:58:52 2018 - (70316)
Z -93 Hsu_HHS475P b c a R = 0.04
RES= 0 -70 X
-25 Y


PLATON-Nov 27 21:49:11 2018 - (70316)
1 Y
Z 178 Hsu_HHS483C 2/c
R = 0.07
RES= 0 -116 X
NOMOVE FORCED
Prob = 50
Temp = 100
HHS-483
F1
C16
C17
C18
C15
C14
C13
C12
N4
C11
N2
N3
S1
O2
O3
C8
C7
C9
C10
C6
C5
C4
O1
N1
C3
C2
C1
F
H
N
N
N
O
S
O
H
N
O
H

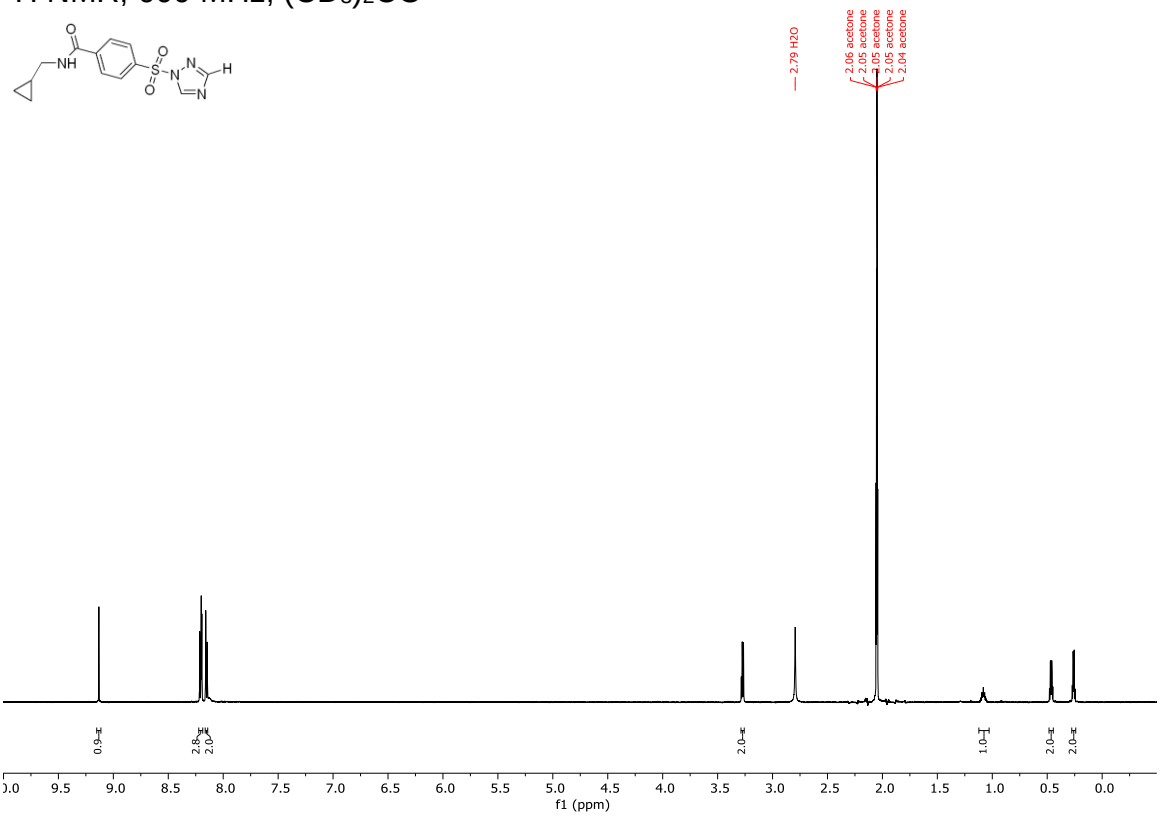
^{1}H NMR, 600 MHz, $(CD_3)_2CO$



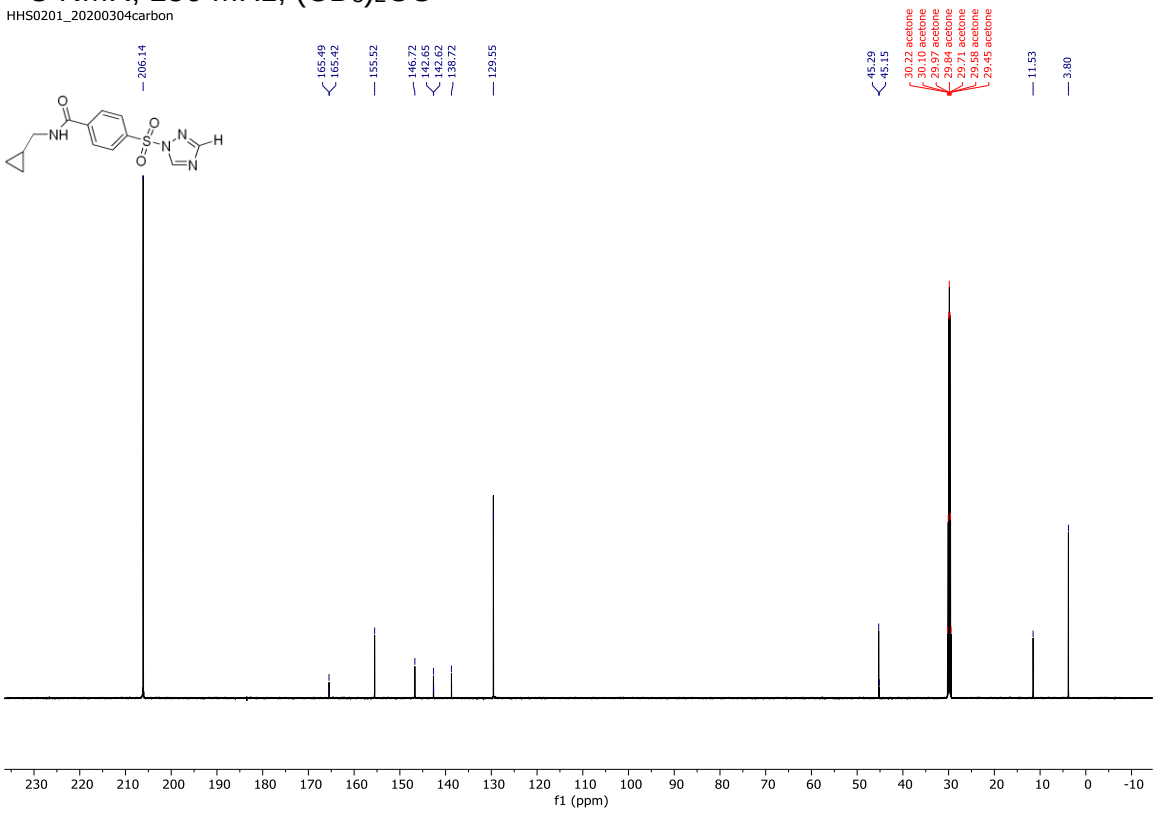
^{13}C NMR, 150 MHz, $(CD_3)_2CO$



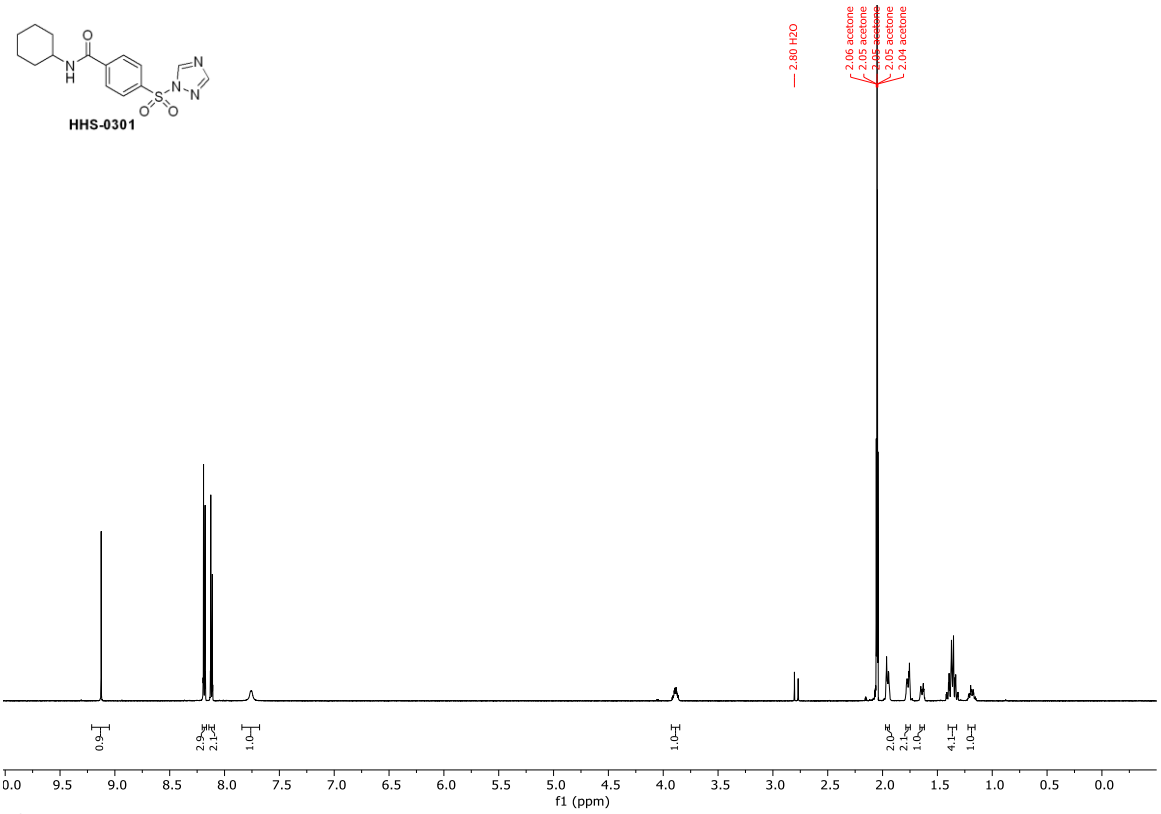
^{1}H NMR, 600 MHz, $(CD_3)_2CO$



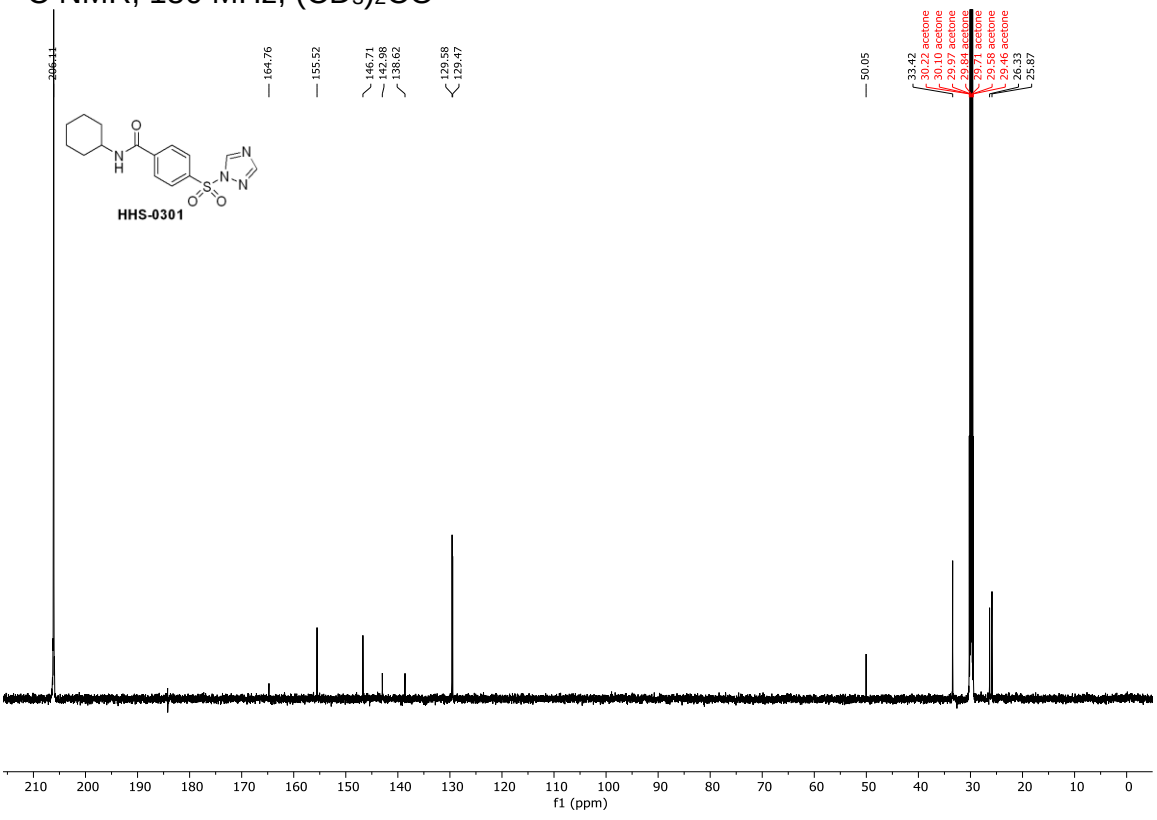
^{13}C NMR, 150 MHz, $(CD_3)_2CO$



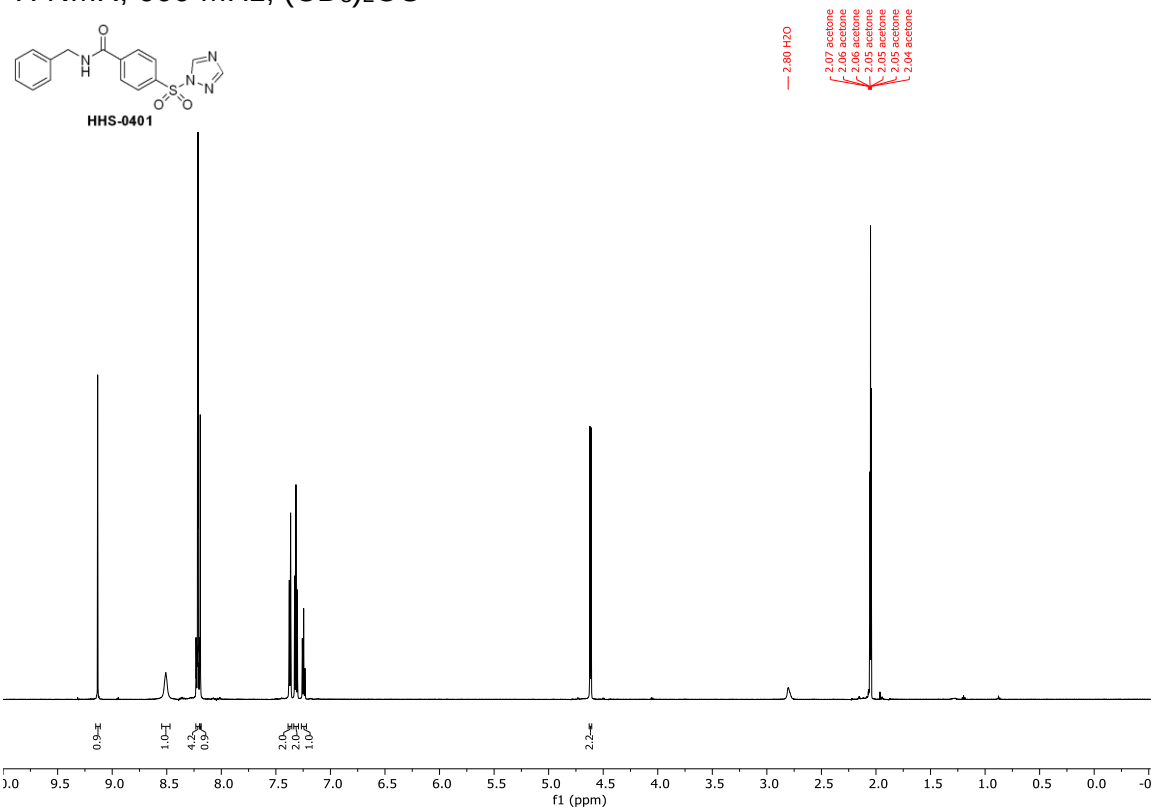
^{1}H NMR, 600 MHz, $(CD_3)_2CO$



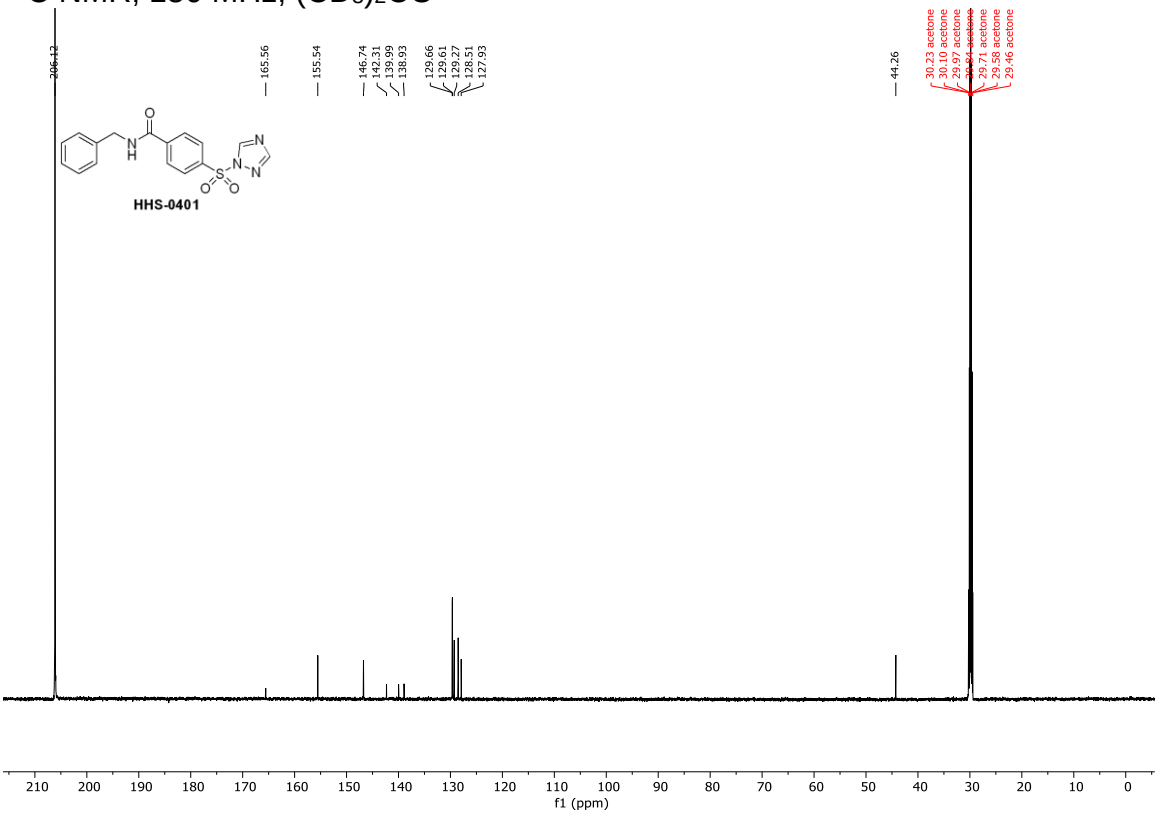
^{13}C NMR, 150 MHz, $(CD_3)_2CO$



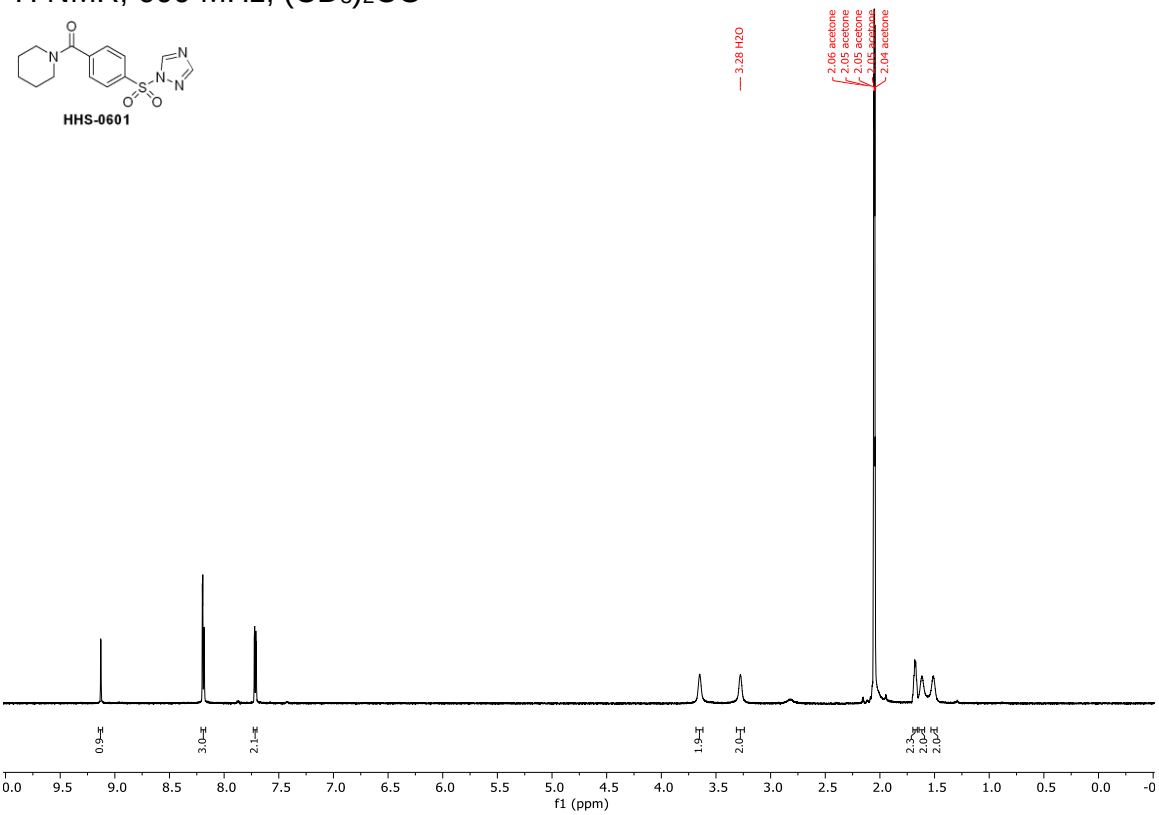
^{1}H NMR, 600 MHz, $(CD_3)_2CO$



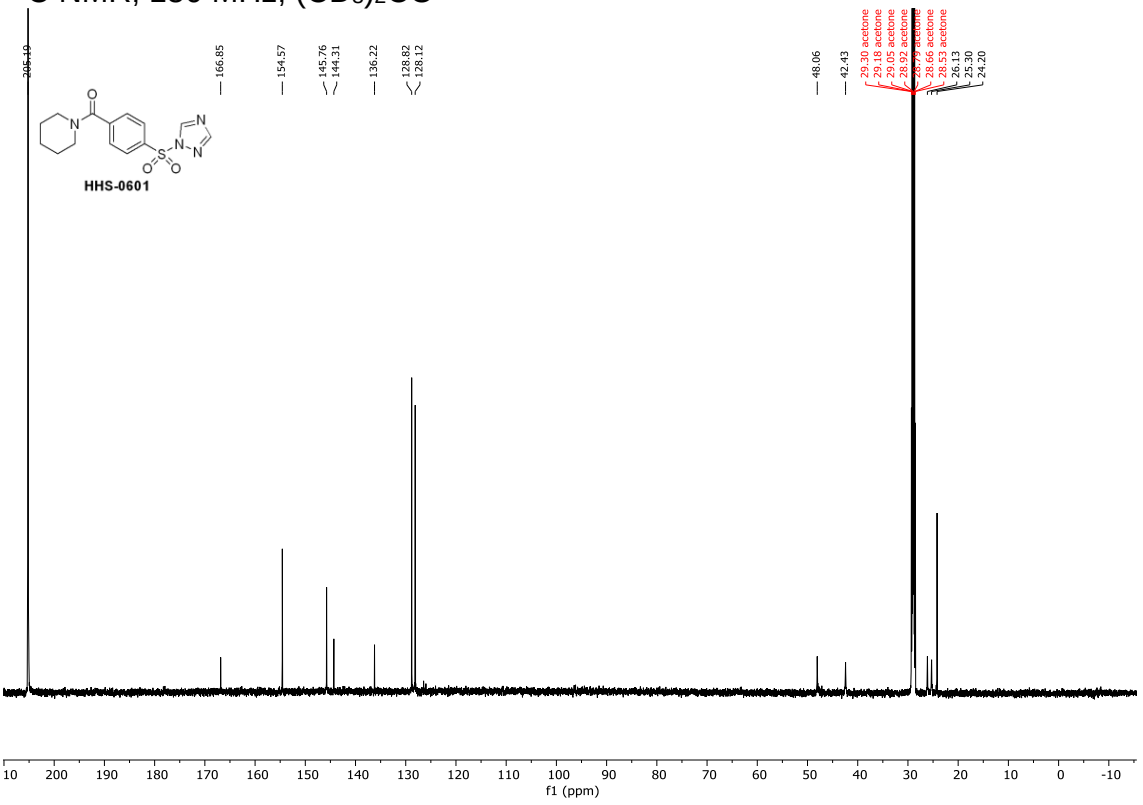
^{13}C NMR, 150 MHz, $(CD_3)_2CO$



^{1}H NMR, 600 MHz, $(CD_3)_2CO$

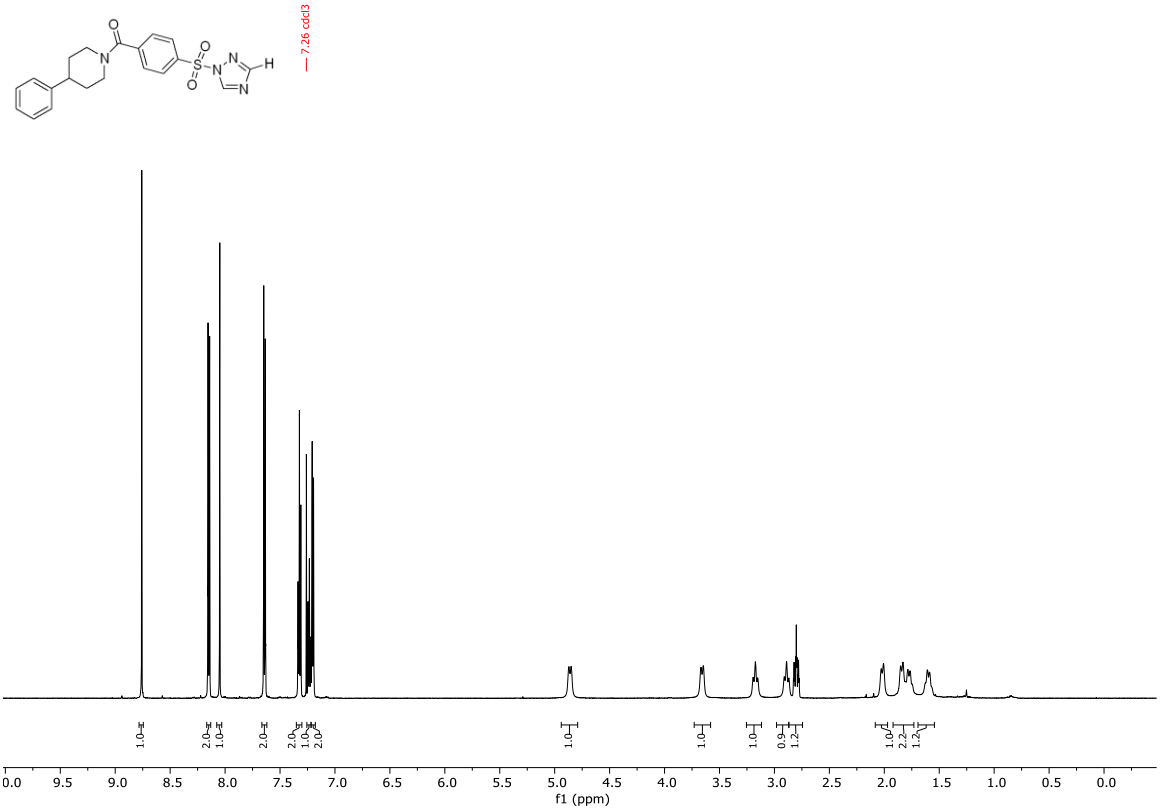


^{13}C NMR, 150 MHz, $(CD_3)_2CO$

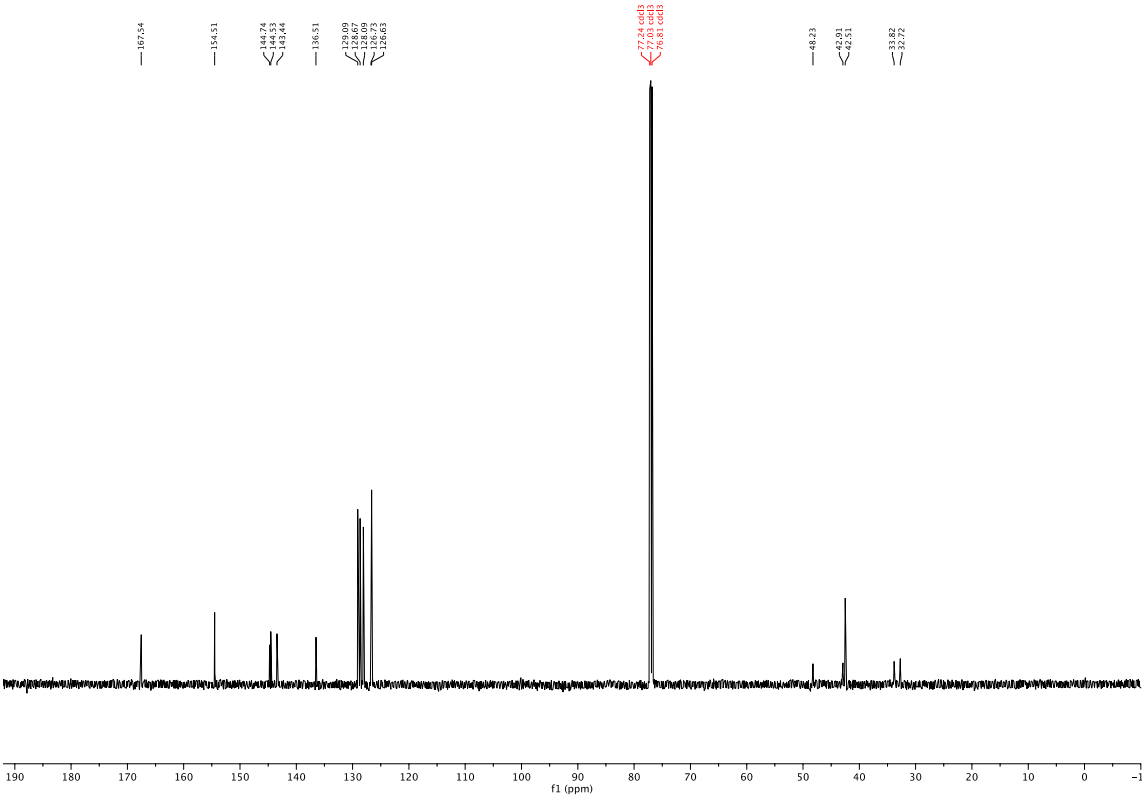


HHS-0701

^{1}H NMR, 600 MHz, $CDCl_3$



^{13}C NMR, 150 MHz, $CDCl_3$ HHS-0701

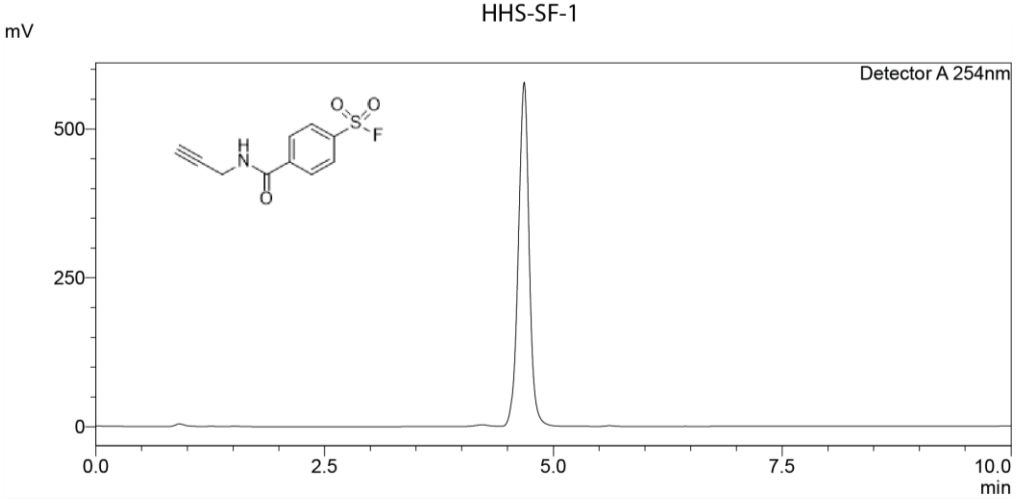


HPLC Analysis of Compound Purity

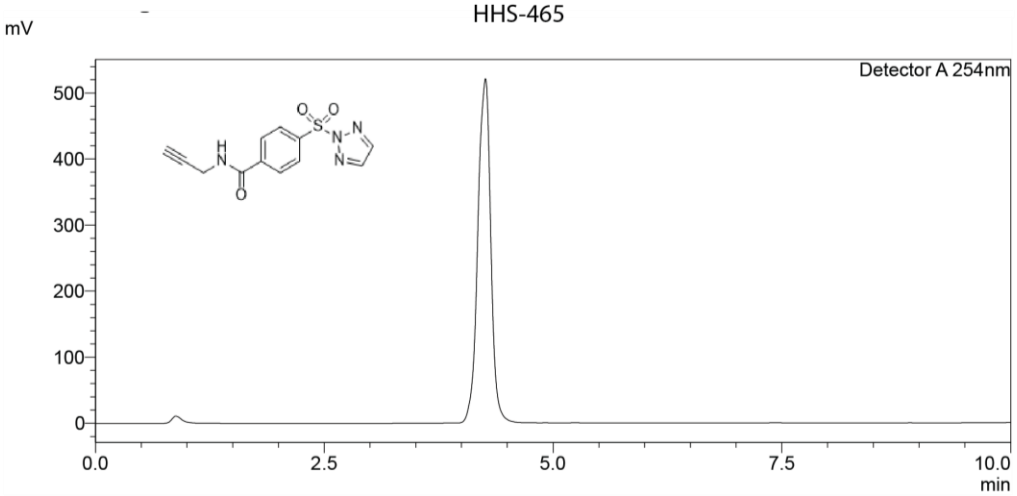
The purity of compounds was determined by HPLC on a Shimadzu Prominence series HPLC instrument with UV detection at 254 nm. Chromatographic separation was performed using a Phenomenex Kinetex C18 column (2.6 µm, 50 mm x 4.6 mm). Mobile phases A and B were composed of H_2O and CH_3CN, respectively. Using a constant flow rate of 0.4 mL/min, the mobile phase gradient was as follows: 0-1 min, 25% B; 1-6 min 25-100% B (linear gradient); 6-8 min 100% B; 8-9 min 100-25% B; 9-10 min 25% B. All final compounds were determined to be >95% pure by this method.

Representative spectra:

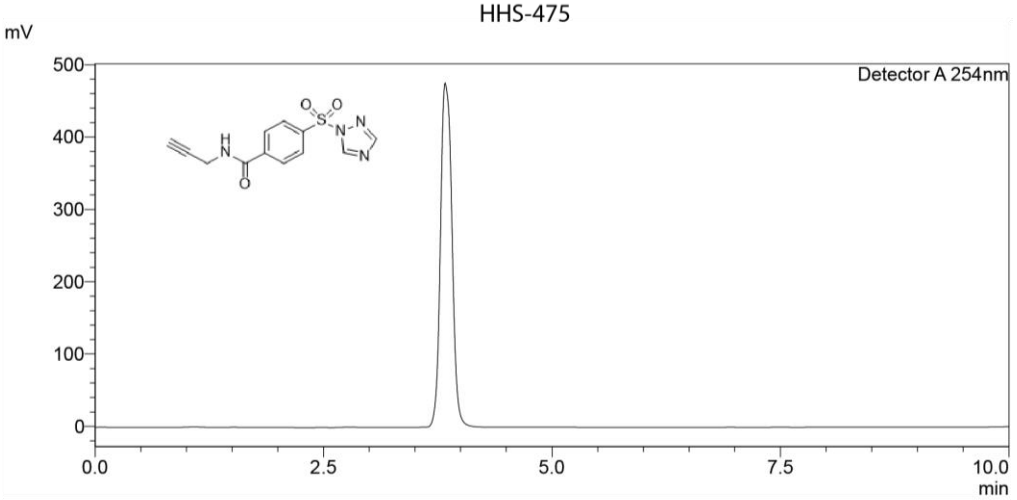
HHS-SF-1



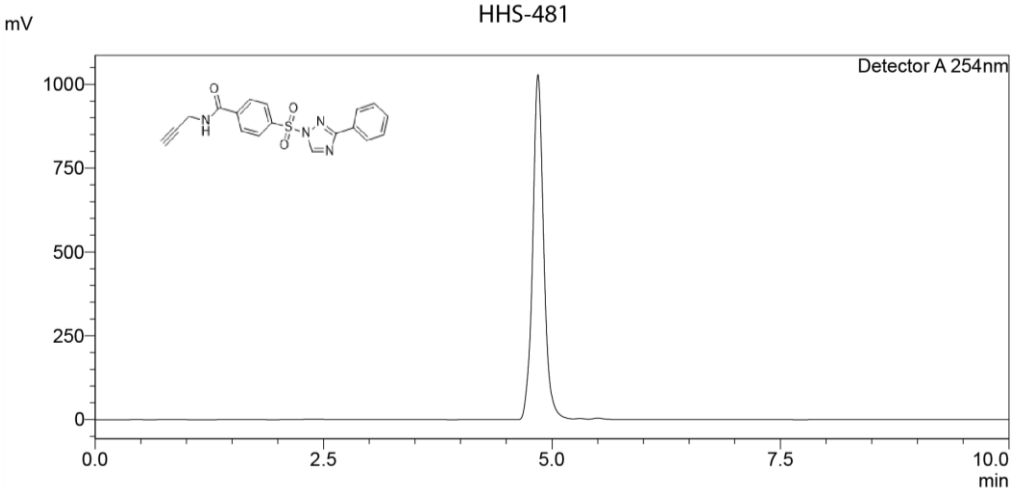
HHS-465



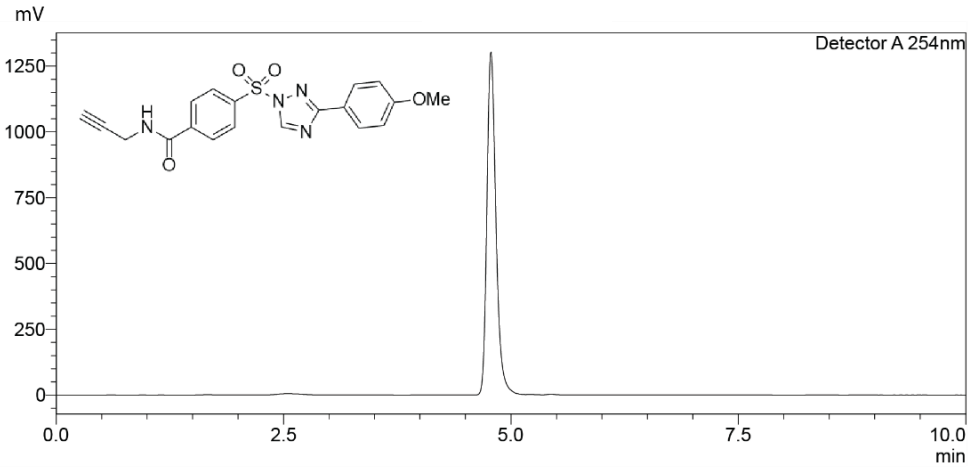
HHS-475



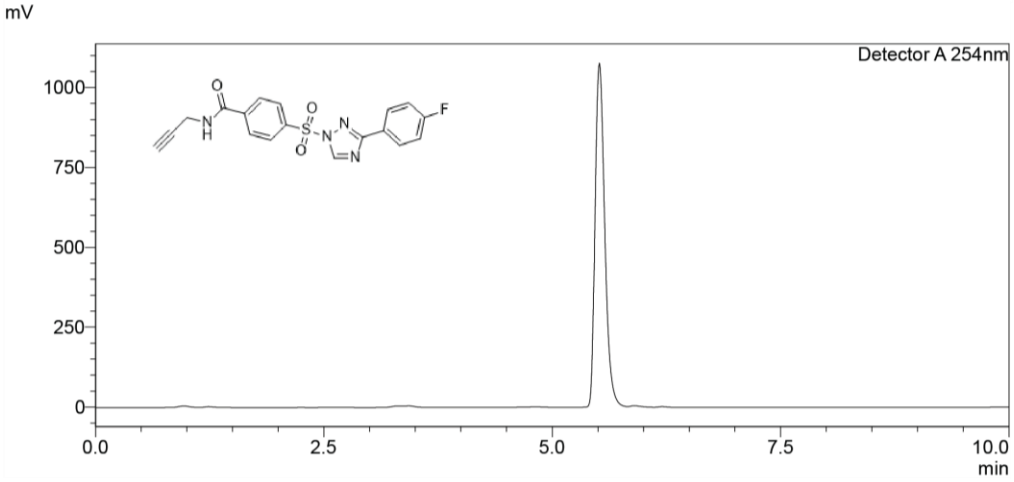
HHS-481

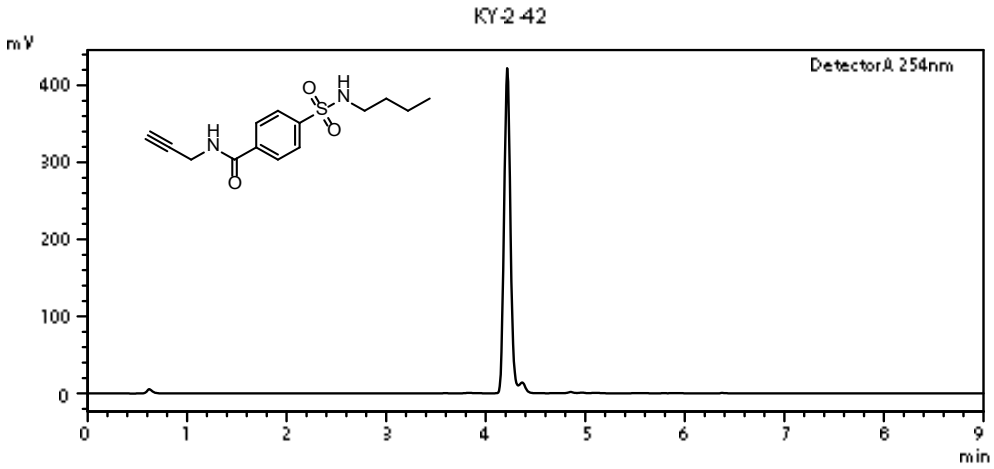


HHS-482



HHS-483




KY-2-42
mV
Detector A 254nm
400
300
200
100
0
0
1
2
3
4
5
6
7
8
9
min
O
S
O
H
N
H
N
O