Masthead Logo

University of Pennsylvania
ScholarlyCommons

Master of Chemical Sciences Capstone Projects

Department of Chemistry

5-2019

Reverse-Polarity Activity-Based Protein Profiling

Suzanne Dettling
sdett@sas.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/mcs_capstones

Reverse-Polarity Activity-Based Protein Profiling

Abstract

Reverse-polarity activity-based protein profiling (RP-ABPP) is a chemical proteomics approach that uses clickable, nucleophilic hydrazine probes to capture and identify protein-bound electrophiles in cells. The RP-ABPP approach is used to characterize the structure and function of reactive electrophilic PTMs and the proteins that harbor them, which may uncover unknown or novel functions of proteins in an endogenous setting. RP-ABPP has demonstrated utility as a versatile method to monitor metabolic regulation of electrophilic cofactors, as was done with the pyruvoyl cofactor in S-adenosyl-L- methionine decarboxylase (AMD1) and discover novel types of electrophilic modifications on proteins in human cells, as was done with the glyoxylyl modification on secernin-3 (SCRN3). These cofactors cannot be predicted by sequence and as such this area is relatively undeveloped. RP-ABPP is the only global unbiased approach to discover these electrophiles. Here, the utility of these experiments is described and a detailed protocol is provided for de novo discovery, quantitation, and global profiling of electrophilic functionality of proteins through the use of nitrogenous nucleophilic probes deployed directly to living cells in culture.

Keywords
reverse-polarity activity-based protein profiling, chemical proteomics, enzyme activity, enzyme cofactors, post-translational modifications, de novo PTM discovery

AN ABSTRACT OF THE CAPSTONE REPORT OF

Suzanne E. Dettling for the degree of Master of Chemical Sciences

Title: _Reverse-Polarity Activity-Based Protein Profiling_

Project conducted at: _University of Pennsylvania Chemistry Department_
231 S. 34 Street, Philadelphia, PA 19104-6323
Supervisor: _Megan L. Matthews, Assistant Professor of Chemistry_
Dates of Project: _5/21/18-4/24/19_

Abstract approved:

Prof. Megan L. Matthews, Academic Advisor

Reverse-polarity activity-based protein profiling (RP-ABPP) is a chemical proteomics approach that uses clickable, nucleophilic hydrazine probes to capture and identify protein-bound electrophiles in cells. The RP-ABPP approach is used to characterize the structure and function of reactive electrophilic PTMs and the proteins that harbor them, which may uncover unknown or novel functions of proteins in an endogenous setting. RP-ABPP has demonstrated utility as a versatile method to monitor metabolic regulation of electrophilic cofactors, as was done with the pyruvoyl cofactor in S-adenosyl-L- methionine decarboxylase (AMD1) and discover novel types of electrophilic modifications on proteins in human cells, as was done with the glyoxylyl modification on secernin-3 (SCRN3). These cofactors cannot be predicted by sequence and as such this area is relatively undeveloped. RP-ABPP is the only global unbiased approach to discover these electrophiles. Here, the utility of these experiments is described and a detailed protocol is provided for de novo discovery, quantitation, and global profiling of electrophilic functionality of proteins through the use of nitrogenous nucleophilic probes deployed directly to living cells in culture.

Reverse-Polarity Activity-Based Protein Profiling
by
Suzanne E. Dettling

A CAPSTONE REPORT

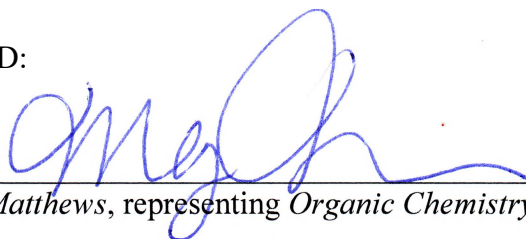submitted to the

University of Pennsylvania

in partial fulfillment of
the requirements for
the degree of

Master of Chemical Sciences

Presented *April 24, 2019*
Commencement *May 20, 2019*

<u>Master of Chemical Sciences</u> Capstone Report of *Suzanne E. Dettling* presented on *April 24, 2019*.

APPROVED:

Megan L. Matthews, representing *Organic Chemistry*

I understand that my Capstone Report will become part of the permanent collection of the University of Pennsylvania Master of Chemical Sciences Program. My signature below authorizes release of my final report to any reader upon request.

Suzanne E. Dettling, Author

Acknowledgements

Table of Contents

List of Figures

Method Review of Reverse-Polarity Activity-Based Protein Profiling

Chapter 1. Introduction

1.1 The 'Electrophilome'

Major aspects of protein function, structure, and regulation are often mediated by post-translational modifications (PTMs) (e.g. protein-protein interactions, cell stability, localization of proteins, etc.).[1, 2] Recent development of a new chemical proteomics technology, reverse-polarity activity-based protein profiling (RP-ABPP), utilizes nucleophilic probes to capture reactive electrophile PTMs in cells, thus allowing global discovery of electrophilic functionality in proteins.[2] Diverse electrophilic modifications are well known to confer function and are usually acquired through covalent installation[3, 4] or exogenous cofactor binding.[5] Proteins have been known to incorporate more than ten classes of electrophiles for essential functions, including catalysis.[6] Two examples of functional electrophiles are pyruvoyl cofactors, which facilitate catalysis of decarboxylation by forming a schiff base with the substrate, and formylglycyl cofactors, which are necessary for sulfate catalysis of hydrolysis of sulfate esters.[4] Such functionality is often difficult or impossible to predict by primary structure, yet has been found on proteins associated with disease (e.g. KEAP1 protein in cancer, APP protein in Alzheimer's disease, FTO protein in obesity, etc.).[2] RP-ABPP can also be used for de novo discovery of novel PTMs, as was done in the discovery of the glyoxylyl modification on SCRN3.[2] Additionally, some of the electrophilic sites that demonstrated high reactivity towards the nucleophilic probes represent cofactors critical for enzyme catalysis.

Characterizing the structure and function of electrophile PTMs and the proteins that harbor them can elucidate unknown or novel functions of proteins in an endogenous setting, thereby improving our understanding of disease mechanisms, potentially even revealing new drug targets. Furthermore, the probes not only function as discovery tools but also as potent inhibitors and as such may eventually serve as a launching point for development of selective small-molecule inhibitors. Despite the importance of discovering and characterizing electrophile PTMs, there has yet to be a review compiling the experiments, analyses, and significance involved in such work. As such, the goal of this review is to collect and expound on this method, its purpose, and where it fits into chemical proteomics as a whole.

Chemical proteomics methods are vastly diverse and require a broad range of knowledge. RP-ABPP is no different as these experiments incorporate biological, chemical, instrumentation, and computational aspects. RP-ABPP experiments incorporate cell culture, SDS-PAGE, western blots, cloning, mutagenesis, organic synthesis, liquid chromatography-tandem mass spectrometry, computational search strategies, peptide synthesis, and other techniques. This review will integrate all these techniques into their respective experiments and discuss how they fit into the method and what they are used to accomplish.

1.2 Development from activity-based protein profiling (ABPP)

RP-ABPP was developed from activity-based protein profiling (ABPP), a related chemical proteomics method that characterizes nucleophilic reactivity of amino acids in the proteome. ABPP is a method that utilizes electrophile probes to target functional nucleophiles,[7] while RP-ABPP works conversely, using nucleophilic hydrazine probes to target active-electrophiles as depicted in **Figure 1A**, where purple Ns represent nucleophiles and yellow Es represent electrophiles. **Figure 1B** shows the structures of the nucleophilic hydrazine probes used previously.[2] Hydrazines were selected as the nucleophilic group due to their known ability to react with oxidative and electrophilic cofactors which are shown in **Figure 1C**. Hydrazines are known to inhibit enzymes through covalent reaction with pyruvoyl,[8] aspartimide,[9] quinone,[3] and formylglycyl[10] electrophilic cofactors (the first four structures in **Figure 1C**) as well as alkylation of oxidative prosthetic groups (covalent heme cofactors[11] and covalent flavin cofactors[12]) (the final two structures in **Figure 1C**). Additionally, these hydrazine probes are deployed in living cells to focus the results on electrophiles that are functional and strongly reactive in an endogenous setting.[2]
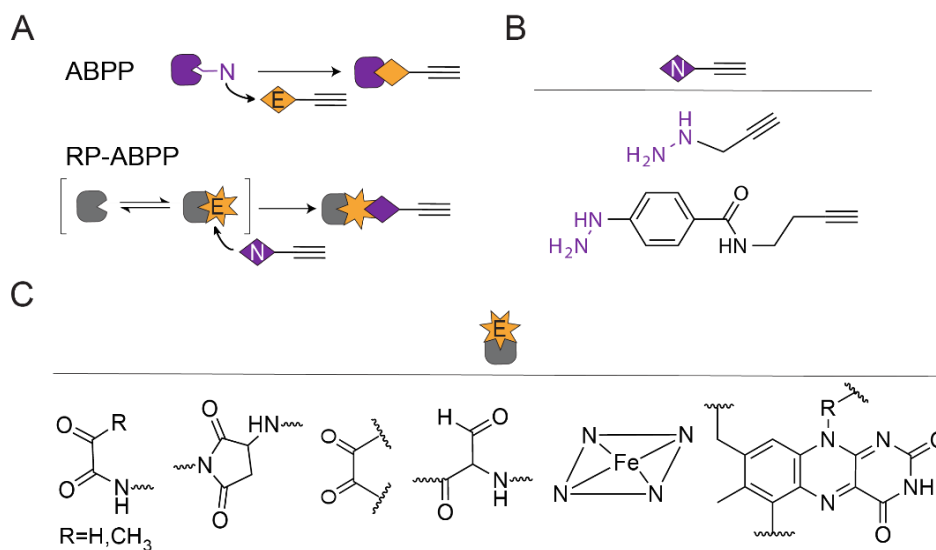
Figure 1. (A) Schematic of ABPP and its reverse polarity counterpart. N represents nucleophilic reactivity on an amino acid and E represents an electrophile probe. In the RP-ABPP schematic, E represents an electrophile PTM and N represents a nucleophile probe. **(B)** Structures of hydrazine probes used previously.[2] **(C)** Enzyme cofactors, PTMs, and prosthetic groups known to react with hydrazines.[2,3, 8-12]

The 'electrophilome', or other 'half' of the reactive proteome, has been largely ignored prior to the advent of this method due to the lack of electrophilic reactivity on un-modified amino-acids. However, initial results from this unbiased screen have demonstrated that there are functional and active electrophiles incorporated onto proteins post-translationally in the cell.[2] In ABPP, nucleophilic sites and proteins can be

predicted from sequence because specific amino acids are inherent nucleophiles. This prediction allows profiling of anticipated activity at the predicted site and protein, whereas RP-ABPP experiments must begin by discovering protein, site, and structure of the electrophile because there is no prediction to rely upon. Thus, in addition to the applications of ABPP, RP-ABPP can also be applied for de novo discovery.

All RP-ABPP experiments utilize a chemical probe, which includes a nucleophilic hydrazine warhead and an alkyne handle for detection and enrichment via 'click' chemistry (Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC)) to azide reporter tags. In this 'click' reaction, azides and alkynes react to form triazoles, enabling probes (alkyne) to bind to tags (azides) for further processing. For downstream analysis via mass spectrometry, biotin-azide tags are used as they have a very high affinity for streptavidin, allowing for detection, isolation, and purification of probe-bound proteins and peptides as depicted in the top portion of **Figure 2**. For gel experiments, fluorophore tags (e.g. rhodamine) are used to enable visual detection of probe-bound proteins via SDS-PAGE gel as depicted in the bottom portion of **Figure 2**.
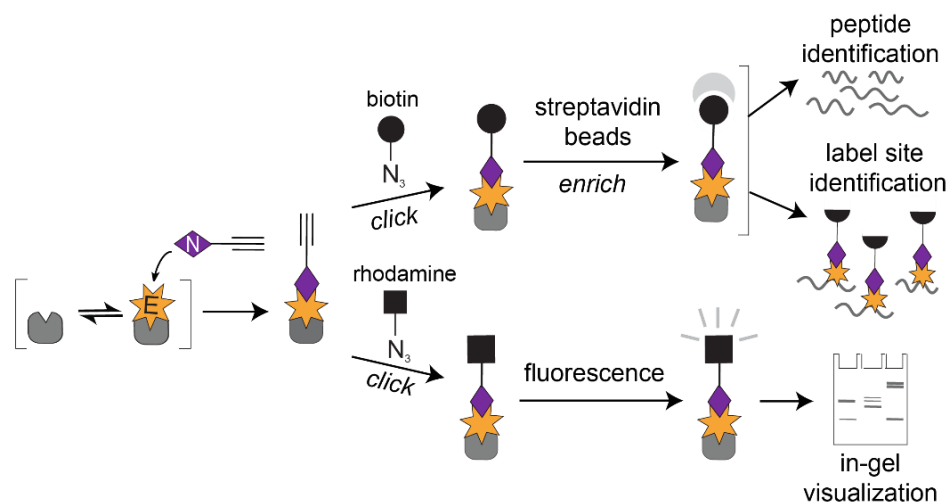
Figure 2. Schematic of utility of tags conjugated to alkyne probe. For mass spectrometry (MS) experiments, biotin tags are used to enable enrichment of probe-bound targets on streptavidin resin and conversely, for gel-based detection, a rhodamine-azide tag is used to visualize the bound electrophile using in-gel fluorescence.

1.3 Goals and overview

The overall objective of this review is to compile all experiments involved in discovering 'the electrophilome' and put these experiments into context regarding results, analysis, and significance. To achieve this, the three major aims of the method discussed and explored are: (1) identifying and quantifying proteins with probe reactivity from native systems, (2) resolving the site where the electrophile is located, and (3) determining the structure of the electrophile. These aims are accomplished through a series of experiments as depicted by the flowchart in **Figure 3**. All protocols and procedures are adapted from Matthews, M.; He, L.; Horning, B.; Olson, E.; Correia, B.; Yates III, J.; Dawson, P.; Cravatt, B. Chemoproteomic profiling and discovery of protein electrophiles in human cells. *Nat. Chem.* **2017,** *9,* 234.
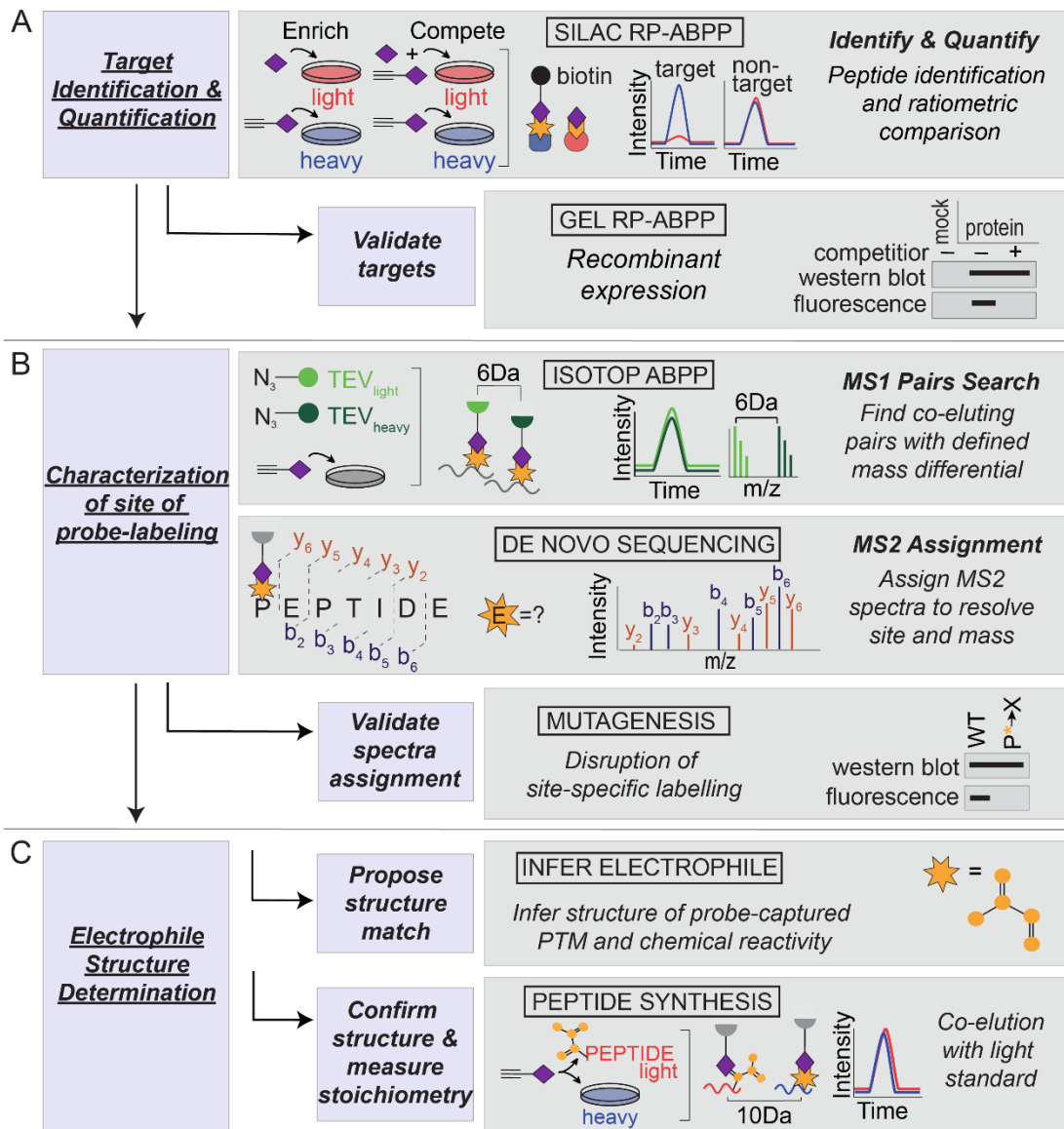
Figure 3. Schematic flowchart of RP-ABPP experiments. **(A)** Target identification and quantification is performed using SILAC RP-ABPP and targets are validated by gel RP-ABPP western blots and RP-ABPP data for hydrazine probe-treated transfected cells expressing a protein target.[2] The first lane corresponds to a control transfection ('mock') with the appropriate empty expression vector. **(B)** The site of probe-labeling is characterized using IsoTOP ABPP experiments to determine co-eluting isotopically-differentiated peptide pairs and de novo sequencing the ions to resolve the modified site.[2] The site of probe-labeling is validated through comparison of mutation and wild-type (WT) probe-labeling and expression profiles.[2] **(C)** The electrophile is determined, confirmed, and stoichiometrically quantified by inferring the electrophile and coelution of heavy-Arg/Lys-labeled transfected cells treated with probe (followed by processing by isoTOP-ABPP) with light amino acid-labeled synthetic standards.[2]

In target identification and quantification, enrichment and competition experiments are performed by treatment with probe on heavy SILAC (stable isotope labeling by amino acids in cell culture) cells and either competitor or probe with excess competitor on light SILAC cells as shown in **Figure 3A**. A competitor is a non-clickable analog of the probe without the alkyne handle. Probe-labeled proteins are conjugated to biotin, enriched on streptavidin, digested, and analyzed by LC-MS/MS examining co-eluting isotopically differentiated proteomes.[2] From these spectra, heavy to light ratios can be used to determine and quantify high-reactivity targets. These targets are then validated by visualizing competition and enrichment in cells overexpressing the target compared to 'mock' transfected controls.[2] Chapter 2 will discuss this in more detail.

To characterize the site of the electrophile, isotopic tandem orthogonal proteolysis (isoTOP) ABPP is performed to determine the peptide labeled, de novo sequencing of isoTOP ABPP data is performed to determine the specific residue labeled, and the site is validated by site-specific mutagenesis as depicted in **Figure 3B**. IsoTOP ABPP experiments involve probe-treatment of non-SILAC cells and conjugation of probe to a heavy or light tag. This isotopic differentiation on corresponding peptides allows for identification of the peptide-labeled by examining coeluting and specific mass differentiated peptides. In de novo sequencing, the MS2 spectra are assigned to specific amino acid residues to determine which residue is labeled and this result is validated by mutating that specific site and visualizing a loss of probe-labeling.[2] Chapter 3 will discuss this in more detail.

Finally, the electrophile structure is determined by first proposing a match using mass information from de novo sequencing as shown in **Figure 3C**. The structure is then confirmed by coeluting a light synthesized peptide containing the proposed probe-captured electrophile structure with heavy-labeled peptides purified from SILAC cells.[2] Similarly, light synthetic peptides of an internal standard and a modified peptide are coeluted with heavy-labeled peptides purified from SILAC cells to measure the fraction of the proteome containing the modification. In this experiment, the internal standard effectively represents the total protein (both modified and unmodified forms). By using known quantities of modified and internal standards, absolute quantification of the endogenous modified and total protein can be determined, thus enabling calculation of the fractional occupancy. This review will delve into the specifics of the experiments briefly outlined here, providing detailed protocols of RP-ABPP experiments. Chapter 4 will discuss this in more detail.

Chapter 2. Target identification and quantification

The first aim of RP-ABPP is to determine probe-reactive proteins and quantify their reactivity. Probe-reactive proteins are identified and quantified by performing labeling, purification, and mass spectrometry (MS) analysis by stable isotope labeling by amino acids in cell culture (SILAC) RP-ABPP, and validation via recombinant expression protocols. Upon identification and quantification of these proteins, high-reactivity targets can be further interrogated. Identification is an essential first step to RP-ABPP because proteins harboring electrophilic PTMs are generally unknown and cannot be predicted based on sequence. Quantification of this reactivity allows for exclusion of weakly reactive or low stoichiometry electrophiles to bias toward functional sites. Thus allowing further pursuit of highly reactive target proteins containing functional electrophiles.

2.1 Labeling and preparation

RP-ABPP experiments begin with probe, competitor, or probe and competitor mixture treatment and proteome harvesting and separation. These steps generate labeled soluble and membrane proteomes at a known concentration from cell culture. These proteomes are then used as the starting material for the rest of the RP-ABPP experiments. Below, sample procedures are described for in situ labeling (Section 2.1.1) and proteome preparation (Section 2.1.2).

2.1.1 In-situ labeling

To begin RP-ABPP, working stock solutions (~0.2–3 M) of probe and competitor compounds are prepared in water from the hydrazinium chloride salts and the pH is titrated to ~6.5–7 with concentrated sodium hydroxide. These solutions can then be stored in aliquots at –80 °C. After solution preparation, low-passage human adherent cell lines are grown at 37 °C in a humidified 5% CO_2 atmosphere and expanded in media containing high-glucose, L-glutamine, and pyruvate supplemented with 10% (v/v) fetal bovine serum (FBS), and 1% Penicillin/Streptomycin antibiotic/antimycotic.[13] For gel experiments, one 6 cm cell culture plate can be used and for MS samples, two 10 cm cell culture plates should be used.[2]

The treatment begins once the cells have reached near complete confluence, cells are washed with cold phosphate buffered saline (PBS, pH 7.4) and replenished with serum-free media (~15% of normal passage volume, e.g. a 10 cm plate normally passaged with 10 mL media would receive 1.5 mL) supplemented with 10 mM Na-HEPES buffer (pH 7.5). Then, cells are incubated with probe in the absence or presence of 10-fold excess non-clickable analogs (competitors) for 30 minutes at 37 °C. Probe and competitor should be premixed before co-administering to the cells when applicable. Following treatment, cells are washed with cold PBS to remove the probe-containing media, harvested by scraping, collected by centrifugation (1,400 g, 3 minutes, 4 °C), and washed again by resuspension in cold PBS. Unless proceeding directly to proteome preparation, cells should be pelleted and frozen at –80 °C.

2.1.2 Proteome preparation

To prepare the proteome for use in mass spectrometry-based and gel-based experiments, cell pellets are resuspended on ice in PBS (100–500 µL) and lysed by a Branson Sonifier probe sonicator (2 × 6–10 pulses, 50% duty cycle, output setting = 3–5). For this step, resuspension volume and sonication power should be adjusted accordingly for cell pellet yield. To separate soluble and membrane proteomes, samples are ultracentrifuged (100,000 g, 30–45 minutes) and the protein concentrations for each fraction are determined using the DC protein assay, which is similar the Lowry assay, and read on a microplate reader.

2.2 SILAC RP-ABPP

SILAC was initially developed in 2002, and is a widely applicable tool for quantitative proteomics.[14] In this method, two groups of cells are used: (1) cells grown in 'light' media which includes all natural abundance amino acids and (2) cells gown in 'heavy' media which contains one or more heavier isotope amino acids (e.g. amino acid(s) containing ^2H vs ^1H, ^{13}C vs ^{12}C, ^{15}N vs ^{14}N, etc.).[15] This method is a simple and easy way to isotopically differentiate samples because as the proteins incorporate the amino acids from the culture media, peptides will contain a known mass shift. This is sampler than other methods of isotopic differentiation used (e.g. reductive dimethylation (ReDiMe)) because it allows the heavy and light samples to be mixed prior to all major downstream processing. As such, this essentially allows the samples to halved when compared to other methods in which heavy and light samples must be processed in parallel and isotopic differentiation is incorporated later in the process, reducing experimental variation between the samples.

The defined mass shift created using SILAC media can then be analyzed via mass spectrometry to quantify by ratiometric comparison differences in abundance of peptides in the respective proteomes. Coelution of isotopically differentiated peptides is a common strategy in RP-ABPP as will be demonstrated throughout this review. This strategy is used with experimental samples in SILAC RP-ABPP and isoTOP ABPP as depicted in **Figure 4.** In **Figure 4A**, heavy and light cells, proteomes, and peptides are depicted in blue and red, respectively. As shown in **Figure 4A**, isotopic differentiation is utilized in SILAC RP-ABPP as protein targets are determined based on their heavy to light ratios. In **Figure 4B**, heavy and light tagged proteins and peptides are depicted in dark and light green , respectively. As is shown in **Figure 4B**, the isotopic differentiation strategy is utilized in isoTOP ABPP to determine the peptide labeled by finding coeluting peptides with a specific mass differential.
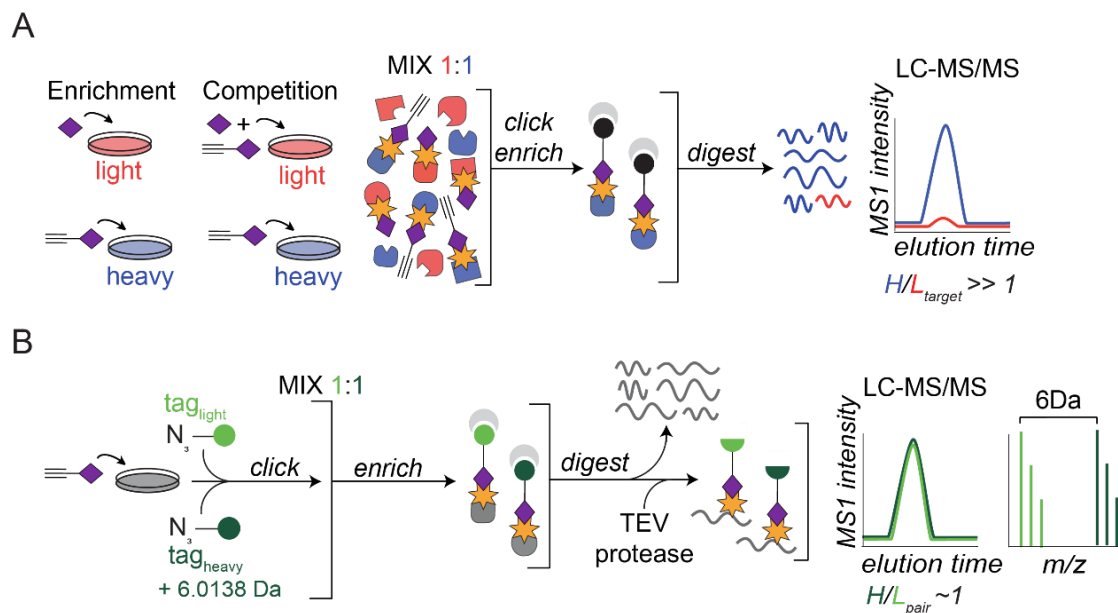
Figure 4. (A) Schematic for MS-based quantitative (SILAC) proteomics experiments (enrichment and competition) as described in the text. Heavy (H) and light (L) cells, proteomes, and peptides are depicted in blue and red, respectively. Biotin tags and streptavidin resin are depicted in black and gray, respectively. **(B)** Characterization of probe labeled peptides using the isoTOP-ABPP method as described in the text. Heavy and light tagged proteins and peptides are depicted in dark and light green, respectively. Biotin-TEV tags and streptavidin resin are depicted in green and gray respectively.

SILAC can be utilized in RP-ABPP enrichment[16] and competition[17] experiments, as shown in **Figure 4A**.[2, 18] In enrichment experiments, heavy cells are treated with probe and light cells with competitor and in competition experiments, heavy cells are treated with probe and light cells with probe and an excess of competitor. Enrichment and competition can also be shown using gel experiments. For MS-based analysis, labeled proteomes are conjugated to biotin tags via click chemistry and enriched on streptavidin beads to remove unlabeled proteins. Proteins are then digested and analyzed via LC-MS/MS. Proteins with high heavy to light ratios for competition and enrichment are deemed to be high-reactivity targets and the reactivity is quantified using these ratios. The experimental steps and analysis described for SILAC RP-ABPP are pictorially represented in **Figure 4A.** Below, sample procedures are described for MS-based analysis (Section 2.2.1), liquid chromatography-tandem MS (Section 2.2.2), and determination of high-reactivity targets (Section 2.2.3). Profiling experiments were initially adapted from those applied to other probes.[19-22]

2.2.1   MS-based analysis of probe-labeled proteins

For SILAC experiments, the *in situ* labeling protocol (Section 2.1.1) is followed, with the alteration of passaging each cell line a minimum of six times in SILAC media (lysine- and arginine-free) containing dialyzed FBS supplemented with either isotopically enriched L-[$^{13}C_6^{15}N_2$]lysine hydrochloride and L- [$^{13}C_6^{15}N_4$]arginine hydrochloride or

natural abundance isotopologues (100 μg/mL each, 550 μM and 475 μM, respectively). This allows the cells to incorporate the heavy or light amino acids. For both enrichment and competition experiments, isotopically heavy cells are treated with the probe. For enrichment experiments, isotopically light cells are treated with non-clickable analog (at the same concentration used for the probe in the heavy cells) and for competition experiments, probe in the presence of 10-fold excess non-clickable analog as a competitor are used. This ultimately allows for high-reactivity targets to bind to probe in heavy cells and bind to competitor in light cells, ensuring high heavy to light ratios in high-reactivity targets.

After labeling, isotopically heavy and light whole cell lysates are mixed in equal proportions and the proteome preparation protocol (Section 2.1.2) is followed as described above. Following the determination of the protein concentration, the fractionated equimolar mixture of heavy and light proteomes (~1−1.5 mg) is diluted to 1 mL in PBS. To conjugate the protein-bound probes to biotin tags, 110 μL of a freshly prepared 'click' reagent mixture containing 0.1 mM tris(benzyltriazolylmethyl)amine (TBTA) (60 μL/sample, 1.7 mM in 4:1 DMSO:t-BuOH), 1 mM $CuSO_4$ (20 μL/sample, 50 mM in H_2O), 100 μM biotin-azide (10 μL/sample, 10 mM in DMSO), and freshly prepared 1 mM tris(2-carboxyethyl)phosphine (TCEP) (20 μL/sample, 50 mM in PBS or H_2O) is added to each sample (1 mL) and the mixture is vortexed. The 'click' reaction is allowed to proceed for 1 hour at ambient temperature on a rotator. After the reaction mixture has rotated for 1 hour, the 'click' reaction is quenched with sequential addition of and mixture with pre-chilled methanol (MeOH, 2 mL), chloroform ($CHCl_3$, 0.5 mL), and PBS (1 mL) on ice. To fractionate the interphase protein from the organic and aqueous solvent layers, the precipitated proteome is centrifuged (5,000 g, 10 min, 4 °C) and to ensure the 'click' reagents are efficiently removed, the protein pellet is washed with cold 1:1 MeOH:$CHCl_3$ (3 × 1 mL) followed by mild sonication in cold 4:1 MeOH:$CHCl_3$ (2.5 mL).

Following conjugation to biotin tags, the samples are denatured, disulfides reduced, and resulting thiols alkylated. This is performed by pelleting the remaining precipitate by centrifugation (5,000 g, 10 min, 4 °C) and re-dissolving by mild sonication in a freshly prepared solution of proteomics-grade urea (500 μL, 6M in PBS). Disulfides are reduced with TCEP (9 mM) pre-neutralized with potassium carbonate (27 mM) for 30 minutes at 37 °C and the resulting thiols are alkylated with iodoacetamide (45 mM) for 30 minutes at ambient temperature protected from light. To ensure complete denaturation, SDS [2% (w/v)] is added.

The final steps in the SILAC RP-ABPP MS-based analysis enrich samples on streptavidin, digest samples with trypsin, and acidify with formic acid. To accomplish this, first the solution is diluted to ~0.2% SDS with PBS (~5 mL) and incubated with pre-equilibrated streptavidin agarose resin (50 μL column volume, 100 μL 1:1 slurry) for ~1.5−2 hours at ambient temperature on a rotator. Then, to remove unbound protein, excess detergent, and small molecules, streptavidin beads are collected by centrifugation (1,400 g, 1−2 min) and sequentially washed with 0.2% SDS in PBS (3 × ~10 mL), detergent-free PBS (3 × ~10 mL) and H_2O (3 × ~10 mL). To perform the digestion, the resin is transferred to a Protein LoBind tube and bound proteins are digested on-bead overnight at 37 °C in ~200 μL total volume containing sequencing grade porcine trypsin (2 μg, Promega) in the presence of urea (2 M in PBS) and $CaCl_2$ (1 mM). In the final

step, the proteolyzed supernatant is transferred to a fresh Protein LoBind tube and acidified with formic acid (5%) to inactivate trypsin and stored at −80 °C.

2.2.2 Liquid chromatography-tandem mass spectrometry (LC/LC-MS/MS)

Numerous mass spectrometry protocols exist and historically MudPIT (multidimensional protein identification technology) mass spectrometry protocols have been used for ABPP and RP-ABPP.[23-26] These protocols utilize loading acidified Trypsin products onto a C18 silica and strong cation exchange (SCX) resin column and eluting peptides with 5 salt 'bumps' followed by increasing acetonitrile (ACN). However, a more efficient method without the salt 'bumps' has recently been incorporated into RP-ABPP. [27] While MudPIT protocols enable further fractionation, they require significantly more time on the MS instrument. Utilizing another method, such as the sample protocol described below, is preferred because significantly less instrument time is required while obtaining equivalent results. In proteomics, the most limiting factor in almost all experiments is instrument time, as such, any way to reduce instrument time without compromising the quality of results is preferred.

Taking the digested acidic peptide mixture, C18 Stage Tips are used to desalt samples. Desalted samples are concentrated under reduced pressure in an evacuated centrifuge (SpeedVac) and re-dissolved in 10 µL of diluent (98% H_2O, 2% acetonitrile, 0.1% formic acid) for nanoLC-MS/MS analysis. A 3–5 µL aliquot of this solution is injected via a nano-LC system onto a 75 µm (inner diameter) fused-silica capillary column hand-packed with C18 resin and laser-pulled tip in solvent A (0.1% formic acid in H_2O). The column is developed with a 60 minute gradient of 5%–100% solvent B (20% H_2O, 80% acetonitrile, 0.1% formic acid). Peptides are ionized in positive-ion mode with a flow rate of 300 nL/min and an applied voltage of 2.3 kV. Spectra are collected in a data dependent mode such that each scan cycle involves a single high resolution (30,000) full MS spectrum of parent ions (MS1 scan from 400–1800 m/z) collected in the orbitrap coupled to a 30 CID-induced fragmentation (MS2) scans in the ion trap of the 30 most abundant parent ions from the MS1 scan. Dynamic exclusion is enabled (repeat count of 1, exclusion duration of 20 s) as is monoisotopic precursor selection. Parent ions with unassigned of +1 charge stated by the instrument are excluded for fragmentation. All other parameters are left as default values.

2.2.3 Determination of high-reactivity targets

Once mass spectrometry data collection has been performed on the samples from MS-based analysis (2.2.1), the MS2 spectra are extracted for all fragmented parent ions (from .ms2 file) from each of the .raw files generated by the instrument (Xcalibur software) using RAW Xtract or RawConverter[28] with monoisotopic selection. After extraction, each .ms2 file is searched using the ProLuCID algorithm against a reverse-concatenated, nonredundant database of the human proteome and filtered using DTASelect 2.0 within the Integrated Proteomics Pipeline (IP2) software. In this search, a static modification on cysteine residues for carboxyamidomethylation (+57.02146 Da) and up to one differential modification on methionine residues for oxidation (+15.9949 Da) are included. In the database search, peptides are also required to have at least one

tryptic terminus but allowed an unlimited number of missed cleavages. Additionally, the exact mass shift of heavy atoms on specific amino acids from the SILAC media is searched (e.g. if using ^{13}C and ^{15}N on Lysine and Arginine, K +8.0142 Da and R +10.0082 Da). This can be done by performing a coupled 'heavy' search on each dataset for both light and heavy isotopologues of the same peptide by specifying the mass shift of heavy residues as static modifications on lysine (+8.0142 Da) and arginine (+10.0082 Da).

To further filter the data, the parent ion mass tolerance for a minimum envelope of three isotopic peaks is set to 50 ppm, the minimum peptide length set to six residues, the false positive rate set to 1%, and at least 2 peptides of a protein must be detected. Heavy and light parent ion chromatograms associated with successfully identified peptides are extracted and compared using CIMAGE software.[25] Data is filtered so that at least one ion of a co-eluting heavy-light pair must be accurately identified from a fragmentation event that occurred within the retention time window (± 10 minutes) of parent ion elution. Furthermore, to ensure that the correct pair of peaks is quantified, chromatograms within a 10 ppm error tolerance of the predicted *m/z*, single-to-noise ratios greater than 2.5, and 'co-elution correlation scores' and 'envelope correlation scores' R^2 values greater than or equal to 0.8 are extracted. To further eliminate false positives and stochastic variability in the data, protein ratios are determined by median peptide ratio derived from three or more unique qualified peptides. To provide final values, protein ratios that comply with these criteria from a single experiment are averaged with ratios acquired from at least three replicates.

2.3 Validation via recombinant expression

Once high-reactivity targets have been determined, targets can be validated by demonstrating hydrazine reactivity as an intrinsic property of the protein targets that is shared by both the endogenous and recombinant forms of these proteins. This validation is performed by treating transfected cells with probe, followed by conjugation to a fluorophore tag (e.g. rhodamine), and visualization of a strong fluorescent band at the appropriate molecular weight. This band should be absent in 'mock' transfected cells and in cells treated with excess non-clickable agents as this should block probe-labeling of each protein. Additionally, recombinant expression of each protein (and lack of expression in 'mock' transfected control cells) is confirmed by western blotting. In order to perform these transfections, target genes need to be obtained in mammalian expression vectors. These vectors can be purchased or made using cloning protocols.[2] Sample procedures for transfection and gel-based analysis (Section 2.3.1) are described below.

To begin the target validation experiment, standard growth conditions are followed and cells are grown to ~40% confluence. The appropriate expression vector [control cells ('mock') receive an equal amount of the appropriate empty vector] is added as well as polyethyleneimine (PEI) 'MAX' (MW 40,000) as a transfection reagent under standard transfection conditions [3:1 vector/PEI (w/w) ratio]. Cells are incubated with transfection reagents for ~48 h before labeling *in situ* (2.1.1) and preparation of the proteome (2.1.2).

2.3.1 Gel-based analysis of probe-labeled proteins

Following transfection (Section 2.3) and labeling (Section 2.1.1) and proteome preparation (Section 2.1.2), proteomes from treated cells are diluted to 1 mg/mL. To conjugate the fluorophore to probe-labeled proteins, 6 µL of a freshly prepared "click" reagent mixture containing 0.1 mM TBTA (3 µL/sample, 1.7 mM in 4:1 DMSO:t-BuOH), 1 mM $CuSO_4$ (1 µL/sample, 50 mM in H_2O), 25 µM azide-rhodamine (1 µL/sample, 1.25 mM in DMSO), and freshly prepared 1mM TCEP (1 µL/sample, 50 mM in PBS or H_2O) is added to each sample (50 µL). Immediately upon addition of the click mixture, the samples are mixed by vortexing and the mixture is allowed to react at ambient temperature for 1 hour. The 'click' reaction is quenched with SDS loading buffer ($4\times$ stock, 17 µL) and proteins are resolved (~25 µg total protein loaded per gel lane) by SDS-PAGE (10% acrylamide gels). Labeling is visualized using in-gel fluorescence scanning on a flatbed fluorescence scanner (e.g. BioRad ChemiDoc MP). The same gel is transferred to nitrocellulose membrane and western blotted using standard protocols.[2]

2.4 Conclusion

As seen above and will be evidenced throughout the review, most RP-ABPP experiments build upon on another. For example, isoTOP ABPP experiments start with *in-situ* labeling (Section 2.1.1) and proteome preparation (Section 2.1.2) and parts of MS-based analysis (Section 2.2.1) are repeated as well. Additionally, mutagenic analysis relies upon gel-based analysis (Section 2.3.1) and begins with *in-situ* labeling (Section 2.1.1) and proteome preparation (Section 2.1.2) as well. It is important to note that these experiments are very rigorous. They require planning and attention to detail and have few acceptable pause points. One of the most critical portions of target identification and quantification is the equilibration and washing of the SILAC RP-ABPP samples on streptavidin beads. This step is critical because if the beads are washed very vigorously, some beads may be lost, causing low peptide signal. Conversely, if the beads are not washed vigorously enough, too many contaminants (e.g. unbound protein, excess detergent, small molecules, etc.) may be present, which decreases the ability to detect proteins of interest. These risks could be mitigated through use of magnetic beads, which are generally used for high-throughput applications. However, these beads are very expensive and require additional equipment.

The results generated from target identification and quantification are quite robust. When RP-ABPP was performed in two cell lines with two probes, eleven high-reactivity targets were identified and only two of these proteins were previously known to harbor an electrophile.[2] Due to the promise of these initial results, performing identification and quantification with other probes and cell lines is likely to widely expand the knowledge of known functional electrophiles.

Chapter 3. Characterization of the probe-captured site

The second aim of RP-ABPP is to determine the site of electrophilic reactivity on the protein. The peptide labeled is found using isoTOP-ABPP[25, 29] experiments and the specific amino-acid residue labeled is found using de novo sequencing. These results are validated using site-specific mutagenic analysis. Characterization of the electrophilic site yields information about where the reactivity is located within the greater context of the protein. Once this site is determined, this paves the way for a large variety of downstream experiments including monitoring the electrophile, examining its installation, and potentially investigating its function. Additionally, after the site harboring the electrophile is identified, this area can be further interrogated to determine the structure of the electrophile.

3.1 IsoTOP ABPP

IsoTOP ABPP is used to determine the peptide labeled, which is depicted in **Figure 4B**. Following probe treatment of non-SILAC cells (shown in gray in **Figure 4B**), this method utilizes conjugation to heavy and light cleavable biotin-TEV tags (shown in green in **Figure 4B**). After conjugation to heavy and light tags, labeled proteomes are enriched on streptavidin beads and on bead digestion is performed. After digestion, all unlabeled peptides are discarded as shown in **Figure 4B**. The remaining labeled peptides are released from the beads through cleavage of the tag (by TEV protease), ultimately generating probe-labeled peptides as mass differentiated pairs. Using the coeluting pair with heavy and light tags (plotted in dark and light green, respectively, in **Figure 4B**), the peptide harboring the labeled residue can be found. Determining which peptide is labeled can give information about where the reactivity is located in the sequence of the protein (e.g. active site, N-terminus, etc.). These searches enable differentiation between probe-labeled peptides and other peptides in the sample regardless of identity or mass. A sample isoTOP-ABPP procedure (Sections 3.1.1-3.1.2) is described below.

3.1.1 IsoTOP ABPP sample preparation to isolate probe-captured peptides

IsoTOP ABPP begins with probe treatment of wild-type or transfected cells *in-situ* (Section 2.1.1) and proteome preparation (Section 2.1.2). After performing these protocols, soluble proteomes (2 mg total protein) are diluted to 1 mL in PBS. To conjugate half of the proteome (0.5 mL) to the light TEV tag and the other half to the heavy TEV tag, click reactions are scaled accordingly to maintain final concentrations of 0.1 mM TBTA, 1 mM $CuSO_4$, 100 μM of light or heavy biotin-TEV-azide (5 mM in DMSO) and 1 mM TCEP. The mixture is vortexed and placed on a rotator at ambient temperature. Once the mixture has rotated for 1 hour, it is centrifuged (16,000 *g*, 5 minutes, 4 °C) and resulting pellets are mildly sonicated in ice-cold methanol (0.5 mL). The light- and heavy-labeled proteomes are combined and centrifuged once more. To solubilize the proteomes, 1.2% SDS (1 mL in PBS) is added and samples are stored at –80 °C overnight.

After leaving the samples in –80 °C overnight, samples are diluted to ~0.2% SDS with PBS (~5 mL) and incubated with pre-equilibrated streptavidin agarose resin (100 μL

1:1 slurry) for ~2–3 hours at ambient temperature. To remove un-labeled proteins, resin is washed as described above in the SILAC RP-ABPP MS-based analysis procedure (Section 2.2.1), and after the washes, the resin is transferred to clean Eppendorf tubes and resuspended in urea (500 μL, 6M in PBS). Cysteines are then reduced and alkylated with TCEP and iodoacetamide, respectively, as described above in the SILAC RP-ABPP MS-based analysis procedure (Section 2.2.1). To remove reagents, resin is washed once with PBS and bound proteins are digested with sequencing grade porcine trypsin (Promega, 2 μg) for 8–12 h at 37 °C in the presence of 2 M urea (200 μL, in PBS) and $CaCl_2$ (1 mM).

After trypsin digestion, sequential washes with PBS (5 × 0.5 mL) and H_2O (5 × 0.5 mL) are performed to remove unmodified peptides, urea, and trypsin. The resin is transferred to fresh tubes and equilibrated with TEV buffer (50 mM Tris, pH 8). To release the remaining immobilized peptides, TEV protease is added (~1–2 μM in ~200 μL TEV buffer at 30 °C for 3–5 hours). Heavy- and light-labeled peptides are collected and recovered from the resin with H_2O (2 × 50 μL). Samples are stored at –80 °C and must be analyzed within several days.

3.1.2 Characterization of probe-labeled peptides by isoTOP ABPP

To analyze the samples generated from the isoTOP preparation, data is collected from isolated probe-captured peptides using mass spectrometry protocols as described above in the SILAC RP-ABPP procedure (Section 2.2.2). Then, the data is searched on the MS1 level for paired spectra of mass differentiated coeluting peaks as shown on the right in **Figure 5**. Every recorded monoisotopic precursor mass is searched 6.0138 Da (± 5 ppm) upstream and downstream in each total ion (MS1) spectrum for a possible isotopic partner taking into account +2 and +3 charge states (3.0069 and 2.0046 Da, respectively). The relative intensity of the monoisotopic peaks is required to be greater than or equal to 5% of the base peak of each spectrum and each isotope profile (envelope) must have at least three peaks. Additionally, the Euclidean distance between two isotope profiles must be greater than or equal to 0.2. Pairs with the same m/z values (± 5 ppm) and retention times (± 10 minutes) are grouped to eliminate duplicates. Pairs of parent ions from transfected versus mock-transfected cells are analyzed and pairs of parent ions from three biological replicates should be analyzed as well.

3.2 Fragmentation spectra assignment by de novo sequencing

To further resolve the site labeled and the mass of the probe-captured PTM, the MS2 spectra from the isoTOP experiments is assigned via de novo sequencing by extraction of b and y ions and assignment spectra to their respective amino acids. The residue containing the modification can be identified as it will contain the known mass shift of the fragmented heavy/light tag, containing covalently bound probe and conjugated biotin-TEV tag. After all peaks are identified as either an amino acid or amino acid with fragmented tag, the mass of the bound electrophile can be calculated. A sample procedure is described below.
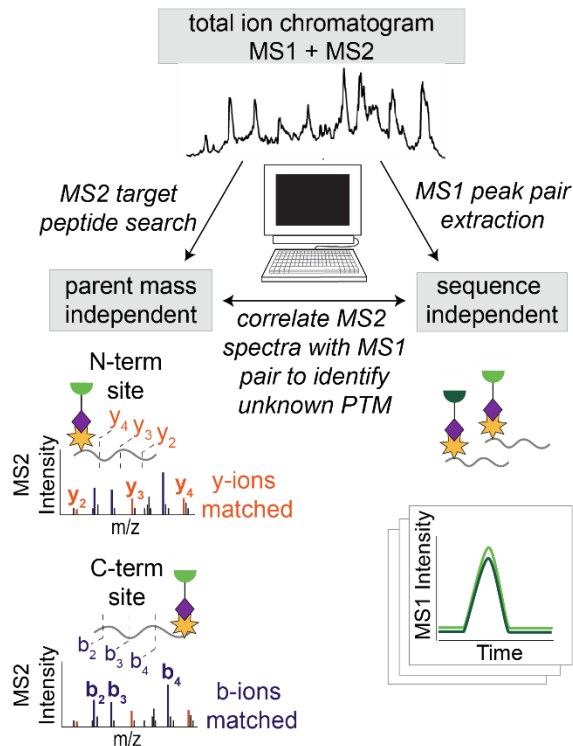
14

Figure 5. Search strategy to determine probe-labeled site from isoTOP ABPP data. To determine the peptide containing the modification, MS1 searches for coeluting pair with the specified mass difference are performed. To determine the residue labeled, MS2 spectra are assigned by de novo sequencing.

Using the MS2 spectra generated from the isoTOP experiments, spectra are assigned manually as depicted on the left in **Figure 5**. It may be necessary for the instrument method to be modified such that both the parent and fragment ions are measured in the orbitrap to gain high-resolution data for both to be certain of the peak assignments and charge state of the tag-specific b-ions. Only the five most abundant parent ions for fragmentation per cycle (versus 30 if the spectra are collected in the ion trap) are selected to account for increased scan time. The number of MS2 spectra collected per cycle is extended to ten if the data is generated on the Fusion Orbitrap. The MS2 spectra is searched for diagnostic ions of the peptide. Highly abundant fragments that represent diagnostic markers for the unmodified portion of the peptide and other fragmentation products representing the portion that contains the modification and that is differentially labeled by the isotopic tags should be found. The spectra for the heavy and light peptides are compared and the ion shift should be observed (6.0138 Da). RawConverter can be used to extract accurate monoisotopic *m/z* values for each MS2 spectra. Downstream structure determination will be based on the correct precursor mass calculated, which will be further discussed in Chapter 4.

3.3 Mutagenic analysis

After the amino acid containing the reactive electrophile has been identified, the next step is validation and confirmation of the location by mutation of the specific residue labeled. The site is confirmed if mutation of the site enables visualization of a loss of probe labeling. Additionally, expression should be present in both the wild-type and mutant protein and confirmed via western blot. To start this experiment, mutants are generated from the plasmids used in the recombinant expression procedure using QuikChange[30] site-directed mutagenesis with primers containing the desired mutations and their respective complements. Once the mutation is confirmed, cells are transfected (Section 2.3), proteomes are labeled (Section 2.1.1) and prepared (Section 2.1.2), and gel-based analysis (Section 2.3.1) is performed as described above.

3.4 Conclusion

Characterizing the site labeled encompasses very difficult experiments. The isoTOP ABPP sample preparation is particularly difficult and requires extensive planning as the samples are very sensitive and should be analyzed quickly after performing the preparation. Furthermore, as was the case in identification and quantification, incubation and washes of the samples with streptavidin resin is a critical step. In addition to the reasons mentioned above, excess urea not removed during washes can impact TEV protease activity, thereby causing difficulty in releasing immobilized peptides.[29] Further evidence of this issue is that two common sources of problems in these experiments are incomplete trypsin digestion and incomplete TEV digestion.[29] Urea is used as a denaturing agent in these experiments, but high urea concentrations can reduce trypsin and TEV protease is sensitive to even trace amounts of urea, as such, these steps represent critical points in this method.[29] Additionally, the search strategys used to determine the residue labeled are quite complex, but are designed with intent. The strengths in these searches, laid out in **Figure 5,** are that for the MS1 search, no sequence information is required to make the determination and for the MS2 search, no parent mass information is required to make the determination. Once the site has been determined and mutations at the electrophilic site have been generated, this allows for extensive further experimentation. In downstream experiments, the wild-type protein harboring the modification and the mutant that cannot contain the modification can be used to interrogate installation, function, and other aspects of the modification.

Chapter 4. Structure determination of probe-captured electrophile

The third and final goal of RP-ABPP is to determine the structure of the electrophile PTM. A structure match is initially proposed using the mass information about the PTM from de novo sequencing and chemical intuition regarding the residue and chemical reactivity of probe used. Information about enzyme class or similar motifs may also be useful. Peptides are coeluted with synthetic standards to: (1) confirm the proposed structure as well another confirmation of the site of labeling and (2) quantify the fraction of protein bearing the modification.

4.1 Validation of proposed structure by coelution with synthetic standard

Validation of electrophilic site and structure utilizes coelution of isotopically differentiated samples, light synthesized peptide with proposed probe-captured electrophile PTM and probe-labeled heavy SILAC cells. A sample procedure (Sections 4.1.1-4.1.2) is described below.

4.1.1 Synthetic standard

To generate the synthetic standard, solid phase peptide synthesis is used to prepare the peptide labeled and identified by isoTOP ABPP, including the proposed electrophile PTM bound to probe. The obtained synthetic peptides are conjugated to already cleaved tags, in contrast to the isoTOP sample preparation in which the tags are conjugated prior to cleavage. To cleave the tags, stocks of the heavy and light biotin-azide tags (0.25 μmol of each) are diluted to 250 μM with 1 mL 50 mM Tris, pH 8 supplemented with 1 mM DTT in the presence of 0.4 μM TEV protease and the reaction is incubated overnight at 30 °C. Once the reaction has incubated overnight, the reaction is concentrated to ~200 μL and the protease is precipitated with an equal volume of acetonitrile and pelleted. The supernatants containing the cleaved tags are then purified by RP-HPLC. To conjugate the synthetic hydrazone alkyne peptide to the cleaved heavy and light tags, a 'click' reaction is performed with addition of TBTA (0.1 mM), $CuCl_2$ (1 mM), and TCEP (1 mM) to a 0.5 mL mixture of the hydrazone alkyne peptide (~0.3 mM, ~0.15 μmol) and either form of cleaved tag (~60 μM, ~0.03 μmol of each) in phosphate buffer (60 mM, pH 7). The reactions are incubated at ambient temperature for approximately 2 hours and purified by RP-HPLC. The product is then neutralized with phosphate buffer (25 mM, pH 7), lyophilized, and aliquots frozen at –80 °C.

4.1.2 Coelution with synthetic standard

The samples used for coelution are generated using the isoTOP sample preparation procedure in Section 3.1.1, with the exception of growing cells in standard SILAC media. Once probe-labeled peptide pairs conjugated to tags have been isolated, the standard is diluted in water, the concentration is verified spectrophotometrically, and 0.5 pmol is added to the digested sample just prior to column loading for analysis. If the proposed structure is correct, the peptides should co-elute and appear identical in the data analysis with the only exception being the previously defined mass shift.

4.2 Determination of absolute stoichiometry with synthetic isotopolgues

The fraction of modified protein can be determined by coeluting (light) synthetic modified and internal standards with (heavy) endogenous modified and internal peptides as depicted in **Figure 6**. In these experiments, the internal peptide represents the total protein in the cell (both modified and unmodified protein). Both the endogenous modified and internal peptides absolute quantities are able to be calculated using the known amount of respective standards added to the sample and the relative peak areas of the standard to the endogenous peptides as shown by fractional occupancy in **Figure 6**. A sample procedure for these experiments is described below.
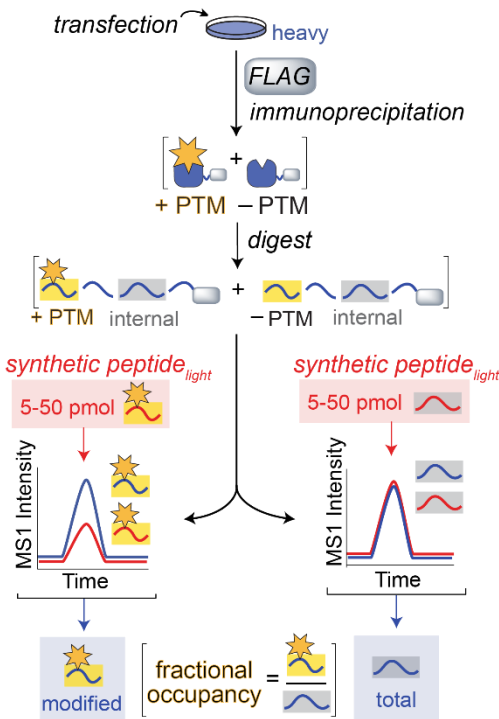
Figure 6. Scheme of stoichiometry measurement for modified peptide. Using a known quantity of synthetic standards of modified and internal peptides enables absolute quantification of the endogenous modified and total protein, allowing for calculation of fractional occupancy.

To generate the endogenous peptides in stoichiometric quantification, cells grown in heavy SILAC media are transfected and probe treated (Section 2.1.1) using procedures described above. The plasmids transfected must contain genes for the protein and FLAG tag for downstream purification. Cells are lysed in PBS (pH 7.4), fractionated by ultracentrifugation (100,000 g, 30–45 minutes), and the samples diluted to 1 mL with 50 mM Na-HEPES buffer (pH 7.5) supplemented with 500 mM NaCl and 1% Triton X-100. Unsolubilized protein remaining in the sample is pelleted by centrifugation and denatured with a small volume 10% SDS for 1 hour at 37 °C. The completely resolubilized sample is incubated at 4 °C overnight by rotation with anti-FLAG resin. The resin is washed by

resuspension and centrifugation with the same buffer supplemented with 500 mM NaCl (5 × 1 mL) followed by 100 mM NaCl (2 × 1 mL). The bound protein is eluted by incubating the beads for ~1 h at 37 °C in PBS containing 8 M urea (2 × 50 µL). Cysteines are reduced with TCEP (10 mM pre-neutralized with 30 mM potassium carbonate for 30 minutes at 37 °C) and alkylated with iodoacetamide (20 mM under the same conditions but protected from light). The samples are diluted to 2M urea with PBS and digested at 37 °C overnight with 2 µg trypsin supplemented with 1 mM CaCl$_2$. Trypsin is inactivated with 5% formic acid. The natural abundance probe-labeled peptide standard generated above is diluted in PBS and the digested protein sample with the modified standard as well as an internal peptide that represents the total protein (5–50 pmol of each) are doped in just prior to analysis by the same method used for proteomic profiling. Absolute amounts of standards should be adjusted to achieve nearly comparable peak intensities for quantitation.

4.3 Conclusion

Throughout RP-ABPP, coelution of isotopically differentiated samples is thoroughly utilized. This principle is used for SILAC, isoTOP, and synthetic standard experiments. RP-ABPP has been used for de novo discovery of a new previously unknown modification[2] and there is the potential that other human proteins may harbor previously unknown electrophilic modifications. As such, electrophilic structure determination experiments could yield novel and far-reaching results.

Chapter 5. Conclusions

5.1 Previous results and significance

Electrophilic modifications are not easily predicted by sequence and sequence predictions do not yield information about function or activity. Initial results from RP-ABPP have yielded eleven targets and these targets are largely functional and strongly implicated in disease.[2] Additionally, nine of the eleven targets were not previously known to harbor an electrophile and on one of these targets, SCRN3, a novel previously undiscovered modification, the glyoxylyl, was discovered.[2] The results of the RP-ABPP experiments used to discover the glyoxylyl are depicted in **Figure 7. Figure 7A** depicts the probe reaction with the gyloxylyl. Identification and quantification of SCRN3 as a target is shown in **Figure 7B** with extracted parent ion chromatograms and corresponding heavy to light ratios for tryptic peptides of SCRN3 protein probe-treated cells quantified in enrichment and competition (left) and quadrant plot of average competition versus enrichment SILAC ratios from quantitative proteomics experiments (right).[2] In **Figure 7C,** isoTOP ABPP MS1 ion chromatograms (left) and isotopic envelopes (right) demonstrate coelution and specific mass differentiation of the labeled SCRN3 peptide.[2] **Figure 7D** depicts de novo sequencing to determine the residue labeled and the mutagenesis gel and western blot validate the result that both C6 and D7 must be present for labeling to occur, because the y-ions resolve the modified site (**Figure 7***) to the N-terminal cysteine and/or adjacent aspartate and mutation profiles of Cys6-to-Ala6 (C6A) and Asp7-to-Phe7 (D7F) mutant SCRN3 proteins compared to wild-type (WT) SCRN3 show a lack of probe labeling.[2] In **Figure 7E,** heavy-Arg/Lys-labeled SCRN3-transfected cells treated with probe, followed by processing by isoTOP-ABPP, yields an isotopically differentiated probe-labeled SCRN3 peptide pair (light and dark green), which co-elutes with a light amino acid-labeled probe-Glyoxylyl6-Arg20 standard (also an isotopically differentiated peptide pair; light and dark gray).[2] Inset chromatogram shows all four traces scaled to the same intensity to show co-elution of endogenous and standard probe-Glyoxylyl6-Arg20 SCRN3 peptides.[2] The discovery of the glyoxylyl is significant because it demonstrates the utility of RP-ABPP in discovering new unknown electrophilic modifications. The results of these initial experiments indicate that other human proteins may harbor electrophilic groups and that there are other unknown modifications to be discovered. As such, the advent of this method is critical due to RP-ABPP being the only global unbiased approach to discover electrophilic cofactors.
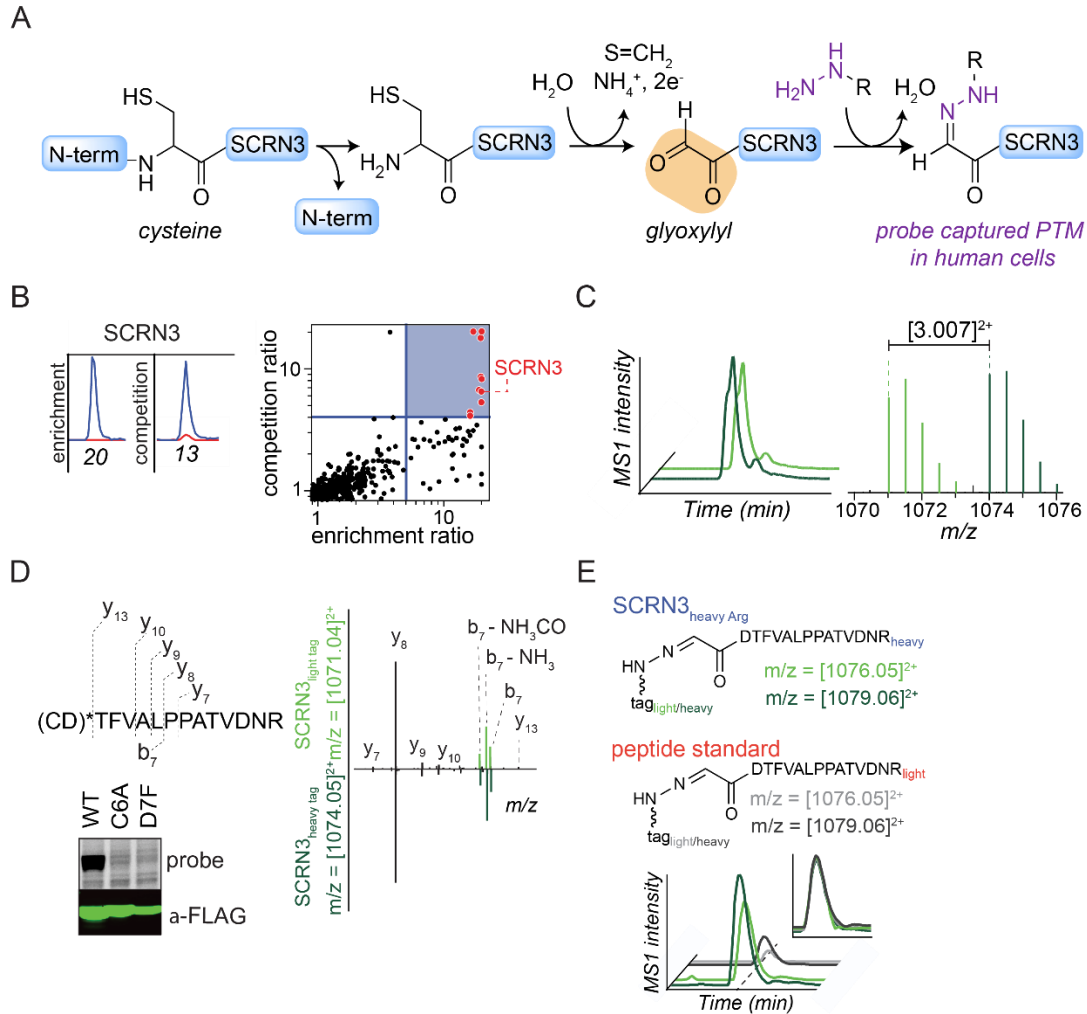
Figure 7. Sample data from RP-ABPP experiments. Adapted from Matthews, M.; He, L.; Horning, B.; Olson, E.; Correia, B.; Yates III, J.; Dawson, P.; Cravatt, B. Chemoproteomic profiling and discovery of protein electrophiles in human cells. *Nat. Chem.* **2017,** *9,* 234. **(A)** Reaction scheme of probe captured glyoxylyl on SCRN3. **(B)** SCRN3 data from SILAC RP-ABPP identification and quantification experiments. **(C)** Extracted MS1 ion chromatograms (left) and corresponding isotopic envelopes (right) for co-eluting heavy- and light-tagged peptides labeled by probe (in dark and light green, respectively). **(D)** Comparison of high-resolution MS2 spectra generated from light- versus heavy-tagged parent ions. Through de novo sequencing, the modified (*) site is determined and is confirmed by examining probe-labeling and expression profiles of mutant SCRN3 proteins compared to wild-type (WT) SCRN3. **(E)** The glyoxylyl structure and site were confirmed by coelution with synthetic standards as described in the main text.

5.2 Future applications

RP-ABPP is a recently developed method, and there are many potential future applications. These experiments are not limited to cell culture, they could easily be adapted for use in bacteria, pathogens, tissue, or in living animals. This method opens up a new area, the entire other half of the reactive proteome, the 'electrophilome.' Based on

targets generated from only two probes in two human cell lines,[2] there are more electrophile-bound proteins to be discovered and disease-relationships to be explored. In addition to discovery of protein targets and electrophilic modifications, the tools in this method can also be used to further investigate various aspects of the proteins and modifications discovered. For example, discovery of installation mechanisms and functions of these electrophiles could yield new information about the proteins and mechanisms in these cells. Additionally, as many of the protein targets discovered have strong associations to poorly understood disease mechanisms (e.g. APP – Alzheimer's Disease, FTO – Obesity, etc.) these modifications' installation and function could play an important role in or yield information about the pathogenesis of these diseases. Due to hydrazines' ability to covalent inhibit enzymes through reaction with their electrophilic cofactors, these experiments could ultimately be used to discover inhibitors and launch drug trials.

References

(1) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto Jr, G. J. Protein posttranslational modifications: the chemistry of proteome diversifications. **2005,** *44* (45), 7342-7372.

(2) Matthews, M. L.; He, L.; Horning, B. D.; Olson, E. J.; Correia, B. E.; Yates, J. R.; Dawson, P. E.; Cravatt, B. F. Chemoproteomic profiling and discovery of protein electrophiles in human cells. *Nat. Chem.* **2017,** *9* (3), 234-243.

(3) Klinman, J. P.; Bonnot, F. Intrigues and Intricacies of the Biosynthetic Pathways for the Enzymatic Quinocofactors: PQQ, TTQ, CTQ, TPQ, and LTQ. *Chemical Reviews* **2014,** *114* (8), 4343-4365.

(4) Okeley, N. M.; Van Der Donk, W. A. Novel cofactors via post-translational modifications of enzyme active sites. **2000,** *7* (7), R159-R171.

(5) Phillips, R. S. Chemistry and diversity of pyridoxal-5′-phosphate dependent enzymes. *BBA - Proteins and Proteomics* **2015,** *1854* (9), 1167-1174.

(6) Walsh, C. *Posttranslational modification of proteins: expanding nature's inventory.* Roberts and Company Publishers: 2006.

(7) Cravatt, B. F.; Wright, A. T.; Kozarich, J. W. Activity-based protein profiling: From enzyme chemistry. *Annual Review of Biochemistry* **2008,** *77*, 383-414.

(8) Shantz, L. M.; Stanley, B. A.; Secrist III, J. A.; Pegg, A. E. Purification of human S-adenosylmethionine decarboxylase expressed in Escherichia coli and use of this protein to investigate the mechanism of inhibition by the irreversible inhibitors, 5'-deoxy-5'-[(3-hydrazinopropyl) methylamino] adenosine and 5'-{[(Z)-4-amino-2-butenyl] methylamino}-5'-deoxyadenosine. *Biochemistry* **1992,** *31* (29), 6848-6855.

(9) Klaene, J. J.; Ni, W.; Alfaro, J. F.; Zhou, Z. S. Detection and Quantitation of Succinimide in Intact Protein via Hydrazine Trapping and Chemical Derivatization. *Journal of Pharmaceutical Sciences* **2014,** *103* (10), 3033-3042.

(10) Carlson, B. L.; Ballister, E. R.; Skordalakes, E.; King, D. S.; Breidenbach, M. A.; Gilmore, S. A.; Berger, J. M.; Bertozzi, C. R. Function and structure of a prokaryotic formylglycine-generating enzyme. *Journal of Biological Chemistry* **2008,** *283* (29), 20117-20125.

(11) Augusto, O.; Kunze, K. L.; de Montellano, P. O. N-Phenylprotoporphyrin IX formation in the hemoglobin-phenylhydrazine reaction. Evidence for a protein-stabilized iron-phenyl intermediate. *Journal of Biological Chemistry* **1982,** *257* (11), 6231-6241.

(12) Binda, C.; Wang, J.; Li, M.; Hubalek, F.; Mattevi, A.; Edmondson, D. E. Structural and mechanistic studies of arylalkylhydrazine inhibition of human monoamine oxidases A and B. *Biochemistry* **2008,** *47* (20), 5616-5625.

(13) Paul, J. Cell and tissue culture. **1970,** (4th Edition).

(14) 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.

(15) Mann, M. Functional and quantitative proteomics using SILAC. *Nature Reviews Molecular Cell Biology* **2006,** *7* (12), 952-958.

(16) Lannine, B. R.; Whitby, L. R.; Dix, M. M.; Douhan, J.; Gilbert, A. M.; Hett, E. C.; Johnson, T.; Joslynl, C.; Kath, J. C.; Niessen, S.; Roberts, L. R.; Schnute, M.

E.; Wang, C.; Hulce, J. J.; Wei, B. X.; Whiteley, L. O.; Hayward, M. M.; Cravatt, B. F. A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nature Chemical Biology* **2014,** *10* (9), 760-767.

(17) Ong, S.-E.; Schenone, M.; Margolin, A. A.; Li, X.; Do, K.; Doud, M. K.; Mani, D.; Kuai, L.; Wang, X.; Wood, J. L. J. P. o. t. N. A. o. S. Identifying the proteins to which small-molecule probes and drugs bind in cells. **2009,** *106* (12), 4617-4622.

(18) Chen, X.; Wong, Y. K.; Wang, J. G.; Zhang, J. B.; Lee, Y. M.; Shen, H. M.; Lin, Q. S.; Hua, Z. C. Target identification with quantitative activity based protein profiling (ABPP). *Proteomics* **2017,** *17* (3-4).

(19) Adibekian, A.; Martin, B. R.; Wang, C.; Hsu, K.-L.; Bachovchin, D. A.; Niessen, S.; Hoover, H.; Cravatt, B. F. Click-generated triazole ureas as ultrapotent in vivo-active serine hydrolase inhibitors. *Nature Chemical Biology* **2011,** *7* (7), 469-478.

(20) Martin, B. R.; Wang, C.; Adibekian, A.; Tully, S. E.; Cravatt, B. F. Global profiling of dynamic protein palmitoylation. *Nature methods* **2012,** *9* (1), 84.

(21) Hulce, J. J.; Cognetta, A. B.; Niphakis, M. J.; Tully, S. E.; Cravatt, B. F. Proteome-wide mapping of cholesterol-interacting proteins in mammalian cells. *Nature methods* **2013,** *10* (3), 259.

(22) Niphakis, M. J.; Lum, K. M.; Cognetta III, A. B.; Correia, B. E.; Ichu, T.-A.; Olucha, J.; Brown, S. J.; Kundu, S.; Piscitelli, F.; Rosen, H. A global map of lipid-binding proteins and their ligandability in cells. *Cell (Online)* **2015,** *161* (7), 1668-1680.

(23) Kline, K. G.; Wu, C. C. MudPIT analysis: application to human heart tissue. *Membrane Proteomics* **2009**, 281-293.

(24) Galmozzi, A.; Dominguez, E.; Cravatt, B. F.; Saez, E. Application of Activity-Based Protein Profiling to Study Enzyme Function in Adipocytes. *Methods in Emzymology* **2014,** *538*, 151-169.

(25) Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B. D.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **2010,** *468* (7325), 790-U79.

(26) Washburn, M. P.; Wolters, D.; Yates III, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology* **2001,** *19* (3), 242.

(27) Blaesi, E. J.; Palowitch, G. M.; Hu, K.; Kim, A. J.; Rose, H. R.; Alapati, R.; Lougee, M. G.; Kim, H. J.; Taguchi, A. T.; Tan, K. O.; Laremore, T. N.; Griffin, R. G.; Krebs, C.; Matthews, M. L.; Silakov, A.; Bollinger, J. M.; Allen, B. D.; Boal, A. K. Metal-free class Ie ribonucleotide reductase from pathogens initiates catalysis with a tyrosine-derived dihydroxyphenylalanine radical. *Proceedings of the National Academy of Sciences* **2018,** *115* (40), 10022-10027.

(28) He, L.; Diedrich, J.; Chu, Y.-Y.; Yates III, J. R. J. A. c. Extracting accurate precursor information for tandem mass spectra by RawConverter. **2015,** *87* (22), 11361-11367.

(29) Weerapana, E.; Speers, A. E.; Cravatt, B. F. Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP) - a general method for mapping sites of probe modification in proteomes. *Nature Protocols* **2007,** *2* (6), 1414-1425.

(30) Wang, W. Y.; Malcolm, B. A. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange (TM) site-directed mutagenesis. *Biotechniques* **1999,** *26* (4), 680-682.

Appendix A: Acronyms and Abbreviations

ABPP = Activity-Based Protein Profiling
ACN = Acetonitrile
APP = Amyloid Precursor Protein
CuAAC = Cu(I)-catalyzed azide-alkyne cycloaddition ('click' reaction)
DC Assay = Detergent Compatible Assay
DDA = Data Dependent Acquisition
DTT = Dithiothreitol
dFBS = dialyzed Fetal Bovine Serum
FBS = Fetal Bovine Serum
FTO = Fat mass and obesity-associated protein
HEPES = 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IAA = Iodoacetamide
IP2 = Integrated Proteomics Pipeline
IsoTOP = Isotopic Tandem Orthogonal Proteolysis
LC-MS/MS = Liquid Chromatography – Tandem Mass Spectrometry
KEAP1 = Kelch-like ECH-Associated Protein 1
MS = Mass Spectrometry
MS/MS = Tandem Mass Spectrometry
MudPIT = Multi-dimensional Protein Identification Technology
nLC = nano Liquid Chromatography
PBS = Phosphate Buffered Saline
PEI = Polyethyleneimine
PTM = Post-Translational Modification
ReDiMe = Reductive Dimethylation
RP-HPLC = Reversed Phase High-Performance Liquid Chromatography
SCRN3 = Secernin-3
SCX = Strong Cation Exchange
SDS = Sodium Dodecyl Sulfate
SDS-PAGE = Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SILAC = Stable Isotope Labelling by Amino Acids in Cell Culture
TBTA = Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
TCEP = Tris(2-carboxyethyl)phosphine
TEV = Tobacco Etch Virus
WT = Wild-Type