

Discovering Disease-Associated Enzymes by Proteome Reactivity Profiling

Katherine T. Barglow and Benjamin F. Cravatt*

The Skaggs Institute for Chemical Biology and
Department of Cell Biology and
Department of Chemistry
The Scripps Research Institute
La Jolla, California 92037

Summary

Proteomics aims to identify new markers and targets for the diagnosis and treatment of human disease. To realize this goal, methods and reagents are needed to profile proteins based on their functional properties, rather than mere abundance. Here, we describe a general strategy for synthesizing and evaluating structurally diverse libraries of activity-based proteomic probes. Quantitative screening of probe-proteome reactions coupled with bioinformatic analysis enabled the selection of a suite of probes that exhibit complementary protein reactivity profiles. This optimal probe set was applied to discover several enzyme activities differentially expressed in lean and obese (*ob/ob*) mice. Interestingly, one of these enzymes, hydroxypyruvate reductase, which was 6-fold upregulated in *ob/ob* livers, participates in the conversion of serine to glucose, suggesting that this unusual metabolic pathway may contribute to gluconeogenesis selectively in states of obesity.

Introduction

In the postgenomic era, researchers are charged with the task of assigning molecular and cellular functions to thousands of newly predicted gene products. To address this daunting challenge, several strategies have been introduced for the global analysis of genes and gene products (mRNA and proteins). Leading the way have been genomic technologies, such as DNA microarrays [1], which exhibit exceptional throughput and efficiency, permitting the comparative analysis of the entire complement of mRNA molecules in cell or tissue samples in a single experiment. However, genomic methods do not account for a variety of posttranscriptional and posttranslational events that regulate protein abundance and activity; therefore, complementary proteomic strategies are needed to directly characterize the functional state of proteins in cells and tissues.

Proteins present a more formidable analytical challenge than mRNAs, as they are chemically heterogeneous, lack general binding partners, and are nonamplifiable [2]. Accordingly, no single strategy is suitable for the simultaneous characterization of the expression and function of all proteins. For example, commonly used proteomic methods such as two-dimensional gel electrophoresis (2DE) encounter problems characterizing

membrane-associated [3] and low-abundance [4] proteins. Liquid chromatography (LC)-based strategies for proteomics, such as isotope-coded affinity tagging (ICAT) [5] and multidimensional protein identification technology (MUDPIT) [6], have overcome some of these problems; however, these gel-free methods still focus on measurements of protein expression and therefore provide only an indirect estimate of protein activity, which is regulated by a variety of posttranslational events *in vivo* [7].

Activity-based protein profiling (ABPP) is a chemical strategy for proteomics that utilizes active site-directed probes to profile the functional state of enzymes in whole proteomes [8–10]. These probes have been shown to selectively label active enzymes, but not their inactive precursor (e.g., zymogen) [11] or inhibitor-bound [12, 13] forms. To date, activity-based proteomic probes have been developed for several important enzyme classes, including serine hydrolases [8, 11, 14], cysteine proteases [15–17], metalloproteases [13], and phosphatases [18]. In each of these cases, well-known inhibitors and/or affinity labels were exploited to direct probe reactivity toward a specific class of enzyme. However, many enzyme families lack cognate affinity labels/inhibitors and are therefore less straightforward to address by ABPP.

To expand the scope of ABPP, a nondirected (combinatorial) version of this method has been advanced that enables active site-directed profiling agents to be identified for many enzymes in parallel in whole proteomes [19]. The utility of nondirected ABPP was demonstrated with a small library of probes bearing a sulfonate ester (SE) electrophile that was found to label more than six mechanistically distinct enzyme classes in an active site-directed manner [19–21]. Still, a drawback of these original studies was the limited structural diversity of the SE library, whose members were distinguished by simple alkyl/aryl groups, resulting in modest differences in their respective proteome reactivity profiles. We hypothesized that more structurally diverse libraries of electrophilic agents would contain activity-based probes for several additional enzyme classes. Here, we report a streamlined route for the synthesis and evaluation of ABPP probes bearing an α -chloroacetamide (α -CA) reactive group and a variable dipeptide binding group. Importantly, we show that members of this dipeptide probe library can be differentiated prior to the labor-intensive step of target identification by bioinformatic analysis of their respective proteome reactivity profiles, enabling the selection of an “optimal probe set” that was applied to discover several enzyme activities differentially expressed in lean and obese mice. Thus, a general strategy is presented by which libraries of candidate activity-based probes can be pruned to a minimal number of useful profiling agents, thereby increasing the speed and efficiency of ABPP experiments without sacrificing the information content garnered in these investigations.

*Correspondence: cravatt@scripps.edu

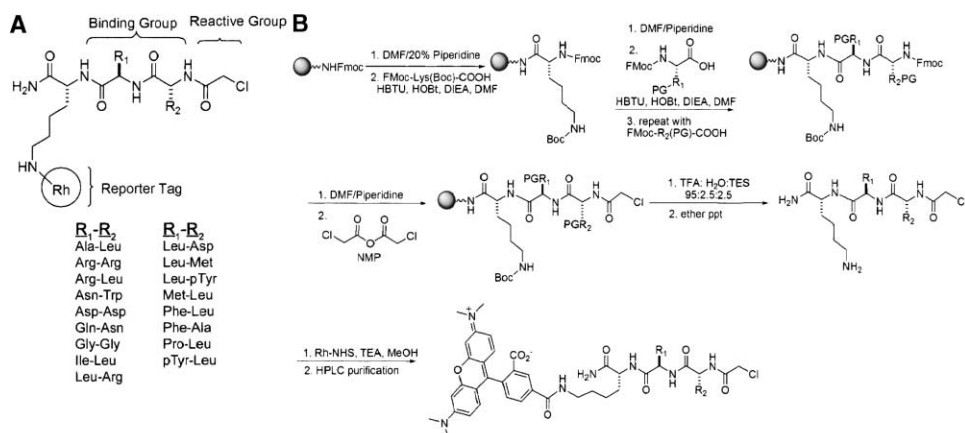


Figure 1. Design and Synthesis of an α -Chloroacetamide Dipeptide Probe Library

(A) Design of the α -chloroacetamide (α -CA) dipeptide probe library, with R₁-R₂ binding group elements listed. pTyr, phosphotyrosine. (B) Solid-phase synthesis of library, performed with Fmoc chemistry and standard peptide protecting groups (PG).

Results and Discussion

Design and Synthesis of a Dipeptide α -Chloroacetamide Probe Library

ABPP probes have been successfully designed for more than a dozen enzyme classes [9, 10]. However, numerous enzymes remain beyond the current scope of this method. The extension of ABPP to additional enzyme classes requires the design, synthesis, and screening of new probe structures. Activity-based probes are, in general, comprised of three elements: (1) a binding group to promote active site-directed interactions with a specific subset of enzymes in the proteome, (2) a reactive group (e.g., electrophile or photocrosslinker) for covalent labeling of probe-bound enzymes, and (3) a reporter tag (e.g., fluorophore or biotin) for visualization and purification of probe-labeled enzymes (Figure 1A). Combinatorial strategies for probe design offer a potentially powerful means to discover activity-based profiling tools for many enzymes in parallel [19]. The success of this nondirected approach for ABPP hinges on the generation of probe libraries of substantial structural diversity and the optimization of efficient methods for screening these reagents with proteomic samples. With these considerations in mind, we designed and synthesized a library of candidate ABPP probes bearing an α -chloroacetamide (α -CA) reactive group and a variable dipeptide binding group (Figure 1A).

The α -CA was selected as the library's reactive group for several reasons. First, this group is sterically small and therefore should not, as a conserved element of the probe library, unduly influence noncovalent probe-protein interactions. Second, due to its tempered electrophilicity, the α -CA is stable under many synthetic chemistry conditions, including standard Fmoc solid-phase peptide synthesis. Finally, there is ample precedent that other carbon electrophiles, such as sulfonate esters (SEs) and epoxides, label a variety of active site residues, including cysteines [15, 22], histidines [23], and glutamates/aspartates [22, 24] in a range of mechanistically distinct enzymes, indicating that the inherent reactivity of the α -CA probe library should not be

strongly biased toward a specific amino acid residue or enzyme class.

Attachment of the α -CA to a dipeptide binding group provided a synthetically straightforward route toward a structurally diverse set of probes, exploiting the intrinsic diversity of commercially available amino acids, including small, bulky, hydrophobic, and charged groups (Figure 1A). One of two types of reporter tags was also appended to this probe library. The initial set of probes contained a rhodamine (Rh) group for visualization of labeled proteins by in-gel fluorescence scanning. A smaller subset of the library was then synthesized as trifunctional agents containing both Rh and biotin substituents for protein detection and affinity purification, respectively [21, 25].

An 18 member probe library was manually synthesized on solid phase using Rink amide MBHA resin and standard Fmoc peptide coupling techniques (Figure 1B). This route was also compatible with automated peptide synthesis. For the Rh-tagged library, the first residue coupled to the beads was lysine, which, in the final step was reacted with an Rh-*N*-hydroxysuccinimidyl ester (Rh-NHS). Combinatorial residues R₁ and R₂ (Figure 1A) were then incorporated into the probes, and the resulting dipeptide reagents were capped with chloroacetic anhydride and then subject to global deprotection/cleavage upon treatment with 95% trifluoroacetic acid, 2.5% water, and 2.5% triethylsilane (as a scavenger). Following solution coupling to Rh-NHS, the free probes were purified by HPLC. Trifunctional probes were synthesized by coupling a biotin-modified lysine to the resin prior to the lysine, R₁, and R₂ residues, and the synthesis was completed as described above. These routes provided sufficient quantities of each probe (2–10 mg) for hundreds of proteomic experiments.

Proteome Reactivity Profiling of the Dipeptide α -CA Probe Library

Initial screening of the dipeptide α -CA library was carried out with three soluble mouse tissue proteomes (brain, heart, and liver) under the following general conditions: 5–40 μ M probe, 2 μ g/ μ l protein, 50 mM Tris-HCl (pH

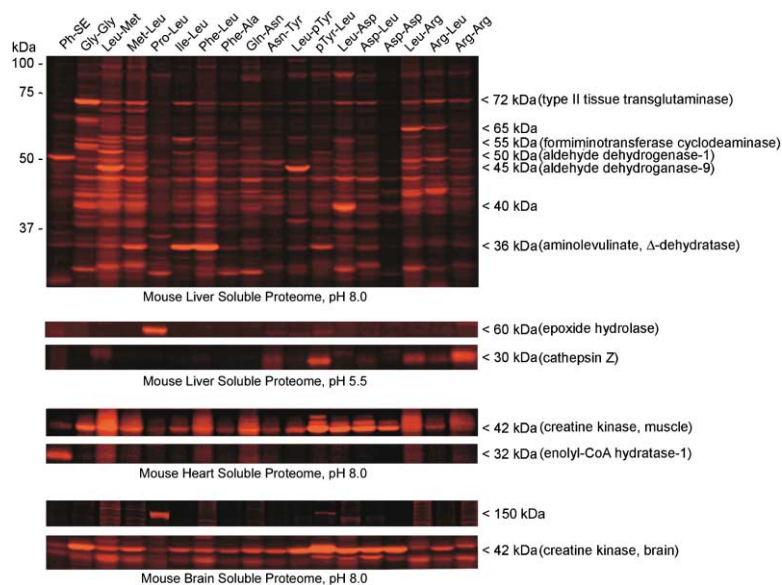


Figure 2. Proteome Reactivity Profiling of the α -Chloroacetamide Dipeptide Library

Representative proteome reactivity profiles of mouse liver, brain, and heart (soluble fractions) with the α -chloroacetamide (α -CA) dipeptide probe library. For full gels, see Supplemental Figure S1. A previously described Rh-tagged phenyl sulfonate ester (Ph-SE) probe, which labels several enzyme classes in an active site-directed manner [19], was included for comparison. Probes were tested at one of four different concentrations to normalize background (heat-insensitive) labeling: 5 (Phe-Leu), 10 (Leu-Met, Met-Leu, and Ph-SE), 20 (Gly-Gly, Ile-Leu, Phe-Ala, Gln-Asn, Asn-Tyr, Leu-pTyr, pTyr-Leu, Leu-Asp, Asp-Leu, Leu-Arg, Arg-Leu, and Arg-Arg), and 40 (Pro-Leu and Asp-Asp) μ M. A subset of probe-labeled enzymes were identified with trifunctional probes and LC-MS/MS analysis and are listed on the right side of the gel.

8.0), 1 hr labeling, at room temperature. Individual probes were found to exhibit different levels of nonspecific (heat-insensitive) proteome reactivity (e.g., see Supplemental Figure S2), and, therefore, an optimal concentration for each probe within the range of 5–40 μ M was first determined so that comparative proteomic profiling could be conducted across the probe library under conditions of normalized background. Mouse tissue proteomes were then treated with the probe library at pH 8.0 and 5.5 (50 mM MES) to accommodate enzymes that exhibit preferred activity under neutral (e.g., cytosolic enzymes) or acidic (e.g., lysosomal enzymes) conditions, respectively. Members of the probe library were found to display markedly distinct proteome reactivity profiles (Figure 2), indicating a strong contribution of the dipeptide binding group to defining specific probe-protein interactions. For example, several proteins showed preferential or exclusive reactivity with a single dipeptide α -CA probe (e.g., 40, 55, 60 [pH 5.5], and 72 kDa liver proteins). In contrast, other proteins exhibited broader reactivity, being labeled by several members of the probe library (e.g., 42 kDa brain and heart proteins). Importantly, no single probe exhibited a dominant proteome reactivity profile, indicating that even the most promiscuous probes targeted only a modest fraction of the proteins labeled by the library as a whole. Finally, little overlap was found between the proteome reactivity profiles of the α -CA dipeptide library and a phenyl SE (Ph-SE) probe [19] (Figure 2), indicating that these two classes of carbon electrophile-based probes targeted distinct portions of the proteome.

Identification and Characterization of Enzyme Targets of the Dipeptide α -CA Probe Library

To assess the target selectivity of the dipeptide α -CA probe library, a representative set of probe-labeled proteins was molecularly characterized. Briefly, protein targets were affinity isolated using trifunctional α -CA probes, digested with trypsin, and the resulting peptides analyzed by liquid chromatography-tandem mass spec-

trometry (LC-MS/MS). Enzymes from several different classes were labeled by the α -CA library (Figure 2), including type II tissue transglutaminase (tTG2), formiminotransferase cyclodeaminase (FTCD), aldehyde dehydrogenase-9 (ALDH-9), aminolevulinatase Δ -dehydratase (ADD), epoxide hydrolase (EH), cathepsin Z (cat Z), and the muscle and brain isoforms of creatine kinase (CK) (Figure 2). A subset of these enzymes represented targets of previously described activity-based probes, including EH and cat Z, which have been labeled by SE [19] and epoxide probes [26], respectively. In contrast, several other enzymes, including ALDH-9, ADD, and CK corresponded to targets labeled exclusively by α -CA probes. These enzymes showed markedly different probe reactivity profiles, highlighting the value of screening proteomes with structurally diverse probe libraries. A striking example of the selectivity of certain probe-enzyme interactions was observed with the Asp-Asp- α -CA probe, which strongly labeled both isoforms of CK, but showed negligible foreground or background reactivity with other proteins in tissue proteomes (Supplemental Figure S2). Finally, the primary Ph-SE targets ALDH-1 and enoyl CoA hydratase 1 (ECH-1) were not targeted by members of the α -CA library.

To confirm and further characterize the unique targets of α -CA probes, the ADD and CK (brain isoform) enzymes were recombinantly expressed in COS-7 cells by transient transfection and treated with a subset of the α -CA probe library. Both recombinant enzymes exhibited probe reactivity profiles similar to those of their respective natively expressed counterparts (Figure 3A). These restricted probe reactivity profiles suggested specific structure-activity relationships for the labeling of each enzyme, consistent with probe modification occurring in the active sites of these proteins. Also consistent with this premise, probe labeling of both ADD and CK were heat-sensitive and competed by known cofactors/inhibitors (Figures 3B and 3C). For example, reaction of the Asp-Asp- α -CA probe with either isoform of CK was blocked by pretreatment with the cofactor ATP

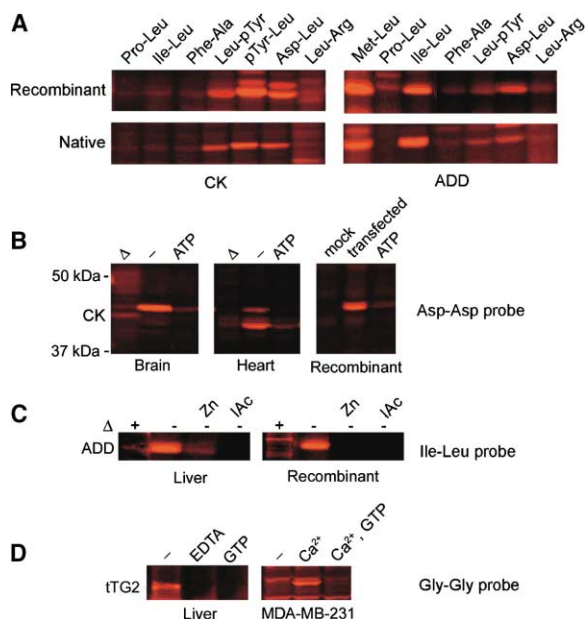


Figure 3. Characterization of Enzyme Targets of α -CA Dipeptide Probe Library

(A) Native and recombinant CK and ADD enzymes exhibit similar probe reactivity profiles. Recombinant enzymes were expressed by transient transfection in COS-7 cells.

(B) Labeling of native and recombinant CK enzymes by the Asp-Asp- α -CA probe is heat sensitive and blocked by treatment with ATP (1 mM).

(C) Labeling of native and recombinant ADD enzymes by the Ile-Leu- α -CA probe is heat sensitive and blocked by treatment with $ZnCl_2$ (1 mM) or iodoacetamide (IAc, 1 mM). Δ , heat-denatured proteomes.

(D) Labeling of tTG2 by the Gly-Gly- α -CA probe is blocked by treatment with EDTA (1 mM) or GTP (1 mM) and increased by $CaCl_2$ (1 mM) (shown for both the mouse and human tTG2 enzymes expressed in mouse liver and the cancer cell line MDA-MB-231 [21], respectively).

(1 mM) (Figure 3B). Similarly, probe labeling of ADD, which is a zinc metalloenzyme [27], was attenuated by the addition of either $ZnCl_2$ (1 mM) or the thiol-reactive reagent iodoacetamide (1 mM) (Figure 3C), indicating that probe modification may occur on one of the enzyme's conserved zinc-chelating cysteine residues, which have previously been found to react with electrophilic inhibitors [28]. Finally, labeling of the Gly-Gly- α -CA target tTG2 was found to be activated by calcium and inhibited by GTP (Figure 3D), accurately reflecting the known effects of these regulatory molecules on tTG2 catalytic activity [29]. Collectively, these data indicate that members of the α -CA probe library modify the active sites of their target enzymes.

Bioinformatic Analysis of the Proteome Reactivity Profiles of Dipeptide α -CA

Probes—Identification of an Optimal Probe Set

A major goal of nondirected ABPP is to identify within a larger library of candidate probes a subset of reagents that can collectively profile the majority of targets labeled by the library as a whole. This “optimal probe set” could then be applied to screen proteomes with maximal efficiency and minimal sample consumption. We postu-

lated that discrimination among the dipeptide α -CA probes could be accomplished without the need for target identification by comparing their relative reactivities with a large number of proteins in tissue proteomes. To test this notion, the probe reactivity profiles of ~ 25 proteins were quantified by in-gel fluorescence scanning and individually normalized to the probe that showed the strongest labeling intensity. These profiles, which represented a 19 probe \times 25 protein matrix, proved too complex to mine by visual inspection for underlying trends. However, bioinformatic analysis using hierarchical clustering analysis [30] revealed that the probes could be classified into at least five discrete subgroups based on shared proteome reactivity profiles (Figure 4A). Common structural features were observed among members of most subgroups that presumably accounted for their similar protein labeling patterns. For example, positively and negatively charged probes segregated into two distinct classes, whereas hydrophobic probes formed a third set which further branched into two subtypes. In contrast to these groups of probes, which each contained multiple members with largely overlapping proteome labeling patterns, the Pro-Leu probe was found to exhibit a strikingly unique proteome reactivity profile comprised of several proteins not targeted by other members of the probe library. These findings underscore the value of empirically determining the proteome reactivity profile of each member of a probe library to uncover agents that display unusual protein labeling patterns. Finally, the Ph-SE probe failed to cluster with any of the α -CA probes, indicating that these two probe types target complementary (rather than redundant) portions of proteomic space. From this analysis, an “optimal probe set” was defined containing representative members of each class of probes (Leu-Met and Ile-Leu [hydrophobic], Leu-Arg [positively charged], Leu-Asp [negatively charged], Pro-Leu, and Ph-SE) that was anticipated to provide $\sim 90\%$ coverage of the proteins targeted by the entire probe library (Figure 4B). We next set out to apply this optimal probe set to profile a model of human disease.

Application of the Optimal Probe Set to a Mouse Model of Obesity and Type II Diabetes

Given the rich diversity of proteins labeled by α -CA probes in the mouse liver (Figure 2), we elected to compare the profiles of this tissue isolated from wild-type (wt) mice and mice lacking the leptin gene (*ob/ob* mice). The *ob/ob* mice exhibit extreme obesity, insulin resistance, and elevated gluconeogenesis and constitute a widely used model of type II diabetes [31, 32]. In addition, liver proteomes from lean (wt) and obese (*ob/ob*) mice have recently been compared by 2DE [33], thus allowing a direct comparison of the results obtained with “conventional” proteomics methods and ABPP. Along with the optimal probe set discussed above, we also analyzed wt and *ob/ob* livers with the serine hydrolase-directed probe fluorophosphonate-rhodamine (FP-rhodamine) [14], which targets numerous esterases, lipases, and proteases [12, 34].

Several differences in the enzyme activity profiles of wt and *ob/ob* livers were observed with the optimal

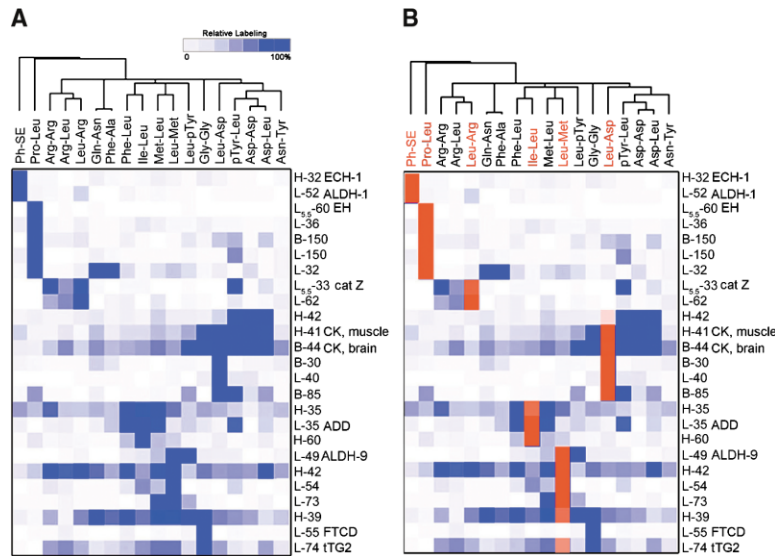


Figure 4. Cluster Analysis of Proteome Reactivity Profiles of α -CA Dipeptide Probes

The probe labeling intensities of representative liver (L), heart (H), and brain (B) proteins were quantified and, for each enzyme, these values were normalized to the probe with the highest labeling intensity (100%) to provide relative probe reactivity profiles. These relative probe reactivity profiles were then analyzed with a hierarchical clustering algorithm [30] (proteins are listed on the vertical axis by mass [kDa] and name; proteins labeled selectively at pH 5.5 are noted), which classified α -CA probes into subgroups based on similar proteome reactivity profiles (A). A representative probe from each of six main subgroups was selected to form an optimal probe set for biological experiments (profiles of the six selected probes are shown in red in [B]). (tTG2, type II tissue transglutaminase; FTCD, formiminotransferase cyclodeaminase; ALDH, aldehyde dehydrogenase; ADD, aminolevulinic acid Δ -dehydratase; EH, epoxide hydrolase; cat Z, cathepsin Z; CK, creatine kinase; and ECH-1, enoyl-CoA hydratase-1).

probe set, a representative subset of which were identified and are shown in Figure 5A (for full enzyme activity profiles, see Supplemental Figures S3 and S4). Quantitation of the probe labeling intensity of each of these enzyme activities by in-gel fluorescence scanning provided a measure of their relative expression in wt and

ob/ob livers. These proteomic data were highly reproducible, permitting the straightforward classification of individual lean and obese mice into distinct subgroups by hierarchical clustering analysis (Figure 5B; see Supplemental Figure S5 for bar graphs of the labeling data utilized in this cluster analysis).

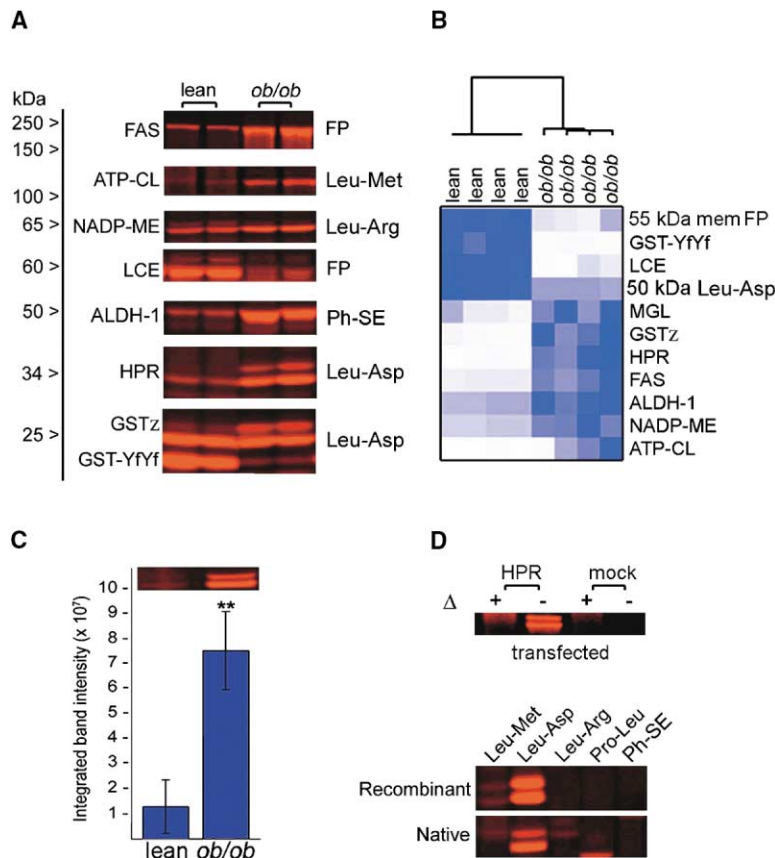


Figure 5. Application of Optimal Probe Set to Identify Enzymes Differentially Expressed in Lean (wt) and Obese (*ob/ob*) Mice

(A) A representative subset of probe-labeled enzymes that are differentially expressed in wt and *ob/ob* mouse livers. Protein molecular masses and identities are shown on the left sides of the gels. For each enzyme, the profile shown corresponds to the probe that exhibited the highest labeling intensity (probes are listed on the right side of the gels; α -CA probes listed by dipeptide binding group). (B) Cluster analysis of the probe labeling intensities of differentially expressed enzymes in four lean (wt) and four *ob/ob* mice. For (A) and (B), identified enzymes: FAS, fatty acid synthase; ATP-CL, ATP-citrate lyase; NADP-ME, NADP-dependent malic enzyme; LCE, liver carboxylesterase; ALDH-1, aldehyde dehydrogenase-1; HPR, hydroxypyruvate reductase; GST, glutathione-S-transferase; and MGL, monoglyceride lipase. Unidentified enzymes listed by molecular mass and preferred probe.

(C) Quantitation of the labeling intensity of HPR with the Leu-Asp- α -CA probe in wt and *ob/ob* livers ($n = 4$ per group).

(D) Characterization of recombinantly expressed HPR. Top, HPR-transfected, but not mock-transfected COS-7 cells contained a 35 kDa doublet that was strongly labeled by the Leu-Asp- α -CA probe in a heat-sensitive manner. Δ , heat-denatured proteome. Bottom, native and recombinant HPR enzymes show similar probe reactivity profiles (all probes tested at 20 μ M). α -CA probes are listed by dipeptide binding group.

Table 1. Relative Levels of Enzyme Activities in *ob/ob* and Wild-Type Livers Identified by ABPP

Enzyme	<i>ob/ob</i> :wt	Probe ^a	<i>ob/ob</i> :wt by 2DE ^b
GST-z (maleylacetoacetate isomerase)	8.6	Leu-Met	not detected
Fatty acid synthase	6.1	FP	4.6
Hydroxypyruvate reductase	5.8	Leu-Asp	not detected
ATP citrate lyase	5.4	Leu-Met	not detected
Monoglyceride lipase	4.5	FP	not detected
Malic enzyme	2.7	Leu-Arg	not detected
Aldehyde dehydrogenase-1	2.3	Ph-SE	2.4
Peroxiredoxin	1.6	Leu-Asp	1.5
Epoxide hydrolase	0.5	Pro-Leu	not detected
Liver carboxylesterase	0.1	FP	not detected
GST YfYf	0.04	Leu-Asp	0.12

Data are presented as averages (n = 4 per group).

^aProbe showing the strongest labeling with each enzyme was used for quantitation.

^bResults obtained previously by 2-DE analysis [33] are shown for comparison.

A comparison of the results obtained by ABPP with a previous 2DE study [33] revealed key similarities and differences. For example, multiple enzymes were identified by both methods (e.g., ALDH-1, glutathione S-transferase [GST] YfYf) and, in these cases, remarkably similar estimates of differential expression were obtained (Table 1). In contrast, several enzymes were identified by ABPP that were not observed in 2DE experiments, including targets of the FP probe liver carboxylesterase (LCE) and monoacylglycerol (MAG) lipase, which were 10- and 4.5-fold upregulated in wt and *ob/ob* livers, respectively, and the Leu-Asp- α -CA target hydroxypyruvate reductase (HPR), which exhibited approximately 6-fold greater labeling in *ob/ob* livers (Figure 5C and Table 1). Why this latter group of enzymes evaded detection by 2DE is unclear, but several explanations are possible, including comigration with abundant proteins and/or specific changes in activity, but not abundance. Thus, ABPP of wt and *ob/ob* livers uncovered both known and novel markers of obesity.

MAG lipases and carboxylesterases have been suggested to play key roles in regulating glycerol ester metabolism in vivo [35, 36] and, therefore, their altered expression in wt and *ob/ob* mice may contribute to the severe differences in lipid homeostasis evident in these animals. In this regard, it is worth noting that these hydrolytic enzymes showed opposite expression patterns in wt and *ob/ob* livers, indicating that the gross up- or downregulation of lipases is not likely a cause or consequence of fatty liver syndrome. The upregulation of HPR in *ob/ob* livers was also intriguing, as this enzyme has been shown to participate in an unusual metabolic pathway that converts serine to glucose [37, 38], which could contribute to the elevated gluconeogenic state of *ob/ob* mice. To confirm that HPR was a specific target of the Leu-Asp- α -CA probe, the cDNA for this enzyme was subcloned and transiently transfected into COS-7 cells. Strong probe labeling of a 35 kDa protein was observed in HPR-transfected cells, but not in mock-transfected cells and this labeling event was heat sensitive (Figure 5D, upper gel). Notably, the recombinant HPR enzyme, like native HPR from *ob/ob* livers, exhibited a highly restricted probe labeling profile, only reacting with the Leu-Asp- α -CA member of the optimal probe set (Figure 5D, lower gel). Collectively, these data

demonstrate that HPR is a specific target of the Leu-Asp- α -CA probe that is highly upregulated in *ob/ob* livers, suggesting that further study of the potential role of this enzyme in obesity is warranted. More generally, the selective labeling of HPR by the Leu-Asp- α -CA agent underscores the value of screening proteomes with structurally diverse sets of probes to uncover specific probe-enzyme interactions.

Conclusion

Here, we have described an integrated strategy for the design, synthesis, and analysis of libraries of structurally diverse chemical probes for the functional profiling of enzymes in whole proteomes. Key to the implementation of this nondirected method for ABPP was the selection of suitable binding group and reactive group elements for the probe library. For example, the variable dipeptide portion of the probe structure provided a synthetically facile route toward a large diversity of protein binding elements. Likewise, the α -chloroacetamide (α -CA) reactive group, consistent with the behavior of other moderately reactive carbon electrophiles [15, 19], proved capable of labeling in an active site-directed manner several mechanistically unrelated enzyme classes, thereby further expanding the scope of enzymes addressable by ABPP. In total, more than 10 different classes of enzymes were identified as targets of the α -CA probe library, most of which were not labeled by previously described ABPP probes.

A potential drawback to the utilization of probe libraries for ABPP is that, as these libraries increase in size, correspondingly greater quantities of proteome are required for analysis. For many proteomes of interest, such as tumor biopsies, limited amounts of material are available. Therefore strategies are needed to maximize the information content achieved in ABPP experiments while minimizing sample consumption. Toward this end, we demonstrated that bioinformatic analysis could be used to classify members of the α -CA probe library into subgroups based on shared proteome reactivity profiles. Representative members of each of these subgroups could then be selected to form an "optimal probe set" for biological studies. Importantly, this process for pruning probe libraries could be accomplished prior to

the labor-intensive steps of protein enrichment and identification. Considering further that the initial screening step was carried out using a handful of readily available proteomes (e.g., mouse tissues), we speculate that this time- and cost-effective approach could be applied to probe libraries of much greater size (e.g., 100–1000 probes).

Once having defined an optimal probe set, we applied these reagents to profile liver tissue from lean (wild-type [wt]) and obese (*ob/ob*) mice. Defects in liver function are suspected to be both a cause and consequence of metabolic diseases such as obesity and diabetes [32, 39], and, accordingly, liver proteins that are differentially expressed in lean and obese animals may represent novel markers and targets for the diagnosis and treatment of these diseases. The optimal probe set identified several enzyme activities with altered expression in wt and *ob/ob* livers, including the specific Leu-Asp- α -CA target hydroxypyruvate reductase, which was upregulated approximately 6-fold in fatty livers and has been implicated in the biosynthesis of glucose from serine [37, 38]. These findings suggest that nonclassical pathways for glucose biosynthesis may be operational in obesity-related syndromes like type II diabetes and could contribute to the hyperglycemic phenotype observed in these metabolic diseases [40, 41].

Finally, some of the limitations of functional proteomic analysis by ABPP should also be noted. First, even though, in each case examined here and previously [22], probe labeling was found to occur in enzyme active sites, these events may not always report on the activity state of enzymes. For example, the allosteric modulation of enzyme function at sites distal to the active site may remain undetected by ABPP. On the other hand, if, as is often the case, enzyme activity is regulated *in vivo* by autoinhibitory domains, protein partners, and/or small molecules that sterically obstruct the active site [7], then probes that are sensitive to these molecular interactions can provide an effective readout of the functional state of enzymes in complex proteomes. Second, even though the selection of optimal probe sets should greatly reduce the amount of proteome required for functional profiling experiments, the subsequent steps of affinity-enrichment and identification of probe-labeled enzymes still require significant quantities of proteome [11]. This problem may be most effectively addressed by incorporating an inhibitor/cofactor competition screen into non-directed ABPP, such that the identity of an enzyme can be inferred from the sensitivity of its probe labeling to specific active site-directed reagents [19, 34]. Indeed, as progressively greater amounts of information are garnered on the binding of inhibitors to enzymes, these data, combined with other biochemical parameters (e.g., probe reactivity profile, molecular mass, and subcellular distribution) may enable the identification of enzymes by ABPP using trace amounts of proteome (e.g., micrograms) and without the need for additional time- and sample-consuming affinity enrichment and LC-MS analysis.

Significance

The field of proteomics aims to develop new tools and methods to facilitate the functional analysis of proteins

on a global scale. Activity-based protein profiling (ABPP) offers a powerful means to visualize dynamics in enzyme function in biological samples of high complexity. However, the general and systematic application of ABPP requires the advent of chemical probes that target a large diversity of enzyme classes in an active site-directed manner. Here, we have described an integrated platform for the design, synthesis, and screening of structurally diverse libraries of ABPP probes. Quantitative proteomic screening coupled with bioinformatic analysis enabled the identification within a larger probe library of an optimal set of reagents for biological studies. Members of this optimal probe set exhibited complementary proteome reactivity profiles, thereby maximizing the information content achieved in biological experiments while minimizing time expenditure and sample consumption. The value of this strategy for ABPP was demonstrated by the discovery of several obesity-associated enzyme activities, including proteins involved in lipid metabolism and gluconeogenesis. More generally, the methods reported herein should enable the generation of functional proteomic probes for many enzyme classes, thereby enriching the repertoire of chemical reagents that can be applied to identify new diagnostic markers and therapeutic targets for human disease.

Experimental Procedures

General Materials and Instrumentation

All chemicals were obtained from Aldrich, Acros, Novabiochem, or Molecular Probes, and were used without further purification, except where noted. Triethylamine was distilled over calcium hydride and stored over potassium hydroxide. Solvents were used as received or passed over an activated alumina column (CH_2Cl_2). All reactions were carried out under an argon atmosphere using oven-baked glassware cooled in a desiccator. Purification was performed on a Hitachi D-700 HPLC system. ^1H spectra were recorded on a Bruker AMX-400 MHz spectrometer. Chemical shifts (δ) are reported relative to tetramethylsilane and coupling constants (J) are reported in Hz.

Solid-Phase Synthesis of Dipeptide α -CA Probe Library

Peptides were synthesized on Rink amide MBHA resin (Novabiochem) using standard 9-fluorenylmethoxycarbonyl (Fmoc) solid phase synthesis. Fmoc-protected resin was deprotected with 20% piperidine/dimethylformamide (DMF, 2×2 ml, 8 min) and washed ($3 \times$ DMF, $3 \times$ CH_2Cl_2 , $3 \times$ DMF) to remove excess reagents. Then Fmoc-Lys-COOH (Novabiochem, 5 equivalents) was coupled to the resin (30 min) using 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU; 4.9 equivalents), *N*-hydroxybenzotriazole (HOBt; 5 equivalents), and diisopropylethylamine (DIEA; 8 equivalents) in DMF (1 ml). Complete coupling was confirmed using a Kaiser test. The resin was washed and deprotected as described above, and fully protected Fmoc-amino acids R1 and R2 coupled as described above. Following the final deprotection, capping was effected using 1 M chloroacetic anhydride in *N*-methylpyrrolidinone (NMP, 1 ml). Following extensive washing with DCM, cleavage and deprotection were effected using 95% trifluoroacetic acid (TFA)/2.5% H_2O /2.5% triethylsilane (TES) (1 ml, 1 hr). Peptides containing phosphotyrosine protected as the phosphoramidate (Fmoc-Tyr-(PO(NMe)₂)-OH) were stirred overnight in 88% TFA/10% H_2O /2.5% TES for complete deprotection. Beads were then rinsed in the cleavage solution (2×1 ml), fractions combined, diluted with an equal volume of toluene, concentrated by rotary evaporation under reduced pressure, and dried under high vacuum overnight. Peptides were precipitated from diethyl ether and collected by vacuum filtration, then washed with ether (3×5 ml) and dried under vacuum to provide white solids. All dipeptide α -CA products were character-

ized by LC/MS and carried out without further purification. For the initial library, highly nucleophilic residues such as Lys and Cys were not utilized to avoid possible reactions with the electrophilic chloroacetyl group.

For trifunctional compounds, the first residue coupled on the resin was Fmoc-Lys(Biotin)-COOH (Novabiochem, 5 equivalents) using the same conditions as above, followed by Fmoc-Lys(Boc)-COOH, then R1 and R2 groups. The remainder of the synthesis was carried out as described above.

For synthesis on the peptide synthesizer, each coupling was performed using diisopropylcarbodiimide (DIC) in 0.45 M HOBt/NMP for 2 hr. Cleavage/deprotection was then carried out as described above.

Deprotected peptides (5 mg, 1 equivalent) were coupled at the free amine on the lysine residue to 5-(and-6)-carboxytetramethylrhodamine succinimidyl ester (5 mg, ~1 equivalent) dissolved in methanol (1 ml), and triethylamine (10 μ l, 5 equivalents). After stirring 4 hr, the solvent was removed by rotary evaporation under reduced pressure. HPLC chromatography (5%–100% acetonitrile [CH_3CN] in H_2O with 0.05% TFA), removal of CH_3CN under reduced pressure and lyophilization of the H_2O gave the final product as a dark red solid. All probes were characterized by high resolution MS (see Supplemental Data) and their purity assessed by analytical HPLC.

Proteome Sample Preparation and Probe Labeling

Mouse tissues were either purchased (Pel-Frez) or freshly isolated from animals sacrificed by CO_2 asphyxiation and flash frozen (liquid N_2) immediately upon removal. Proteomes were prepared by Dounce-homogenization of tissues in Tris buffer (50 mM Tris-HCl [pH 8.0]) or MES buffer (50 mM MES [pH 5.5]) followed by sonication. Soluble fractions (supernatant) were isolated by centrifugation at $100,000 \times g$ (45 min) to remove membrane fractions (pellet). Soluble proteomes from lean (wt) and obese (*ob/ob*) liver samples were run over two sequential PD-10 size-exclusion columns to remove excess lipids prior to analysis. Total protein concentration was determined using a protein assay kit (Bio-Rad). Samples were diluted to 2 mg/ml and stored at -80°C until use. Probe solutions were made up in DMSO (1 mM) and stored at -20°C until use. Probe-proteome reactions (50 μ l volume) were run at 5–40 μM final probe concentration at 25°C . Reactions were gently vortexed and allowed to sit for 1 hr, then quenched with equal volume $2\times$ SDS-PAGE loading buffer (reducing, with protease inhibitor). For cofactor and inhibition studies, protein was incubated with 1 mM cofactor/inhibitor for 1 hr, then labeled with probe as above. Quenched reactions were immediately placed on ice and run on 1D SDS-PAGE gels (30 μg protein/lane) or stored at -80°C . Samples were visualized in-gel using a Hitachi FMBio IIe Flatbed laser-induced fluorescence scanner (MiraiBio, Alameda, CA). Labeled proteins were quantified by measuring integrated band intensity as described previously [12] and averaged for four independent samples.

Cluster Analysis of Proteome Reactivity Profiles

For the results shown in Figure 4, the intensity of labeling of each protein target with each probe was averaged from two independent proteomic reactions and normalized to the probe with the highest labeling intensity (100%). All other values were expressed as a percentage thereof. A hierarchical clustering algorithm was applied to these profiles by average linkage clustering using the Pearson correlation coefficient as the measure of similarity (Gene Cluster computer package [30]). Cluster analysis of wt and *ob/ob* mice shown in Figure 5B was performed as above, normalizing the probe labeling intensity of each protein target to the individual mouse with the strongest labeling.

Protein Enrichment and Identification

For protein identification, 1 ml of proteome at 2 mg/ml concentration (prepared as above) was labeled with a trifunctional probe at room temperature for 2 hr. Labeled proteomes were then applied to a PD-10 size exclusion column to remove excess probe. Samples were diluted to 8 ml and incubated with avidin-agarose beads (Sigma) for 1 hr according to established protocols. Affinity-enriched proteins were separated on SDS-PAGE gels, visualized by in-gel fluorescence scanning, and bands cut using an image of the

scanned gel as a template as previously described [25]. Gel slices were washed 3 times with 50% methanol/50% water, then subjected to in-gel reduction, alkylation, and tryptic digestion. Tryptic peptides were analyzed by nanoLC-MS/MS and results searched against public databases of mouse proteins. The proteomic profiles of trifunctional probes were generally quite similar to those of the corresponding rhodamine probes. However, the rhodamine probes, in general, displayed less background labeling and were therefore used for initial screening experiments.

Recombinant Expression and Characterization of Enzyme Targets of the Dipeptide α -CA Library

cDNAs corresponding to the enzyme targets of the α -CA library were purchased as expressed sequence tags (Invitrogen), sequenced, subcloned into pcDNA3.1, and transiently transfected into COS-7 cells using established molecular biology techniques. Transfected cells were harvested after 48 hr by scraping, then were homogenized in phosphate-buffered saline (PBS), Dounce-homogenized, sonicated, and spun at $100,000 \times g$ for 45 min. Soluble protein (supernatant) was diluted to 1 mg/ml and labeled with probe as described above.

Supplemental Data

Supplemental Figures and Experimental Procedures for this article are available online at <http://www.chembiol.com/cgi/content/full/11/11/1523/DC1/>.

Acknowledgments

We thank Dr. Arul Joseph and E. Gregory Hawkins for assistance with the initial design, synthesis, and screening of the dipeptide probe library and members of the Cravatt group for helpful discussions and critical reading of this manuscript. This work was supported by the National Institutes of Health (Grant no. CA087660), the Skaggs Institute for Chemical Biology, and a Howard Hughes Medical Institute predoctoral fellowship (to K.T.B.).

Received: July 12, 2004

Revised: August 10, 2004

Accepted: August 17, 2004

Published: November 26, 2004

References

1. Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E., and Davis, R.W. (1998). Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* **16**, 301–306.
2. Kodadek, T. (2001). Protein microarrays: prospects and problems. *Chem. Biol.* **8**, 105–115.
3. Santoni, V., Molloy, M., and Rabilloud, T. (2000). Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054–1070.
4. Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F., and Sanchez, J.C. (2000). The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* **21**, 1104–1115.
5. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999.
6. Washburn, M.P., Wolters, D., and Yates, J.R., III. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247.
7. Kobe, B., and Kemp, B.E. (1999). Active site-directed protein regulation. *Nature* **402**, 373–376.
8. Liu, Y., Patricelli, M.P., and Cravatt, B.F. (1999). Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. USA* **96**, 14694–14699.
9. Jessani, N., and Cravatt, B.F. (2004). The development and application of methods for activity-based protein profiling. *Curr. Opin. Chem. Biol.* **8**, 54–59.
10. Speers, A.E., and Cravatt, B.F. (2004). Chemical strategies for activity-based proteomics. *ChemBiochem* **5**, 41–47.

11. Kidd, D., Liu, Y., and Cravatt, B.F. (2001). Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **40**, 4005–4015.
12. Jessani, N., Liu, Y., Humphrey, M., and Cravatt, B.F. (2002). Enzyme activity profiles of the secreted and membrane proteome that depict cancer invasiveness. *Proc. Natl. Acad. Sci. USA* **99**, 10335–10340.
13. Saghatelian, A., Jessani, N., Joseph, A., Humphrey, M., and Cravatt, B.F. (2004). Activity-based probes for the proteomic profiling of metalloproteases. *Proc. Natl. Acad. Sci. USA* **101**, 10000–10005.
14. Patricelli, M.P., Giang, D.K., Stamp, L.M., and Burbaum, J.J. (2001). Direct visualization of serine hydrolase activities in complex proteome using fluorescent active site-directed probes. *Proteomics* **1**, 1067–1071.
15. Greenbaum, D., Medzihradsky, K.F., Burlingame, A., and Bogyo, M. (2000). Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **7**, 569–581.
16. Faleiro, L., Kobayashi, R., Fearnhead, H., and Lazebnik, Y. (1997). Multiple species of CPP32 and Mch2 are the major active caspases present in apoptotic cells. *EMBO J.* **16**, 2271–2281.
17. Borodovsky, A., Ovaa, H., Kolli, N., Gan-Erdene, T., Wilkinson, K.D., Ploegh, H.L., and Kessler, B.M. (2002). Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**, 1149–1159.
18. Kumar, S., Zhou, B., Liang, F., Wang, W.Q., Huang, Z., and Zhang, Z.Y. (2004). Activity-based probes for protein tyrosine phosphatases. *Proc. Natl. Acad. Sci. USA* **101**, 7943–7948.
19. Adam, G.C., Sorensen, E.J., and Cravatt, B.F. (2002). Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype. *Nat. Biotechnol.* **20**, 805–809.
20. Adam, G.C., Cravatt, B.F., and Sorensen, E.J. (2001). Profiling the specific reactivity of the proteome with non-directed activity-based probes. *Chem. Biol.* **8**, 81–95.
21. Adam, G.C., Sorensen, E.J., and Cravatt, B.F. (2002). Trifunctional chemical probes for the consolidated detection and identification of enzyme activities from complex proteomes. *Mol. Cell. Proteomics* **1**, 828–835.
22. Adam, G.C., Burbaum, J.J., Kozarich, J.W., Patricelli, M.P., and Cravatt, B.F. (2004). Mapping enzyme active sites in complex proteomes. *J. Am. Chem. Soc.* **126**, 1363–1368.
23. Liu, S., Widom, J., Kemp, C.W., Crews, C.M., and Clardy, J. (1998). Structure of human methionine aminopeptidase-2 complexed with fumagillin. *Science* **282**, 1324–1327.
24. Withers, S.G., and Aebersold, R. (1995). Approaches to labeling and identification of active site residues in glycosidases. *Protein Sci.* **4**, 361–372.
25. Speers, A.E., and Cravatt, B.F. (2004). Profiling enzyme activities in vivo using click chemistry methods. *Chem. Biol.* **11**, 535–546.
26. Greenbaum, D., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradsky, K.F., and Bogyo, M. (2002). Chemical approaches for functionally probing the proteome. *Mol. Cell. Proteomics* **1**, 60–68.
27. Shoolingin-Jordan, P.M., Spencer, P., Sarwar, M., Erskine, P.E., Cheung, K.M., Cooper, J.B., and Norton, E.B. (2002). 5-Aminolaevulinic acid dehydratase: metals, mutants and mechanism. *Biochem. Soc. Trans.* **30**, 584–590.
28. Seehra, J.S., and Jordan, P.M. (1981). 5-Aminolaevulinic acid dehydratase: alkylation of an essential thiol in the bovine-liver enzyme by active-site-directed reagents. *Eur. J. Biochem.* **113**, 435–446.
29. Liu, S., Cerione, R.A., and Clardy, J. (2002). Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. *Proc. Natl. Acad. Sci. USA* **99**, 2743–2747.
30. Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868.
31. Johnson, P.R., Greenwood, M.R., Horwitz, B.A., and Stern, J.S. (1991). Animal models of obesity: genetic aspects. *Annu. Rev. Nutr.* **11**, 325–353.
32. Yoon, J.C., Puigserver, P., Chen, G., Donovan, J., Wu, Z., Rhee, J., Adelmant, G., Stafford, J., Kahn, C.R., Granner, D.K., et al. (2001). Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* **413**, 131–138.
33. Edvardsson, U., von Lowenhielm, H.B., Panfilov, O., Nystrom, A.C., Nilsson, F., and Dahllof, B. (2003). Hepatic protein expression of lean mice and obese diabetic mice treated with peroxisome proliferator-activated receptor activators. *Proteomics* **3**, 468–478.
34. Leung, D., Hardouin, C., Boger, D.L., and Cravatt, B.F. (2003). Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. *Nat. Biotechnol.* **21**, 687–691.
35. Karlsson, M., Contreras, J.A., Hellman, U., Tornqvist, H., and Holm, C. (1997). cDNA cloning, tissue distribution, and identification of the catalytic triad of monoglyceride lipase. Evolutionary relationship to esterases, lysophospholipases, and haloperoxidases. *J. Biol. Chem.* **272**, 27218–27223.
36. David, L., Guo, X.J., Villard, C., Moulin, A., and Puigserver, A. (1998). Purification and molecular cloning of porcine intestinal glycerol-ester hydrolase—evidence for its identity with carboxylesterase. *Eur. J. Biochem.* **257**, 142–148.
37. Xue, H.H., Fujie, M., Sakaguchi, T., Oda, T., Ogawa, H., Kneer, N.M., Lardy, H.A., and Ichiyama, A. (1999). Flux of the L-serine metabolism in rat liver. The predominant contribution of serine dehydratase. *J. Biol. Chem.* **274**, 16020–16027.
38. Xue, H.H., Sakaguchi, T., Fujie, M., Ogawa, H., and Ichiyama, A. (1999). Flux of the L-serine metabolism in rabbit, human, and dog livers. Substantial contributions of both mitochondrial and peroxisomal serine:pyruvate/alanine:glyoxylate aminotransferase. *J. Biol. Chem.* **274**, 16028–16033.
39. Lind, P. (2004). Interdependence of hepatic lipid and glucose metabolism: novel pharmacological targets for diabetes. *Curr. Opin. Investig. Drugs* **5**, 395–401.
40. Kurukulasuriya, R., Link, J.T., Madar, D.J., Pei, Z., Richards, S.J., Rohde, J.J., Souers, A.J., and Szczepankiewicz, B.G. (2003). Potential drug targets and progress towards pharmacologic inhibition of hepatic glucose production. *Curr. Med. Chem.* **10**, 123–153.
41. Ross, S.A., Gulve, E.A., and Wang, M. (2004). Chemistry and biochemistry of type 2 diabetes. *Chem. Rev.* **104**, 1255–1282.