nature
chemical biology

© 2008 Nature Publishing Group http://www.nature.com/naturechemicalbiology

Disparate proteome reactivity profiles of carbon electrophiles

Eranthie Weerapana, Gabriel M Simon & Benjamin F Cravatt

Insights into the proteome reactivity of electrophiles are crucial for designing activity-based probes for enzymes lacking cognate affinity labels. Here, we show that different classes of carbon electrophiles exhibit markedly distinct amino acid labeling profiles in proteomes, ranging from selective reactivity with cysteine to adducts with several amino acids. These data specify electrophilic chemotypes with restricted and permissive reactivity profiles to guide the design of next-generation functional proteomics probes.

The field of activity-based protein profiling (ABPP) applies reactive chemical probes to profile the functional state of enzymes in native proteomes[1]. Original ABPP probes incorporated well-defined affinity labels as reactive groups to target enzyme classes such as the serine[2] and cysteine[3] hydrolases. Many enzymes, however, do not have cognate affinity labels, and the design of ABPP probes for these proteins remains challenging. Structural insights into the substrate-binding pocket of enzyme classes can reveal nucleophilic residues for targeting with appropriate electrophiles. Recent work in the design of protein kinase probes positioned α-fluoromethylketone and acylphosphate electrophiles within an ATP scaffold to exploit the nucleophilicity of proximal cysteine[4] and lysine[5] residues, respectively. Differentiating among electrophilic chemotypes that show restricted and permissive amino acid reactivity profiles should streamline such endeavors to design ABPP probes for a wide range of enzyme classes.

A variety of electrophiles are available for incorporation into ABPP probes. The proteome reactivity profiles of iodoacetamide and maleimide reactive groups have been extensively investigated[6]. Here, we expand on these studies by investigating the reactivity of a panel of carbon electrophiles (**Fig. 1a**) comprising a phenylsulfonate ester (SE, 1), a linear epoxide (EP, 2), an α-chloroacetamide (CA, 3), an α,β-unsaturated ketone (UK, 4) and a spiro-epoxide (SP, 5) in complex proteomes (see **Supplementary Methods** online). An alkyne was incorporated into these electrophilic frameworks to provide a click chemistry handle for gel and mass spectrometric analysis[7]. Application of these electrophiles to a soluble mouse liver proteome, followed by click chemistry with a rhodamine azide (Rh-N_3) reporter tag and visualization of labeled proteins by SDS-PAGE and in-gel fluorescence scanning, demonstrated that the panel of electrophiles exhibit a range of protein reactivities (see **Supplementary Fig. 1** online). The highest reactivity was observed for the UK probe, which demonstrated substantial protein labeling at concentrations as low as 1 μM. The CA and SE electrophiles demonstrated moderate levels of reactivity, whereas the EP and SP probes displayed little to no protein labeling, even at concentrations up to 20 μM.

We then examined in greater depth the protein and amino acid labeling profiles for the three probes that displayed the highest levels of proteome reactivity (SE, CA and UK). To address this question, we used a mass spectrometry platform referred to as tandem orthogonal proteolysis (TOP)-ABPP for simultaneous identification of protein targets and exact sites of probe modification[8]. The probes were applied

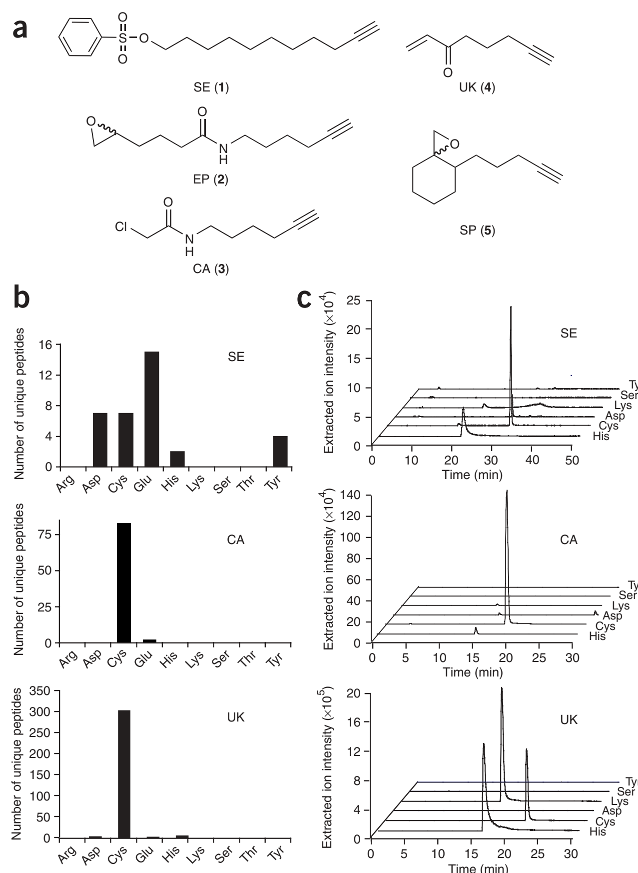

Figure 1 Proteome and solution reactivities of carbon electrophile probes. (**a**) Panel of probes used in this study. (**b**) Proteome reactivity profiles for SE, CA and UK probes demonstrating the number of assigned peptides with unique labeling sites on nine nucleophilic amino acid residues (top, SE; middle, CA; bottom, UK). (**c**) Solution reactivity profiles for SE, CA and UK probes (top, SE; middle, CA; bottom, UK). Representative extracted ion chromatograms are shown for product formation upon reacting the probes with 20 equivalents of each amino acid derivative in PBS for 12 h.

The Skaggs Institute for Chemical Biology and Department of Chemical Physiology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. Correspondence should be addressed to B.F.C. (cravatt@scripps.edu).

Received 20 March; accepted 29 April; published online 18 May 2008; corrected online 29 May 2008 (details online); doi:10.1038/nchembio.91

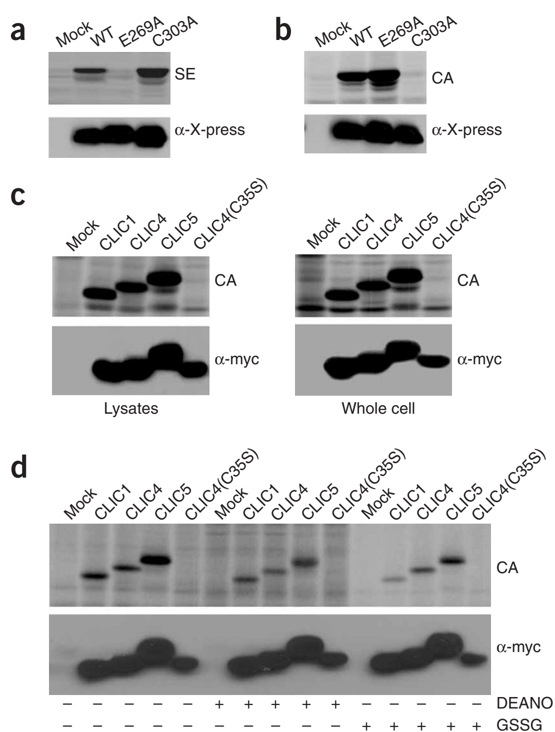

Figure 2 Labeling of ALDH and the CLIC protein family with carbon electrophile probes. (**a**) Labeling of WT, E269A and C303A ALDH-1 enzymes with the SE probe. (**b**) Labeling of WT, E269A and C303A ALDH-1 enzymes with the CA probe. Top, fluorescent gel images shown in grayscale demonstrating the selective labeling of the E269A and C303A mutants with the CA and SE probes, respectively. Bottom, western blots confirming equivalent expression of WT, E269A and C303A ALDH-1 enzymes using α-X-press antibodies. (**c**) Labeling of CLICs with the CA probe in lysates and in whole cells. Top, fluorescent gel images shown in grayscale. Bottom, western blots confirming expression of CLICs using α-myc antibodies. (**d**) Nitric oxide and oxidized glutathione treatment of CLICs. Lysates were treated with either 5 mM of the nitric oxide donor diethylamine nitric oxide sodium salt (DEANO) or 2 mM oxidized glutathione (GSSG). Top, fluorescent gel images shown in grayscale. Bottom, western blots confirming expression of CLICs using α-myc antibodies.

to four different mouse tissue proteomes (soluble fractions of heart, kidney and liver, and the membrane fraction of liver; 50 μM probe, 2 h, phosphate-buffered saline (PBS), pH 7.4; $n = 2$ per tissue). The tandem MS datasets generated by TOP-ABPP were analyzed by the SEQUEST algorithm, specifying a differential modification corresponding to the masses of each probe on nine potentially nucleophilic amino acids. Assignments were screened for peptides uniquely labeled on a single amino acid residue. A very restricted reactivity profile was observed for the CA and UK probes, which selectively labeled cysteine residues in the proteomes (**Fig. 1b**). In contrast, the SE probe demonstrated unique labeling events on several amino acids, including aspartate, glutamate, cysteine, tyrosine and histidine residues (**Fig. 1b**).

We next investigated whether the distinct proteome labeling profiles of the SE, CA and UK probes could be discerned from their reactivities with isolated amino acids in solution. Each probe was exposed to 20 equivalents of amino acid derivatives under buffer conditions that mimic the proteomic environment. Product formation was monitored by LC/MS and revealed similar reactivity profiles for the SE and CA probes (**Fig. 1c**), with cysteine adducts representing the predominant products and minor amounts of the corresponding histidine, lysine and aspartate adducts also being observed. Of particular interest was the relatively low reactivity of the SE probe with carboxylic acid groups in solution (**Fig. 1c**), which contrasted sharply with the large number of aspartate and glutamate labeling sites observed for this probe in proteomes (**Fig. 1b**). In contrast, the UK probe displayed substantial reactivity with cysteine, lysine and histidine groups in solution (**Fig. 1c**) but only showed adducts with cysteine residues on proteins within complex proteomes (**Fig. 1b**). These data thus indicate that electrophilic probes can display unanticipated trends of reactivity in proteomes that are not easily extrapolated from solution studies.

To further explore the disparate reactivity of these probes, we analyzed the labeling profile of aldehyde dehydrogenase-1 (ALDH-1), an enzyme containing both a cysteine nucleophile (Cys303) and catalytic glutamate base (Glu269). Each catalytic residue was mutated to alanine, and the probe reactivity profiles of mutant enzymes were compared to that of wild-type (WT) ALDH-1 following transient transfection in COS-7 cells. Consistent with previous studies[9], the SE probe was found to label both WT and C303A mutant ALDH-1 but not the E269A mutant (**Fig. 2a**). Notably, the cysteine-reactive CA probe showed the opposite profile, labeling both WT and E269A ALDH-1 but not the C303A mutant (**Fig. 2b**). These data thus demonstrate that different classes of carbon electrophiles can display mutually exclusive amino acid reactivity profiles, even when placed within the same enzyme active site.

A survey of the residues labeled by the SE, CA and UK probes revealed a strong enrichment for functional residues that play roles in catalysis, substrate binding and post-translational regulation. Representative examples of these functional residues are listed in **Table 1** (see **Supplementary Table 1** online for an expansive list of functional residues labeled and **Supplementary Tables 2–4** online for complete lists of labeling sites for SE, CA and UK probes, respectively). SE-labeled proteins include the previously reported δ-3,5-δ-2,4-dienoyl-CoA isomerase and the acyl-CoA dehydrogenases, which were labeled on catalytic glutamate and aspartate bases, respectively[9]. In addition, the expansive tissue profiling demonstrated several tyrosine-specific labeling events, including the active site tyrosine of corticosteroid-11 β-dehydrogenase and a tyrosine residue from the dual specificity tyrosine phosphorylation-regulated kinase 1A that is known to be dynamically phosphorylated[10]. The proteome coverage of the CA and UK probes spanned a variety of enzymes with cysteine nucleophiles including fatty acid synthase, UDP-glucose-6-dehydrogenase and multiple nitrilases. Additionally, labeling was observed on several metal-coordinating residues, such as cysteines that bind iron and zinc in aconitate hydratase, betaine-homocysteine S-methyltransferase and alcohol dehydrogenase. Cysteine residues within nucleotide binding domains were also labeled, including GTP- and NAD-binding residues in phosphoenolpyruvate carboxykinase, lactate dehydrogenase and isocitrate dehydrogenase. Notably, more than half of the proteins targeted by the CA and UK probes were exclusively labeled by one of these two agents (**Supplementary Fig. 2** and **Supplementary Table 5** online), which indicates that different cysteine-reactive electrophiles target specific subsets of the proteome. All three probes labeled cysteine residues involved in oxidative regulation; this observation was exemplified by residues in hemoglobin, glyceraldehyde-3-phosphate dehydrogenase and peroxiredoxin, which were modified on known sites of oxidation or S-nitrosylation[11].

The proteome coverage of the panel of carbon electrophiles included several proteins previously inaccessible to the current suite of ABPP probes. Some of these proteins, such as the chloride intracellular channel (CLIC) family, were labeled on conserved

© 2008 Nature Publishing Group http://www.nature.com/naturechemicalbiology

Table 1 Subset of functional residues targeted selectively by the SE, CA and UK probes

IPI number	Description	Peptide	UK	SE	CA	Site of labeling	Function
SE-labeled residues							
IPI00130804	δ-3,5-δ-2,4-dienoyl-CoA isomerase	EVDMGLAAD*VGTLQR	–	755	–	Asp204	Active site proton donor
IPI00119114	Long-chain specific acyl-CoA dehydrogenase	GFYYLMQELPQE*R	–	26	–	Glu291	Active site proton acceptor
IPI00115599	Corticosteroid-11 β-dehydrogenase 1	MTQPMIAPY*SASK	–	4	–	Tyr183	Active site proton acceptor
CA-labeled residues							
IPI00118344	UDP-glucose 6-dehydrogenase	ASVGFGGSC*FQK	–	–	18	Cys276	Active site nucleophile
IPI00119945	Nitrilase 2	VGLGIC*YDMR	–	–	125	Cys153	Active site nucleophile
IPI00135977	Chloride intracellular channel protein 4	AGSDGESIGNC*PFSQR	–	–	74	Cys35	Site of nitrosylation
UK-labeled residues							
IPI00221400	Alcohol dehydrogenase 1	IDGASPLDKVCLIGC*GFSTGYGSAVK	13	–	–	Cys175	Zinc binding
		VIPLFSPQC*GECR	18	–	–	Cys98	Zinc binding
		VIPLFSPQCGEC*R	22	–	–	Cys101	Zinc binding
IPI00229510	L-lactate dehydrogenase B chain	ITVVGVGQVGMAC*AISILGK	45	–	–	Cys36	NAD binding domain
IPI00125135	Ubiquitin-conjugating enzyme E2 D2	IYHPNINSNGSIC*LDILR	29	–	–	Cys85	Active site glycyl thioester intermediate

Asterisks after residues denote probe-labeled residues; numerical values denote total spectral counts for each probe across all proteomic samples. IPI, International Protein Index.

residues implicated in the post-translational regulation of activity[12]. CLIC proteins are ion channels that are able to assume both soluble and membrane-bound forms and are distantly related to the glutathione S-transferase superfamily[12,13]. Our mass spectrometric data revealed that the CA probe labeled CLIC4 at Cys35, a conserved residue among CLIC family members that has been shown to be a site for nitrosylation and oxidative regulation[12]. To test whether the CA probe might serve as a general profiling tool for CLICs, we over-expressed three members of the mouse CLIC family (CLIC1, CLIC4 and CLIC5) with a C-terminal myc/His tag in COS-7 cells. All three CLICs were labeled by the CA probe in transfected cell proteomes (**Fig. 2c**). A C35S mutant of CLIC4 was not labeled by the CA probe (**Fig. 2c**), confirming that reactivity occurred specifically on the conserved Cys35 residue. Pretreatment of cell lysates with oxidizing agents such as nitric oxide or oxidized glutathione resulted in substantial reduction of labeling of the CLICs, which supports the notion that the conserved cysteine residue targeted by the CA probe is susceptible to oxidation (**Fig. 2d** and **Supplementary Fig. 3** online). Notably, the CA probe also proved capable of labeling CLICs in living cells (**Fig. 2c**), which suggests a potential route to monitor the post-translational regulation of these proteins *in vivo*.

In order to expand the number of protein classes addressable by ABPP, structural knowledge of active sites needs to be paired with an understanding of small-molecule reactivity. Here, we show that distinct classes of reactive carbon electrophiles demonstrate widely divergent amino acid preferences in proteomes. The promiscuity of the SE probe designates it as a highly versatile electrophile for ABPP, and potentially for related chemical biology endeavors such as ligand-guided protein surface labeling[14], which aims to convert reversible ligands into covalent probes by proximity-induced reactivity with nucleophilic amino acids neighboring protein active sites. One could envision improving the target selectivity of SE probes by combining this electrophile with high-affinity binding groups for individual proteins of interest. In contrast, the CA and UK probes, by displaying selective reactivity with cysteine residues, constitute powerful electrophiles for enzymes that require this amino acid for function and/or post-translational regulation. Furthermore, the bias toward cysteine reactivity demonstrated by the CA and UK groups suggests a similar reactivity profile for structurally related but less electrophilic groups, such as the acrylamide and α-fluoromethylketone. These latter electrophiles could prove particularly useful for generating target-selective ABPP probes, as has been demonstrated for the epidermal growth factor (EGF) receptor[15] and RSK kinases[4], respectively. More generally, the preponderance of functional residues labeled by carbon electrophiles in proteomes suggests that these sites display sufficiently enhanced nucleophilicity (possibly dictated by the local protein microenvironment) to be experimentally discriminated from the large excess of nonfunctional residues in the proteome. Future ABPP studies incorporating the SE, CA and UK electrophiles, as well as others, into substrate and/or inhibitor scaffolds should facilitate the development of functional proteomics probes for a wide range of proteins.

Note: Supplementary information and chemical compound information is available on the Nature Chemical Biology website.

ACKNOWLEDGMENTS
This work was supported by the US National Institutes of Health (CA087660) and the Skaggs Institute for Chemical Biology.

AUTHOR CONTRIBUTIONS
E.W. and B.F.C. conceived and designed the experiments. E.W. carried out the experiments. E.W., G.M.S. and B.F.C. analyzed the experimental data. The manuscript was written by E.W. and B.F.C.

Published online at http://www.nature.com/naturechemicalbiology
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Evans, M.J. & Cravatt, B.F. *Chem. Rev.* **106**, 3279–3301 (2006).
2. Kidd, D., Liu, Y. & Cravatt, B.F. *Biochemistry* **40**, 4005–4015 (2001).
3. Kato, D. *et al. Nat. Chem. Biol.* **1**, 33–38 (2005).
4. Cohen, M.S., Zhang, C., Shokat, K.M. & Taunton, J. *Science* **308**, 1318–1321 (2005).
5. Patricelli, M.P. *et al. Biochemistry* **46**, 350–358 (2007).
6. Shin, N.Y., Liu, Q., Stamer, S.L. & Liebler, D.C. *Chem. Res. Toxicol.* **20**, 859–867 (2007).
7. Speers, A.E. & Cravatt, B.F. *Chem. Biol.* **11**, 535–546 (2004).
8. Speers, A.E. & Cravatt, B.F. *J. Am. Chem. Soc.* **127**, 10018–10019 (2005).
9. Adam, G.C., Burbaum, J., Kozarich, J.W., Patricelli, M.P. & Cravatt, B.F. *J. Am. Chem. Soc.* **126**, 1363–1368 (2004).
10. Rush, J. *et al. Nat. Biotechnol.* **23**, 94–101 (2005).
11. Forrester, M.T., Foster, M.W. & Stamler, J.S. *J. Biol. Chem.* **282**, 13977–13983 (2007).
12. Littler, D.R. *et al. FEBS J.* **272**, 4996–5007 (2005).
13. Cromer, B.A., Morton, C.J., Board, P.G. & Parker, M.W. *Eur. Biophys. J.* **31**, 356–364 (2002).
14. Chen, G. *et al. J. Am. Chem. Soc.* **125**, 8130–8133 (2003).
15. Fry, D.W. *et al. Proc. Natl. Acad. Sci. USA* **95**, 12022–12027 (1998).

© 2008 Nature Publishing Group http://www.nature.com/naturechemicalbiology