NIH-PA Author Manuscript

NIH Public Access
Author Manuscript
J Am Chem Soc. Author manuscript; available in PMC 2006 August 9.

Published in final edited form as:
J Am Chem Soc. 2005 July 20; 127(28): 10018–10019.

A Tandem Orthogonal Proteolysis Strategy for High-Content Chemical Proteomics

Anna E. Speers and **Benjamin F. Cravatt**
The Skaggs Institute for Chemical Biology and The Departments of Chemistry and Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Proteomics aims to assign molecular and cellular functions to the numerous proteins encoded by eukaryotic and prokaryotic genomes.[1] The daunting size and diversity of the proteome has inspired efforts to enrich specific classes of proteins based on shared functional properties.[2] Within this realm of "targeted proteomics", chemical strategies have proven particularly valuable. For example, chemical probes have been created that label proteins based on their catalytic properties [e.g., activity-based protein profiling (ABPP)[3]] and post-translational modifications (PTMs).[4] These chemical methods offer key insights into protein function, distinguishing, for example, active from inactive enzymes in cells and tissues.[3]

The theoretical information content in chemical proteomic experiments greatly exceeds the actual data procured, due in large part to limitations in existing analytical technologies. Ideally, the identities of all probe-labeled proteins and their sites of modification could be determined in a single experiment. However, the low abundance of probe-modified peptides, coupled with their conjugation to large affinity tags, complicates mass spectrometry (MS) analysis, especially in the background of whole proteome proteolytic digestions.[5] Strategies have been introduced for the enrichment of probe-labeled peptides,[4,6a] but these methods discard the rest of the proteome digest, which makes it difficult to differentiate proteins of high similarity, renders protein assignments less statistically significant, and prohibits molecular analysis of entire protein sequences. We sought to address these issues by designing a tandem orthogonal proteolysis (TOP) strategy for the parallel characterization of probe-labeled proteins and sites of probe modification.

The TOP method was combined with ABPP by exploiting click chemistry (CC) techniques,[7] as outlined in Scheme 1. Following proteome labeling with an alkynyl ABPP probe, CC is used to introduce a biotin tag with a tobacco etch virus protease (TEV)[8] cleavage site. Tagged proteins are then subject to streptavidin enrichment and on-bead trypsin digestion. The supernatant is isolated by filtration, and the probe-labeled peptides are eluted from the beads by incubation with TEV. The trypsin and TEV digests are then analyzed in sequential MudPIT[9] experiments to characterize probe-labeled proteins and site(s) of probe modification, respectively.

For initial experiments, a small library of biotinylated TEV-N_3 tags **1-4** with variable spacer regions (Figure 1A) was synthesized and reacted under CC conditions with a mouse heart proteome pretreated with an alkynyl phenyl-sulfonate ester ABPP probe (PS≡ Scheme 1).[7b]Enoyl-CoA hydratase-1 (ECH1), a target of PS probes, is abundant in this proteome, and its labeling site has been characterized as D204.[6]

Supporting Information Available: Synthesis, experimental, and data analysis protocols; detailed lists of TOP-ABPP data. This material is available free of charge via the Internet at http://pubs.acs.org.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Avidin blotting confirmed that all tags were conjugated to ECH1 (Figure 1A). After proteome enrichment with streptavidin beads and sequential trypsin and TEV digestions, liquid chromatography (LC)-MS/MS revealed strong mass peaks for the probe-modified ECH1 active site peptide for **3** and **4**, but not for **1** or **2** (Figure 1B), suggesting that sufficient distance between the solid support and the bound peptide is required for optimal TEV cleavage. A comparison of TEV versus chemical elution[6a]revealed stronger mass signals for the TEV-cleaved ECH1 peptide (Supporting Information), indicating that removal of the biotin tag enhances peptide ionization.

A large-scale TOP-ABPP analysis of heart proteome was conducted to identify additional targets of PS≡.PS≡-heart reactions were performed in triplicate (4 mg protein, 10 μM probe, 2 h). Following CC with **4**, streptavidin enrichment, trypsinization, and TEV cleavage, tryptic and TEV samples were analyzed by MudPIT,[10] resulting in the identification of 32 PS-labeled proteins[11] (Table 1 and Supporting Information).

PS≡-labeled proteins included, in addition to ECH1, a family of acyl-CoA dehydrogenases (VLCAD, LCAD, MCAD), which were all modified on their Glu catalytic bases. VLCAD was also labeled on a Tyr (Y161) predicted, based on the structure of a related enzyme, to reside within 8 Å of the catalytic Glu.[12] Additional enzymes labeled by PS≡ on active site residues included thiolase and ALDH6, which were modified on Cys nucleophiles. A third set of proteins was identified in which probe labeling occurred on a single peptide, but the functions of the modified residues were not purely catalytic. For example, the labeled D194 of quinine reductase 2 (QR2) is involved in cofactor binding.[13] Two isocitrate dehydrogenases (DHs) were labeled on a Cys residue, which is a site for regulation by glutathione.[14] Finally, a handful of abundant proteins were labeled on multiple peptides (e.g., hemoglobin, GAPDH), suggesting that they were “nonspecific” targets of PS≡ Even for these proteins, however, dominant labeling sites were often observed, and these residues tended to be of functional significance (e.g., a nitric oxide-binding residue in hemoglobins). Collectively, these results indicate that a range of functional residues are targeted by sulfonate ester probes in proteomes.

Of particular interest, PS≡ was also found to specifically label a hypothetical protein (LOC67914). BLAST searches revealed that this protein and its predicted site(s) of labeling (Y246 in particular) are conserved across a host of organisms, from yeast to humans; yet, to date, no members of this class of proteins have been functionally characterized. Considering the number of metabolic enzymes targeted by PS≡ it will be intriguing to determine whether LOC67914 is also an enzyme and, if so, whether PS≡ labeling occurs on a catalytic residue.

All of the aforementioned proteins were also identified in the tryptic digests (Table 1). In contrast, some proteins were only observed in TEV elutions (Figure 2 and Supporting Information). Their absence in tryptic samples suggests that these proteins represent false predictions of Sequest, which can make incorrect protein assignments based on MS^2 spectra of single peptides.[15] Conversely, many proteins were identified in trypsin, but not TEV digests (Figure 2). These proteins likely include unlabeled “background” proteins nonspecifically bound to streptavidin beads,[16] as well as targets whose probe-labeled peptides were not detected by MS. For the latter group, alternative proteases may be used to generate distinct probe-labeled peptides for MS analysis.[15] Overall, these results underscore the value of TOP methods, which provide two independent data sets (whole protein digests and probe-labeled peptides) for the verification of targets of chemical probes.

The TOP strategy exhibits several advantages for chemical proteomic investigations. The trypsin phase, which typically afforded 40-60% sequence coverage of probe-labeled proteins, provides a means to distinguish protein isoforms (e.g., splice variants, PTMs) and to comparatively quantify proteins (e.g., by isotopic labeling[17] or spectral counting[18]).

Conversely, TEV digests identify direct sites of probe labeling, revealing fundamental insights into the molecular basis for specific probe-protein interactions. Perhaps most valuable, however, is the information procured by viewing these complementary data sets in combination. For instance, the detection of proteins in both trypsin and TEV digests facilitates the discovery of unanticipated targets of chemical probes that might otherwise have been discarded as “false assignments” or “background” (e.g., the LOC67914 protein). Another attractive feature of the TOP method is that the cleavable tag can be appended by CC to proteins following probe labeling, thereby avoiding potential negative effects of the tag on cell permeability and/or protein labeling. The versatility of this “tag-free” approach should prove particularly useful for experiments performed in vivo, where discrimination between specific and nonspecific labeling events depends on elucidation of probe modification sites. Such advances offer new avenues for the characterization of proteins and residues of unknown function.

Acknowledgements

Acknowledgment. We thank S. Niessen, J. Yates, J. Hewel, and D. Cociorva for advice on MudPIT. This work was supported by the NCI (CA87660), the Skaggs Institute for Chemical Biology, and a HHMI predoctoral fellowship (A.E.S.).

References

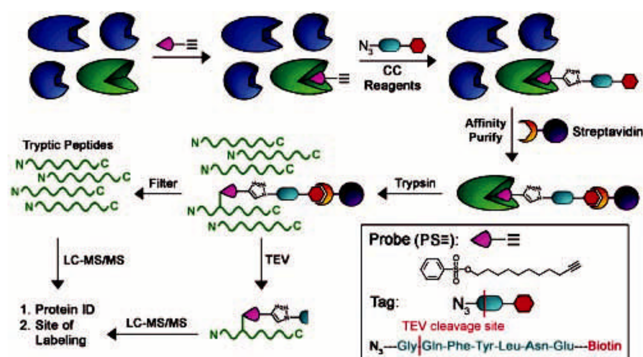
1. Patterson SD, Aebersold R. Nat. Genet 2003;33:311–323. [PubMed: 12610541]
2. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC. Electrophoresis 2000;21:1104–1115. [PubMed: 10786884]
3. (a) Speers AE, Cravatt BF. Chembiochem 2004;5:41–47. [PubMed: 14695510] (b) Liu Y, Patricelli MP, Cravatt BF. Proc. Natl. Acad. Sci. U.S.A 1999;96:14694–14699. [PubMed: 10611275] (c) Jessani N, Liu Y, Humphrey M, Cravatt BF. Proc. Natl. Acad. Sci. U.S.A 2002;99:10335–10340. [PubMed: 12149457]
4. (a) Zhou H, Watts JD, Aebersold R. Nat. Biotechnol 2001;19:375–378. [PubMed: 11283598] (b) Oda Y, Nagasu T, Chait BT. Nat. Biotechnol 2001;19:379–382. [PubMed: 11283599] (c) Vocadlo DJ, Hang HC, Kim EJ, Hanover JA, Bertozzi CR. Proc. Natl. Acad. Sci. U.S.A 2003;100:9116–9121. [PubMed: 12874386] (d) Tai H-C, Khidekel N, Ficarro SB, Peters EC, Hsieh-Wilson LC. J. Am. Chem. Soc 2004;126:10500–10501. [PubMed: 15327282]
5. Hansen KC, Schmitt-Ulms G, Chalkely RJ, Hirsch J, Baldwin MA, Burlingame AL. Mol. Cell. Proteomics 2003;2:299–314. [PubMed: 12766231]
6. (a) Adam GC, Burbaum J, Kozarich JW, Patricelli MP, Cravatt BF. J. Am. Chem. Soc 2004;126:1363–1368. [PubMed: 14759193] (b) Adam GC, Sorensen EJ, Cravatt BF. Nat. Biotechnol 2002;20:805–809. [PubMed: 12091914]
7. (a) Speers AE, Adam GC, Cravatt BF. J. Am. Chem. Soc 2003;125:4686–4687. [PubMed: 12696868] (b) Speers AE, Cravatt BF. Chem. Biol 2004;11:535–546. [PubMed: 15123248]
8. Dougherty WG, Cary SM, Parks TD. Virology 1989;171:356–364. [PubMed: 2669323]
9. Multi-dimensional Protein Identification Technology. Washburn MP, Wolters D, Yates JR. Nat. Biotechnol 2001;19:242–247. [PubMed: 11231557]
10. Sequest (Eng, J. K.; et al. J. Am. Soc. Mass Spectrom. 1994, 5, 976-989) was used to search MS2 spectra against a mouse protein database. For TEV data, a differential modification of 435.3 was assigned to C, D, E, and Y, which were the primary residues modified by PS≡.
11. For proteins to be scored as correct assignments, they needed to be detected in all three tryptic data sets and in at least two of the three TEV data sets (at least four spectra/peptide). See also Supporting Information.
12. Djordjevic S, Pace CP, Stankovich MT, Kim J-J. P. Biochemistry 1995;34:2163–2171.
13. Foster CE, Bianchet MA, Talalay P, Zhao Q, Amzel LM. Biochemistry 1999;38:9881–9886. [PubMed: 10433694]
14. Kil IS, Park JW. J. Biol. Chem 2005;280:10846–10854. [PubMed: 15653693]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

15. MacCoss MJ, Wu CC, Yates JR. Anal. Chem 2002;74:5593–5599. [PubMed: 12433093]
16. Consistent with this premise, many of these proteins were found in control enrichments in which PS≡ was not added to the proteome.
17. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Nat. Biotechnol 1999;17:994–999. [PubMed: 10504701]
18. Liu H, Sadygov RG, Yates JR. Anal. Chem 2004;76:4193–4201. [PubMed: 15253663]JA0532842

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



Scheme 1.
Tandem Orthogonal Proteolysis Strategy for Activity-Based Protein Profiling (TOP-ABPP)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

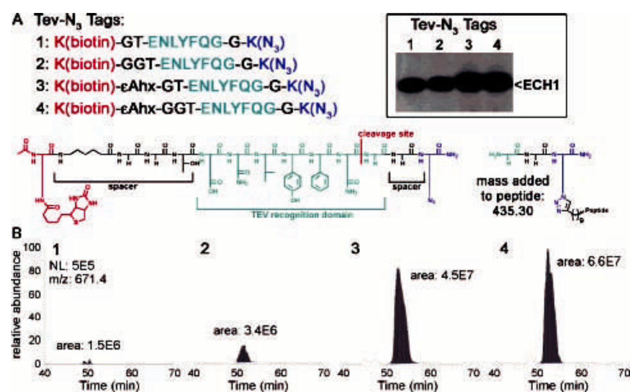


Figure 1.

Comparison of TEV-N_3 tags **1**-**4**. (A) List of tags, full structure of **4**, before and after TEV cleavage. Avidin blot shows successful CC conjugation for all tags to PS≡-labeled ECH1 in heart proteome. (B) Ion extraction of LC trace for probe-labeled ECH1 active site peptide (*m*/*z* 671.4 for +3), showing greater mass signals for longer tags **4** and **5**.

Figure 2.
Proteins identified by TOP-ABPP. Proteins found only in TEV and tryptic digests were regarded as false assignments and background, respectively.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Predicted Sites of Labeling for Select Enzymes[a]

enzyme	labeled peptide	site (s) of labeling	catalytic residue	active site residue	tryp digest % coverage
coverageECH1	K.EVDMGLAAD*VGTLQR.L	D204	yes	yes	69
VLCAD	R.IFE*GANDILR.L	E463	yes	yes	36
	K.ELGAFGLQVPSELGGLGLSNTQ Y*AR.L.	Y161	no	yes	
LCAD	K.GFYYLMQELPQE*R.L.	E291	yes	yes	62
MCAD	K.IYQIY*E*GTAQIQR.L	Y400/E401	yes(E)	yes	42
ALDH6	R.C*MALSTAILVGEAK.K	C317	yes	yes	48
Thiolase	K.DGGQYALVAAC*AAGGQGHAM YVEAYPK.-	C459	yes	yes	42
QR2	K.VLAPQISFGLD*VSSEEER.K	D194	no	yes	29
Isocitrate DH (m)	K.SSGGFVWAC*K.N	C308	no	yes	29
Isocitrate DH (c)	K.SEEGGFIWAC*K.N	C269	no	yes	15
His triad protein	R.ISQAEE*DD*QQLLGHLLLVAK.K	E105/D107	unknown	unknown	31
LOC67914	R.AVLAGIY*NTTE*LVMMQDSSPD FEDTWR.F	Y246/E250	unknown	unknown	13

[a]Predicted sites of probe labeling are indicated with an asterisk (*). In cases where Sequest did not unambiguously distinguish between two potential sites, both are given.