

Labeling Proteins with Small Molecules by Site-Specific Posttranslational Modification

Jun Yin,[†] Fei Liu,[†] Xiaohua Li,[‡] and Christopher T. Walsh^{*†}

Department of Biological Chemistry and Molecular Pharmacology and Institute of Chemistry and Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115

Received April 19, 2004; E-mail: christopher_walsh@hms.harvard.edu

Site-specific labeling of proteins with small synthetic molecules has been an important approach for the elucidation of protein function, mechanism, and interaction networks. For example, incorporation of site-specific fluorescent probes into proteins allows the detection of protein conformational dynamics and the real-time tracking of protein expression, association, and translocation in the living cell.¹ Incorporation of biotin and other small-molecule affinity labels into proteins provides a high-throughput method for protein microarray fabrication and proteomics studies.² Recently, intein-based methods were used to attach a wide range of small molecules including fluorophores, carbohydrates, oligonucleotides, affinity tags, and metal chelators to the C-termini of the target proteins that were expressed as intein fusions. The intein domain was subsequently replaced by a small-molecule cysteine conjugate upon elution from a chitin column.³ Similarly, human *O*⁶-alkylguanine-DNA alkyltransferase (hGAT) has been used for site-specific protein labeling by irreversibly transferring the alkyl group of *O*⁶-benzylguanine derivatives to one of its cysteine residues.⁴ Although these methods have been shown to be capable of labeling proteins with small molecules, the main drawbacks are that the sizes of intein (454 amino acids) and hGAT (207 amino acids) that need to be fused to the target proteins are too large for many applications. Also, the intein-mediated chemical ligation is a relatively slow process requiring overnight incubation and millimolar concentrations of cysteine derivatives.³ On the basis of these observations, we believe that a general and efficient method is still needed for site-specific protein labeling by a variety of small molecules. Such a method would use a peptide tag to direct the specific labeling of a target protein in a complex mixture of cellular proteins. The peptide tag should be as small as possible in size and portable to different proteins in order to be generally useful. Here we report such a method in which target proteins are expressed as fusions to a peptide carrier protein (PCP) excised from a nonribosomal peptide synthetase (NRPS). The Sfp phosphopantetheinyl transferase was used to label PCP site-specifically with small molecule–phosphopantetheinyl (Ppant) conjugate (Figure 1). The PCP domain can be as small as 80 amino acids in length, and the labeling reactions were complete within 30 min in the cell lysate with only micromolar concentrations of small molecule probes. We have also shown that the biotin-labeled PCP fusion proteins in the cell lysate can be directly immobilized on a streptavidin surface for subsequent high-throughput screening.

NRPS PCPs are 8–10 kDa autonomously folded, compact, and stable domains—each offers one specific serine side chain as the nucleophile for covalent phosphorylation with the Ppant moiety of coenzyme A (CoA) (Figure 1).⁵ In *Bacillus subtilis*, the posttranslational modification of PCP is catalyzed by Ppant transferase Sfp⁶ which has been shown to have impressive substrate promiscuity with the thioether or thioester substituents attached to CoA, ranging from

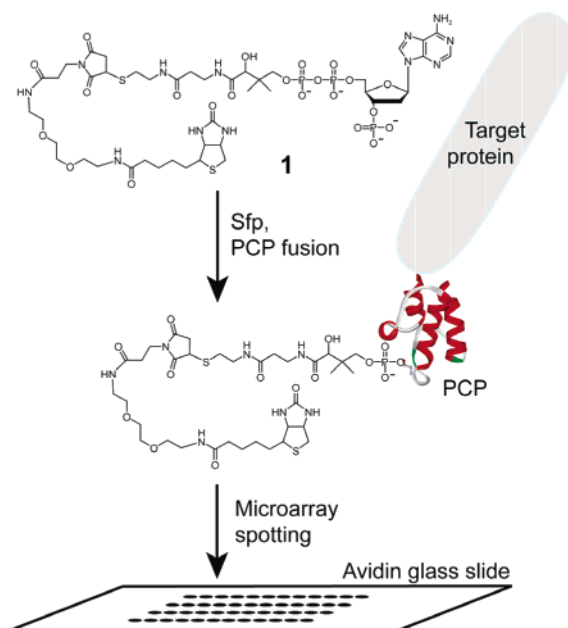


Figure 1. Sfp-catalyzed biotin labeling of PCP fusion proteins and direct spotting of the labeled protein on avidin glass slides.

as small as a proton to as large as decapeptides.⁷ Thus, in principle, many small-molecule ligands with diverse chemical structures can be linked to CoA and then transferred by Sfp to the apo forms of PCPs in a variety of protein fusion contexts. We thus reasoned that a PCP can be an attractive tag for labeling PCP fusion proteins both because of its small size and the high efficiency of the Sfp-catalyzed posttranslational modification of PCP with a wide range of small-molecule CoA conjugates ($k_{cat}/K_m = 1.6 \times 10^4 \text{ s}^{-1} \text{ M}^{-1}$ for CoA loading).^{7a}

To test this idea, fusion proteins were constructed with the PCP domains from NRPS modules EntB (PCP^{EntB}, 98 amino acids)^{8a} and GrsA (PCP^{GrsA}, 80 amino acids)^{8b} fused to the N-termini of the target proteins: enhanced green fluorescent protein (EGFP), glutathione-S-transferase (GST) and maltose binding protein (MBP).⁹ For another set of protein fusions, a 15-amino acid linker with the sequence (Ser-Gly-Gly-Gly-Gly)₃ was introduced between the PCP domain and the target protein and compared to the constructs with no linker inserted. All the fusion proteins were expressed with high yield (20–30 mg/L). The purified fusion proteins were labeled with Ppant biotin in the presence of Sfp and 1,^{7e} and the site-directed labeling of fusion proteins with biotin was confirmed by enzyme-linked immune assays (ELISA) and Western blotting using streptavidin–horseradish peroxidase conjugate as the probe. To test if the protein labeling can proceed in a high-throughput fashion, the fusion proteins were also expressed in 1 mL cultures in the 96-well deep well plates. After overnight induction with IPTG, deoxycholic acid (DCA) and DNase I were added to initiate cell lysis.¹⁰ Ten microliters of cell lysate was transferred from each well

[†] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School.

[‡] Institute of Chemistry and Cell Biology, Harvard University.

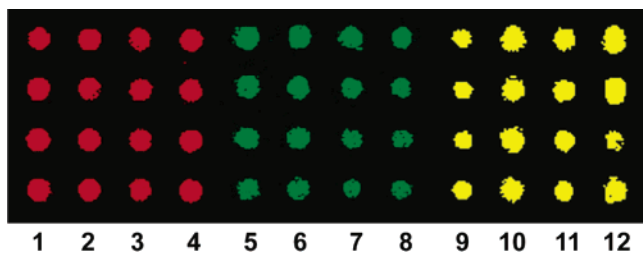


Figure 2. Detection of biotin-labeled GFP (1–4), GST (5–8), and MBP (9–12) fusion proteins on the avidin glass slide by anti-GFP, anti-GST, and anti-MBP antibodies conjugated with fluorophores TRITC, Alexa Fluor 488, and Alexa Fluor 647, respectively. The spots are around 300 μm in diameter, and the lane numbering is the same as in Figure S1(B) in the Supporting Information.

of the deep well plate to 90 μL of the labeling reaction mixture with Sfp and 5 μM of **1** in a standard 96-well plate. Labeling reactions were allowed to proceed for half an hour at 30 $^{\circ}\text{C}$. The Western blotting showed that only the target proteins in the cell lysate were labeled with biotin, suggesting the high specificity of the labeling reactions.⁹ Chasing the reaction mixture with ^3H -acetyl CoA showed that the labeling reaction is more than 95% complete after 30 min, suggesting the high efficiency of the labeling reaction. The labeling-reaction mixtures were then directly spotted on avidin glass slides with a microarray spotter. After washing the glass slides, the immobilized fusion proteins, EGFP, GST, and MBP, were simultaneously detected with a mixture of fluorophore-conjugated antibodies against EGFP, GST, and MBP, respectively. In the control experiment, acetyl CoA was used instead of **1** for PCP labeling, and the reaction mixtures were also spotted on the avidin slide. Only spots corresponding to the biotin-labeled PCP fusion proteins were detected with strong fluorescence intensity, (Figure 2) and the spots corresponding to the acetyl CoA-labeled PCP fusions were not detected. These results suggested that the Ppant linkage is stable for protein labeling and the Sfp-catalyzed small-molecule labeling of the PCP tag is highly specific and efficient, requiring only micromolar concentration of the small molecule. Also the labeling reaction proceeds well in the cell lysate and is well adapted for high-throughput proteomic screening.

To test if the PCP-tagging strategy is amenable to high-throughput enzymatic screening, we constructed N-terminal PCP^{EntB} and PCP^{GrsA} fusion with β -galactosidase and luciferase, respectively.⁹ Expressed fusion proteins showed the same activity as that for the original enzymes. The fusion proteins were also expressed in 1-mL cultures in deep well plates, and after in-well cell lysis by DCA, aliquots of cell lysate were incubated with **1** and Sfp. The immobilization of the biotin-labeled enzyme on the 96-well streptavidin plate can then be assayed by measuring the enzymatic activity retained to each well after wash. Only the wells with biotin-labeled cell lysates showed significant β -galactosidase or luciferase activities, while background activities were found with the wells to which acetyl CoA was added for PCP labeling.⁹ Thus, the PCP tag does not interfere with the function of the target protein, and Sfp-catalyzed biotin labeling can also be applied to high-throughput enzymatic screening.

We did not observe any interference of biotin labeling by the in situ CoA released by cell lysis. Also we found that the PCP fusion proteins expressed from *Escherichia coli* are all in their apo forms without Ppant modification and available for subsequent small-molecule labeling, suggesting that there is no interference from Ppant transferases of the *E. coli* due to their tight substrate specificities.^{6a} The PCP fusions with and without the (Ser-Gly-Gly-Gly)₃ linker showed similar Ppant loading efficiency, antibody binding, and enzymatic activities. This suggested the

versatility of the PCP tag to be compatible with various proteins and that a linker between PCP and the target protein is not necessary. Since in the native context of NRPS, the PCP domain is flanked by both N- and C-terminal domains, presumably PCP can perform equally well as a C-terminal tag. Also, on the basis of the wide substrate tolerance of the functional moiety that attached to CoA for Sfp-catalyzed PCP modification,⁷ we expect that various fluorescent and affinity labels^{7e} besides biotin can be used for PCP labeling. We are currently investigating the possibility of using the PCP tag for labeling proteins in vivo.

In summary we have developed a general strategy for site-specific labeling of proteins with small molecules. The target proteins are expressed as fusions to the PCP tag, and the covalent small-molecule labeling of the PCP tag is catalyzed by Sfp and can be carried out in the cell lysate. The small size of the PCP domain, the portability of PCP to various target proteins, as well as the high efficiency of Sfp-catalyzed PCP posttranslational modification and the compatibility of Sfp with various small-molecule probes presented as CoA derivatives denote the generality of the PCP tag for protein labeling. Furthermore, the use of PCP tag for biotin labeling has been shown to be amenable to high-throughput protein microarray preparation and enzymatic screening.

Acknowledgment. This work was funded by NIH Grants GM-20011 (C.T.W.) and GM66456 (F.L.). We thank Drs. David Vosburg, Hening Lin, and Wei Lu for critical reading of the manuscript.

Supporting Information Available: Results of fusion protein construction, Western blotting, enzymatic assays, and the experimental procedures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) (a) Weiss, S. *Science* **1999**, *283*, 1676–1683. (b) Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 906–918. (c) Miller, L. W.; Sable, J.; Goelet, P.; Sheetz, M. P.; Cornish, V. W. *Angew. Chem., Int. Ed.* **2004**, *43*, 1672–1675.
- (2) (a) Amini, F.; Kodadek, T.; Brown, K. C. *Angew. Chem., Int. Ed.* **2002**, *41*, 356–359. (b) Peluso, P.; Wilson, D. S.; Do, D.; Tran, H.; Venkatasubbiah, M.; Quincy, D.; Heidecker, B.; Poindexter, K.; Tolani, N.; Phelan, M.; Witte, K.; Jung, L. S.; Wagner, P.; Nock, S. *Anal. Biochem.* **2003**, *312*, 113–124. (c) Phizicky, E.; Bastiaens, P. I.; Zhu, H.; Snyder, M.; Fields, S. *Nature* **2003**, *422*, 208–215. (d) Adam, G. C.; Burbaum, J.; Kozarich, J. W.; Patricelli, M. P.; Cravatt, B. F. *J. Am. Chem. Soc.* **2004**, *126*, 1363–1368.
- (3) (a) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705–6710. (b) Tolbert, T. J.; Wong, C. H. *J. Am. Chem. Soc.* **2000**, *122*, 5421–5428. (c) Lesaichere, M. L.; Lue, R. Y.; Chen, G. Y.; Zhu, Q.; Yao, S. Q. *J. Am. Chem. Soc.* **2002**, *124*, 8768–8769. (d) Lovrinovic, M.; Seidel, R.; Wacker, R.; Schroeder, H.; Seitz, O.; Engelhard, M.; Goody, R. S.; Niemeyer, C. M. *Chem. Commun. (Cambridge)* **2003**, 822–823. (e) Lue, R. Y.; Chen, G. Y.; Hu, Y.; Zhu, Q.; Yao, S. Q. *J. Am. Chem. Soc.* **2004**, *126*, 1055–1062. (f) Wood, R. J.; Pascoe, D. D.; Brown, Z. K.; Medicott, E. M.; Kriek, M.; Neylon, C.; Roach, P. L. *Bioconjugate Chem.* **2004**, *15*, 366–372.
- (4) (a) Keppeler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. *Nat. Biotechnol.* **2003**, *21*, 86–89. (b) Kindermann, M.; George, N.; Johnsson, N.; Johnsson, K. *J. Am. Chem. Soc.* **2003**, *125*, 7810–7811.
- (5) (a) Walsh, C. T.; Gehring, A. M.; Weinreb, P. H.; Quadri, L. E.; Flugel, R. S. *Curr. Opin. Chem. Biol.* **1997**, *1*, 309–315. (b) Weber, T.; Baumgartner, R.; Renner, C.; Marahiel, M. A.; Holak, T. A. *Struct. Folding Des.* **2000**, *8*, 407–418.
- (6) (a) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. *Chem. Biol.* **1996**, *3*, 923–936. (b) Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. *Biochemistry* **1998**, *37*, 1585–1595.
- (7) (a) Belshaw, P. J.; Walsh, C. T.; Stachelhaus, T. *Science* **1999**, *284*, 486–489. (b) Sieber, S. A.; Walsh, C. T.; Marahiel, M. A. *J. Am. Chem. Soc.* **2003**, *125*, 10862–10866. (c) Clugston, S. L.; Sieber, S. A.; Marahiel, M. A.; Walsh, C. T. *Biochemistry* **2003**, *42*, 12095–12104. (d) Vitali, F.; Zerbe, K.; Robinson, J. A. *Chem. Commun. (Cambridge)* **2003**, 2718–2719. (e) La Clair, J. J.; Foley, T. L.; Schegg, T. R.; Regan, C. M.; Burkart, M. D. *Chem. Biol.* **2004**, *11*, 195–201.
- (8) (a) Gehring, A. M.; Bradley, K. A.; Walsh, C. T. *Biochemistry* **1997**, *36*, 8495–8503. (b) Stachelhaus, T.; Walsh, C. T. *Biochemistry* **2000**, *39*, 5775–5787.
- (9) See Figures S1, S2, and S3 in the Supporting Information.
- (10) Gildersleeve, J.; Varvak, A.; Atwell, S.; Evans, D.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2003**, *42*, 5971–5973.

JA047749K