Parallel Identification of *O*-GlcNAc-Modified Proteins from Cell Lysates

Hwan-Ching Tai,[†] Nelly Khidekel,[†] Scott B. Ficarro,[‡] Eric C. Peters,[‡] and Linda C. Hsieh-Wilson*,[†]

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, and Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, California 92121

Received April 13, 2004; E-mail: lhw@caltech.edu

Dynamic glycosylation of proteins by *O*-linked β-*N*-acetyl-glucosamine (*O*-GlcNAc) has been increasingly implicated in the regulation of cellular physiology and function.[1] Although discovered more than 20 years ago, an understanding of *O*-GlcNAc as a posttranslational modification has been hampered by the lack of effective tools for its detection and study. Despite recent advances,[2] no method has been reported for the rapid, parallel identification of *O*-GlcNAc proteins from cells. Individual proteins are typically overexpressed or immunoprecipitated to detect the modification.[2b,3] As a result, time-consuming and expensive procedures must be developed for each protein of interest. Even upon isolation, low-abundance regulatory proteins often elude detection due to the limited sensitivity of traditional methods. Here, we report a new approach that permits any protein to be rapidly interrogated for the *O*-GlcNAc modification. Our strategy circumvents the need to purify individual proteins, accommodates any cell type or tissue, and can be extended to the mapping of modification sites. As an illustration of the approach, we identified four new *O*-GlcNAc-glycosylated proteins of low cellular abundance (c-Fos, c-Jun, ATF-1, and CBP) and two new glycosylation sites on the protein *O*-GlcNAc transferase (OGT).

Previously, we described a chemoenzymatic method to tag purified *O*-GlcNAc proteins using an engineered β-1,4-galactosyl-transferase (GalT) and a ketone-containing substrate.[4] We envisioned exploiting this tagging chemistry for the development of a new parallel strategy to identify *O*-GlcNAc-glycosylated proteins from cell lysates. Specifically, the *O*-GlcNAc-modified proteins would be biotinylated and then selectively captured by affinity chromatography (Figure 1). To establish whether a given protein was *O*-GlcNAc glycosylated, one would simply examine whether the protein was captured. Using this approach, multiple proteins could be readily interrogated in parallel by Western blotting using antibodies selective for proteins of interest.

This approach would have several notable advantages. It would accelerate the discovery of *O*-GlcNAc proteins by eliminating the need to purify individual proteins. Virtually any protein could be examined for the modification as a wide variety of antibodies are available for Western blotting. The enhanced sensitivity of our tagging chemistry relative to existing methods would enable identification of even low-abundance regulatory proteins.[4] Moreover, the use of cell lysates rather than intact cells would capture the physiologically relevant glycosylation state of proteins without perturbing metabolic pathways. Finally, the ability to target specific proteins across different tissue or cell types[5] would complement emerging proteomic technologies.[2a]

Implementation of this parallel approach required extension of our tagging chemistry from purified proteins to complex mixtures. HeLa cells were lysed under denaturing conditions to preserve the
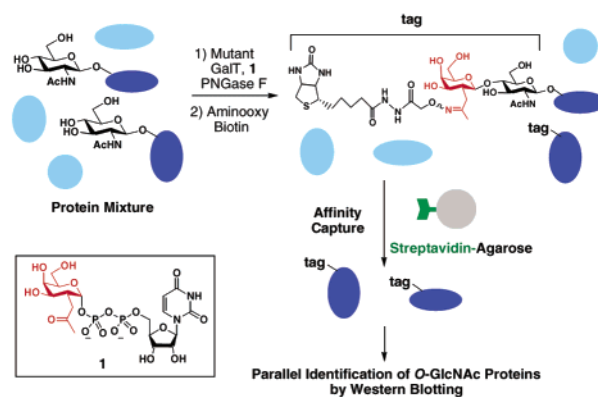

Figure 1. Strategy for identifying *O*-GlcNAc proteins from cell lysates.
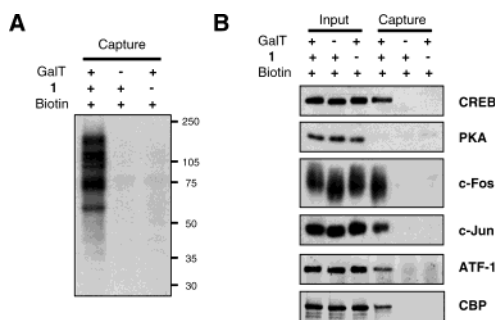

Figure 2. (A) Captured proteins from HeLa cell lysates following labeling as indicated. The blot was probed with streptavidin−HRP to detect biotinylated proteins. (B) Labeled lysates prior to (Input) or following (Capture) affinity capture were probed by Western blotting using antibodies against the indicated proteins.

physiological glycosylation state of the proteins. The cell extract was then labeled with the UDP-ketone analogue **1** and mutant GalT for 12 h at 4 °C. We found that *N*-linked glycans could be removed simultaneously during this incubation period by treatment with PNGase F.[6] Following reaction with an aminooxy biotin, the biotinylated *O*-GlcNAc proteins were captured with streptavidin-agarose beads, resolved by SDS-PAGE, and transferred to nitrocellulose membrane. To determine whether the captured proteins had been biotinylated, the membrane was blotted with streptavidin conjugated to horseradish peroxidase (HRP). A strong chemiluminescence signal was observed, indicating successful labeling of proteins from extracts (Figure 2A). Little signal was detected in the absence of either enzyme or **1**, strongly suggesting that *O*-GlcNAc-modified proteins had been specifically labeled and captured.

To confirm the results, we examined whether the transcription factor cAMP-responsive element binding protein (CREB) was among the captured proteins. CREB is a low-abundance protein that contains only two major *O*-GlcNAc glycosylation sites,[7] and as such, it represents a challenging cellular target. We readily

detected CREB in the captured fraction by Western blotting using an anti-CREB antibody (Figure 2B). In contrast, a protein that lacks *O*-GlcNAc,[8] cAMP-dependent protein kinase (PKA), was not detected. These results demonstrate that low-abundance *O*-GlcNAc proteins from cells can be selectively captured and identified.

We next applied the approach toward the parallel identification of novel proteins. Although the AP-1 transcription factor complex has been shown to be GlcNAc modified,[9] the specific proteins and nature of the glycosidic linkage have remained unresolved. Figure 2 shows that the AP-1 family members c-Fos and c-Jun were captured, indicating that both proteins are *O*-GlcNAc glycosylated. As independent confirmation, we used the traditional approach of UDP-[3H]galactose and GalT,[6] followed by immunoprecipitation of c-Fos. Notably, tritium labeling required 1000 h of exposure to film for strong detection (Supporting Information). In contrast, our strategy permitted detection of c-Fos within minutes.

Importantly, the approach enables study of the *O*-GlcNAc modification across structurally or functionally related protein families. ATF-1, a structural homologue and dimerization partner of CREB,[10] shares only partial sequence identity within the region of CREB glycosylation.[7] Nonetheless, ATF-1 was present in the captured fraction, indicating that both family members are subject to *O*-GlcNAc glycosylation in HeLa cells.

Our strategy also permitted the identification of an entirely new class of *O*-GlcNAc-glycosylated proteins, histone acetyltransferases (HAT). CREB-binding protein (CBP) is a HAT involved in chromatin remodeling and activation of numerous transcription factors.[11] As shown in Figure 2B, we found that CBP is *O*-GlcNAc glycosylated. This finding is interesting in light of recent observations that *O*-GlcNAc transferase (OGT), the enzyme that catalyzes the modification, interacts with a histone deacetylase complex to promote gene silencing.[12] Our results demonstrate that a broader set of transcriptional components are *O*-GlcNAc modified, and they support the notion that *O*-GlcNAc may serve as a general mechanism for transcriptional control.

Finally, we have extended our strategy to the mapping of glycosylation sites. The challenge of identifying specific modification sites has deterred efforts to understand posttranslational modifications, and mass spectrometry enrichment strategies are often required.[13] We reasoned that our tagging chemistry could be applied to the enrichment of *O*-GlcNAc peptides and first demonstrated the approach using CREB. CREB from Sf9 cells was labeled and digested with trypsin. Following avidin chromatography, enrichment of a CREB glycopeptide[7] was observed by MALDI-TOF MS and LC−MS (Figure 3A and Supporting Information). Importantly, the ketone-biotin moiety facilitated the identification of the *O*-GlcNAc peptide by providing a unique fragmentation pattern upon tandem MS. To illustrate the potential of the approach to identify new glycosylation sites, OGT from Sf9 cells was labeled and analyzed as above. Two regions of glycosylation were identified within the catalytic domain of OGT (aa 1037−1046) and the ninth tandem tetratricopeptide repeat (aa 390−406), a highly conserved motif that mediates protein−protein interactions between OGT and its regulatory partners (Figure 3B). The location of these sites within important functional domains suggests that OGT may regulate its own activity via autoglycosylation. Importantly, future extension of this enrichment strategy toward peptides from cell lysates should enable the proteome-wide identification of *O*-GlcNAc-glycosylated proteins.

In conclusion, we have developed a new approach that permits proteins isolated from cell or whole tissue extracts to be rapidly interrogated for the *O*-GlcNAc modification. Our strategy detects
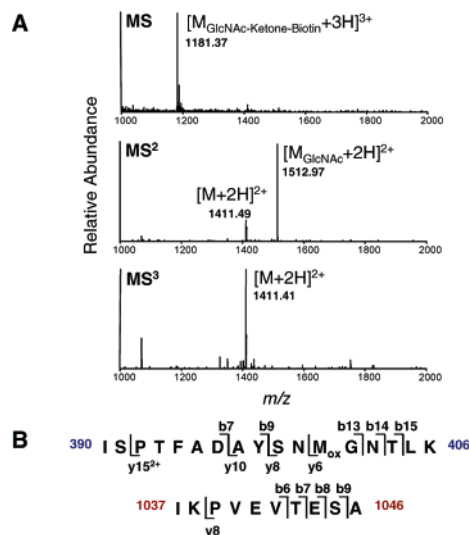

Figure 3. (A) LC−MS[n] signature of the enriched *O*-GlcNAc peptide from CREB. (B) Glycosylated peptides from OGT. Summary of the b and y fragment ions identified by MS[4]. See Supporting Information for full characterization and additional OGT peptides.

low-abundance proteins, circumvents the need to purify individual proteins, and can be extended to the mapping of glycosylation sites. We anticipate that the strategy will accelerate both the discovery of new *O*-GlcNAc-modified proteins and an understanding of the physiological role of this important modification. Finally, the application of similar chemistries may advance the study of other posttranslational modifications.

Supporting Information Available: Experimental procedures and full MS characterization (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Zachara, N. E.; Hart, G. W. *Chem. Rev.* **2002**, *102*, 431−438. (b) Wells, L.; Vosseller, K.; Hart, G. W. *Science* **2001**, *291*, 2376−2378.
(2) (a) Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Mol. Cell. Proteomics* **2002**, *1*, 791−804. (b) Vocadlo, D. J.; Hang, H. C.; Kim, E. J.; Hanover, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9116−9121. (c) Comer, F. I.; Vosseller, K.; Wells, L.; Accavitti, M. A.; Hart, G. W. *Anal. Biochem.* **2001**, *293*, 169−177.
(3) (a) Zhang, F.; Su, K.; Yang, X.; Bowe, D. B.; Paterson, A. J.; Kudlow, J. E. *Cell* **2003**, *115*, 715−725. (b) Chou, T. Y.; Dang, C. V.; Hart, G. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 4417−4421.
(4) Khidekel, N.; Arndt, S.; Lamarre-Vincent, N.; Lippert, A.; Poulin-Kerstien, K. G.; Ramakrishnan, B.; Qasba, P. K.; Hsieh-Wilson, L. C. *J. Am. Chem. Soc.* **2003**, *125*, 16162−16163.
(5) Notably, the strategy has been successfully applied to brain tissue and several mammalian cell lines.
(6) Roquemore, E. P.; Chou, T. Y.; Hart, G. W. *Methods Enzymol.* **1994**, *230*, 443−460.
(7) Lamarre-Vincent, N.; Hsieh-Wilson, L. C. *J. Am. Chem. Soc.* **2003**, *125*, 6612−6613.
(8) Lubas, W. A.; Hanover, J. A. *J. Biol. Chem.* **2000**, *275*, 10983−10988.
(9) Jackson, S. P.; Tjian, R. *Cell* **1988**, *55*, 125−133.
(10) Mayr, B.; Montminy, M. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 599−609.
(11) Vo, N.; Goodman, R. H. *J. Biol. Chem.* **2001**, *276*, 13505−13508.
(12) Yang, X.; Zhang, F.; Kudlow, J. E. *Cell* **2002**, *110*, 69−80.
(13) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198−207.

JA047872B