Moving the O-glycoproteome from form to function

Jacques U. Baenziger[1]

Departments of Pathology and Cell Biology/Physiology, Washington University School of Medicine, St. Louis, MO 63110

Life scientists who study carbohydrate structures are all too accustomed to hearing, "So, what are those things good for anyway?" In PNAS, Schjoldager et al. (1) provide an elegant comeback by using a unique strategy for mapping the location of O-GalNAc–linked carbohydrate structures to show how the addition of GalNAc to specific Ser and Thr residues by one of 20 polypeptide GalNAc-transferases (ppGalNAc-Ts) can modulate the biologic activity of the modified glycoprotein. The consequences can be dramatic, for example, altering cholesterol or triglyceride metabolism and thus exerting a significant effect on human health parameters.

Carbohydrate structures O-glycosidically linked through *N*-acetylgalactosamine (i.e., GalNAc) to Ser or Thr residues were first described on mucins that contain tandem repeats of sequences rich in Ser, Thr, and Pro residues, resulting in glycoproteins with hundreds of O-GalNAc–type structures on a single glycoprotein. Clusters of O-GalNAc structures were then found on other secreted and membrane glycoproteins. With time, glycoproteins with single or limited numbers of O-GalNAc structures were identified, and it became apparent that modification of proteins with O-GalNAc structures is a highly abundant and a complex form of protein glycosylation. In most instances, the O-GalNAc–linked structures are relatively simple, containing a β1,3-linked galactose and one or two negatively charged moieties in the form of sialic acid. The presence of sialic acid on multiple closely spaced O-GalNAc structures has a dramatic effect on the secondary structure of proteins, forcing them to assume a very extended "bottle-brush" conformation. Although simple structures like those shown in Fig. 1 are typical, depending on the repertoire of glycosyltransferases being expressed, a vast array of distinct structures can be synthesized through the sequential addition of individual sugars. Thus, the presence or absence of O-GalNAc structures, the specific location of O-GalNAc structures in the protein sequence, and the structural features of the O-GalNAc structures each have the potential to profoundly affect the biologic properties of glycoproteins bearing this type of modification.

The initial step in mucin-type O-glycosylation, the addition of GalNAc in α-linkage to Ser or Thr, is mediated by a family of UDP-GalNAc: ppGalNAc-Ts.

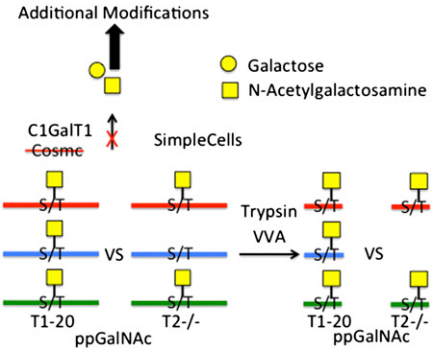

Fig. 1. Identification of O-GalNAc–containing glycoproteins modified by specific ppGalNAcT isoforms. SimpleCells that are devoid of C1GalT1 activity were generated by zinc finger nuclease mutagenesis in liver HepG2 cells by ablating the chaperone Cosmc. Following trypsin digestion, those glycopeptides that have an O-GalNAc were bound by the GalNAc-specific lectin, *Vicia villosa* agglutinin (VVA), and analyzed by MS. O-GalNAc peptides from HepG2 SimpleCells and HepG2 SimpleCells devoid of GalNAc-T2 activity ($T2^{-/-}$) allowed for the identification of proteins and peptides that were the exclusively modified by GalNAc-T2 (blue line) such as ApoC-III.

Twenty homologous genes encoding the ppGalNAc-Ts have been identified in animals (2, 3). The ppGalNAc-Ts have been highly conserved throughout animal evolution and are absent from bacteria, yeast, and plants. In vitro studies have shown that these isoforms of ppGalNAc-Ts have distinct specificities (4). Tissue expression of the individual ppGalNAc-Ts ranges from highly restricted and regulated to ubiquitous. The portrait that emerges is one of highly specific O-GalNAc addition that can be modulated in response to demands occurring during development and metabolism. Genetic evidence supports such roles for O-GalNAc addition. Deficiencies of individual ppGalNAc-Ts have yielded only subtle phenotypes, suggesting that many of the enzymes are redundant and/or that interrogation for the impact of specific modifications will need to be more precisely focused. For example, genome-wide association studies have linked the *GALNT2* gene with altered levels of serum triglyceride and HDL cholesterol. Serum levels of HDL cholesterol responded to increased and decreased expression of GalNAc-T2 in mouse liver, supporting a role for GalNAc-T2 in regulating HDL cholesterol levels in vivo (5–7).

A major obstacle to understanding the biologic significance of O-GalNAc–linked structures has been the inability to map the sites that are modified with these structures. This is particularly critical for mucin-like structures because no simple consensus sequence predicts the location of this modification on glycoproteins. The authors of the current paper recently developed an elegant and effective method to map the O-glycoproteome (8) and have now used the approach to probe the functions of the ppGalNAc-T *galnt2* isoform. The authors have taken advantage of the fact that a single enzyme, C1GalT1, accounts for synthesis of Core 1 O-linked structures by adding Gal in β1,3-linkage to the O-GalNAc originally placed by a ppGalNAc-T isoform (Fig. 1). An X-linked gene, *C1GalT1C1*, produces the chaperone Cosmc that is required for the production of active C1GalT1 (9). By using zinc-finger nuclease gene targeting to inactivate the single functional copy of *C1GalT1C1*, cultured cells deficient in C1GalT1, referred to as SimpleCell lines, can be efficiently generated. Because the Gal cannot be added to the O-GalNAc, all the sites that would have been modified with more complex structures derived from Core 1 are modified with only a GalNAc or GalNAc modified with sialic acid. Removal of the sialic acid and digestion with trypsin yields peptides containing one or more GalNAc residues that can be isolated by using the lectin *Vicia villosa* agglutinin, which is highly specific for α-linked GalNAc. By using electrospray ionization and MS, the sequence and location of the GalNAc moieties on each of these peptides can be determined. The initial study from this research group (8) identified more than 100 glycoproteins with >350 O-GalNAc sites, the vast majority of which were unique.

Schjoldager et al. (1) have now produced SimpleCells from the liver parenchymal HepG2 cell line, HepG2-SC, and identified 45 additional O-GalNAc–modified glycoproteins that had not been identified in their previous study using three different cell lines. Mutating GalNAc-T2 in the HepG2-SC cells allowed for the identification of O-GalNAc sites on glycoproteins that are exclusively modified by GalNAc-T2. Among O-GalNAc–modified glycoproteins linked to HDL or

Author contributions: J.U.B. wrote the paper.

The author declares no conflict of interest.

See companion article on page 9893.

[1]E-mail: baenziger@wustl.edu.

triglyceride metabolism, ApoC-III stood out because it is exclusively glycosylated by GalNAc-T2. Expression of GalNAc-T2 in HepG2-T2$^{-/-}$ cells restored the glycosylation of ApoC-III. Another strength of the strategy used is that expression in WT cells vs. SimpleCells makes it possible to determine whether the entire O-linked structure or just the O-GalNAc is required to restore function. In many instances, it is possible to confirm the specificity of ppGalNAcTs by using in vitro assays and peptide acceptors containing the sequence of interest.

One limitation of the strategy used is that some O-GalNAc–modified glycoproteins may not be sufficiently abundant to be identified as targets for GalNAc addition by using SimpleCells. For example, proteolytic processing via proprotein convertase cleavage at RXXR processing sites is prevented if specific nearby Ser or Thr residues are modified with O-GalNAc (10). Addition of O-GalNAc to the lipase inhibitor protein ANGPTL3 by GalNAc-T2 prevents processing and activation by proprotein convertase. The low abundance of ANGPTL3 in HepG2 cells prevented its identification as a target for GalNAc addition; however, immunoprecipitation from the medium demonstrated that ANGPTL3 was processed in the absence but not the presence of GalNAc-T2. Thus, not all the sites that are modified with O-GalNAc will be immediately revealed by using this strategy; however, if there is reason to believe that O-GalNAc is playing an important role, attention can be focused on the specific protein, as was done with ANGPTL3.

Genome-wide association studies are beginning to reveal unexpected relationships between glycosyltransferases and diseases.

The strategy that has been developed by Schjoldager et al. (1) to map the GalNAc-type O-glycoproteome and to relate this emerging map to the 20 members of the ppGalNAcT family of transferases represents a major advance. Before this approach, sites of O-GalNAc modification were identified as a byproduct of analyzing a glycoprotein or some aspect of its function. Genome-wide association studies are beginning to reveal unexpected relationships between glycosyltransferases and diseases. In the case of ppGalNAcTs, it will now be possible to identify candidate glycoproteins for further analysis as has been done in this study (1). Strategies such as this are moving the field of glycobiology beyond structural studies and mapping to enhancing our understanding of the functional significance of these complex structures and their relevance to human health.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grants R01-CA21923 and R01- HD058474.

1. Schjoldager KT-BG, et al. (2012) Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proc Natl Acad Sci USA* 109:9893–9898.
2. Bennett EP, et al. (2012) Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* 22:736–756.
3. Raman J, Guan Y, Perrine CL, Gerken TA, Tabak LA (2012) UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferases: Completion of the family tree. *Glycobiology* 22:768–777.
4. Gill DJ, Clausen H, Bard F (2011) Location, location, location: New insights into O-GalNAc protein glycosylation. *Trends Cell Biol* 21:149–158.
5. Willer CJ, et al. (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 40:161–169.
6. Teslovich TM, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466:707–713.
7. Kathiresan S, et al. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 40:189–197.
8. Steentoft C, et al. (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* 8:977–982.
9. Ju T, Cummings RD (2002) A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase. *Proc Natl Acad Sci USA* 99:16613–16618.
10. Gram Schjoldager KT, et al. (2011) A systematic study of site-specific GalNAc-type O-glycosylation modulating proprotein convertase processing. *J Biol Chem* 286:40122–40132.