

A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation

Howard C. Hang^{*†}, Chong Yu^{*†}, Darryl L. Kato^{*†}, and Carolyn R. Bertozzi^{*†‡§¶}

^{*}Center for New Directions in Organic Synthesis, Departments of [†]Chemistry and [‡]Molecular and Cell Biology, and [§]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720-1460

Edited by Kai Simons, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany, and approved September 11, 2003 (received for review August 13, 2003)

Mucin-type O-linked glycoproteins are involved in a variety of biological interactions in higher eukaryotes. The biosynthesis of these glycoproteins is initiated by a family of polypeptide *N*-acetyl- α -galactosaminyltransferases (ppGalNAcTs) that modify proteins in the secretory pathway. The lack of a defined consensus sequence for the ppGalNAcTs makes the prediction of mucin-type O-linked glycosylation difficult based on primary sequence alone. Herein we present a method for labeling mucin-type O-linked glycoproteins with a unique chemical tag, the azide, which permits their selective covalent modification from complex cell lysates. From a panel of synthetic derivatives, we identified an azido GalNAc analog (*N*-azidoacetyl-galactosamine, GalNAz) that is metabolized by numerous cell types and installed on mucin-type O-linked glycoproteins by the ppGalNAcTs. The azide serves as a bioorthogonal chemical handle for selective modification with biochemical or biophysical probes using the Staudinger ligation. The approach was validated by labeling a recombinant glycoprotein that is known to possess O-linked glycans with GalNAz. In addition, GalNAz efficiently labeled mucin-type O-linked glycoproteins expressed at endogenous levels. The ability to label mucin-type O-linked glycoproteins with chemical tags should facilitate their identification by proteomic strategies.

Posttranslational modification of proteins is recognized as a major point of diversification that distinguishes the proteomes of higher organisms from their more simple evolutionary ancestors (1). Glycosylation is the most complex form of posttranslational modification and is known to regulate many aspects of protein function (2, 3). Moreover, glycans can possess complex biological functions independent of their protein scaffold (4). Consequently, there is increased attention to system-wide analyses of glycan structure and function (1, 5–8).

The two major forms of glycosylation that transpire in the secretory pathway are N-linked (attached to Asn residues) (9, 10) and O-linked (attached to Ser or Thr residues) (11). N-linked glycans are installed on target proteins by a single enzyme, oligosaccharyl transferase, that recognizes the consensus sequence Asn-Xaa-Ser/Thr (Xaa is any residue but proline) (12). This feature makes N-linked glycosylation amenable to prediction based on primary sequence analysis and disruption by site-directed mutagenesis. The predominant form of O-linked glycosylation is the mucin-type, characterized by an initial *N*-acetylgalactosamine (GalNAc) residue α -linked to the hydroxyl groups of Thr or Ser side chains (Fig. 1*A*) (13). Mucin-type O-linked glycosylation is initiated in the Golgi compartment by a family of polypeptide *N*-acetyl- α -galactosaminyltransferases (ppGalNAcTs) (14). Multiple isoforms exist in higher eukaryotes (≈ 24 in human, ≈ 23 in mouse, 15 in *Drosophila*, and 9 in *Caenorhabditis elegans*), and many of the family members have overlapping tissue distribution and substrate specificity. There is no defined consensus sequence for the ppGalNAcTs, rendering predictions of O-linked glycosylation difficult (15). Semiempirical computational methods allow prediction of domains with clustered O-linked glycans (i.e., mucin domains), but specific sites of O-glycosylation can only be identified with 70% accuracy (16).

Recently, proteomic strategies have been described for the identification of N-linked protein glycosylation. The use of periodate oxidation/hydrazide capture (17) or lectin affinity chromatography (18, 19) enabled the enrichment of N-linked glycopeptides from complex mixtures. Both of these approaches exploited the availability of peptide *N*-glycosidase F (pNGase-F) (20), which specifically cleaves N-linked glycans, thereby liberating polypeptides for proteomic analysis.

By contrast, the identification of proteins bearing mucin-type O-linked glycans remains elusive due to the complex nature of their biosynthesis and the lack of tools comparable to those used to study N-linked glycosylation. Although enzymatic liberation of O-linked glycans can be achieved with a combination of several glycosidases, there is no counterpart to pNGase-F for O-linked glycoproteins. Chemical removal of O-linked glycans can be achieved under alkaline conditions (21), but base-catalyzed β -elimination is not specific to mucin-type O-linked glycans and also cleaves other posttranslational modifications [i.e., phosphates (22, 23), β -O-GlcNAc (24), and thiols (25)]. In addition, the harsh conditions used for β -elimination can damage polypeptide backbones (26). Finally, the only commonality among mucin-type O-linked glycans is their peptide proximal α -GalNAc residue. This residue is elaborated to generate a diversity of structures (13) for which there is no single specific lectin (27). Therefore, the use of lectins to enrich O-linked glycoproteins requires extensive prior digestion with glycosidases and is not generally applicable. Consequently, new approaches are needed for the proteomic analysis of mucin-type O-linked glycosylation.

Herein, we report a strategy for labeling mucin-type O-linked glycoproteins with a bioorthogonal chemical tag that exploits their one commonality: the conserved core GalNAc residue. The strategy involves feeding cells an azido GalNAc analog that is metabolically incorporated into mucin-type O-linked glycoproteins. The glycoproteins bearing azides can then be selectively labeled with phosphine probes via the Staudinger ligation (28). From a panel of azido GalNAc analogs, we demonstrate that *N*-azidoacetyl-galactosamine (GalNAz) is the optimal substrate for metabolic incorporation of azides into mucin-type O-linked glycans. Glycoproteins expressed both recombinantly and at endogenous levels in mammalian cells were labeled with GalNAz, enabling their detection from complex cellular lysates with phosphine probes (Fig. 1*B*). The strategy provides a means to distinguish proteins bearing O-glycans from the remainder of the proteome.

Methods and Materials

Chemical Synthesis. Details for the chemical synthesis of the azido GalNAc analogs can be found in *Supporting Materials and Methods*

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: GalNAc, *N*-acetylgalactosamine; ppGalNAcTs, polypeptide *N*-acetyl- α -galactosaminyltransferases; pNGase-F, peptide *N*-glycosidase F; GalNAz, *N*-azidoacetyl-galactosamine; 2AzGal, 2-azido-2-deoxygalactose; CHO, Chinese hamster ovary; Gal, galactose.

[¶]To whom correspondence should be addressed. E-mail: bertozzi@cchem.berkeley.edu.

© 2003 by The National Academy of Sciences of the USA

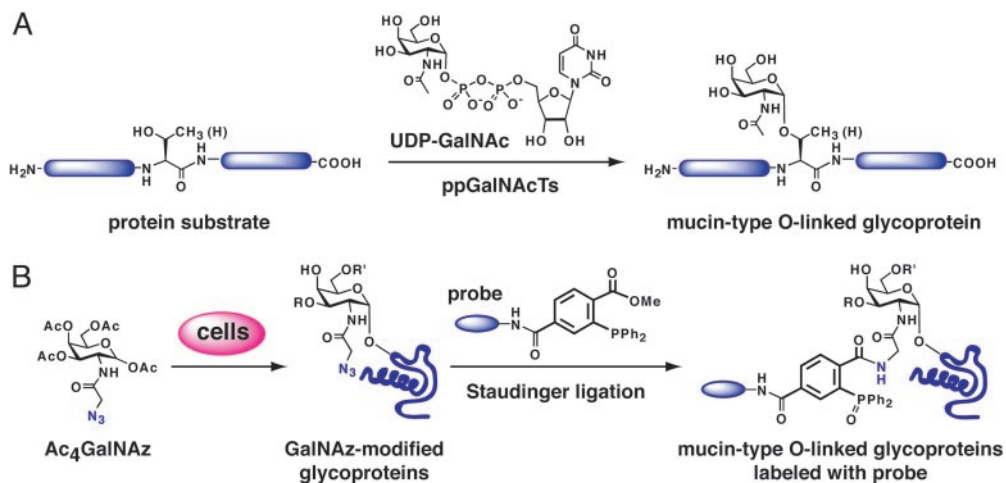


Fig. 1. (A) Initiation of mucin-type O-linked glycan biosynthesis by the ppGalNAcTs. UDP-GalNAc is the nucleotide sugar donor substrate for the ppGalNAcT family. (B) Strategy for metabolic labeling of mucin-type O-linked glycoproteins with an azido GalNAc analog (GalNAz) for proteomic analysis with phosphine probes. This strategy capitalizes on the highly chemoselective Staudinger ligation reaction between azides and phosphines. R and R' are oligosaccharide elaborations from the core α -GalNAc residue.

and Scheme 1, which are published as supporting information on the PNAS web site.

Cell Culture Conditions. Jurkat cells were grown in RPMI medium 1640 supplemented with 10% FCS, 100 units/ml penicillin, and 0.1 mg/ml streptomycin. Chinese hamster ovary K1 (CHO.K1), *ldld*, and *pgsA* cells were grown in Ham's F-12 medium supplemented with 10% FCS. HeLa, NIH 3T3 fibroblasts, and COS-7 cells were grown in DMEM supplemented with 10% FCS. In all cases, cells were incubated in a 5% CO₂ humidified incubator at 37°C.

Analysis of Cell Surface Azides by Flow Cytometry. Cells were seeded at 125,000 cells per well (determined by Coulter cell counter) in 12-well polystyrene tissue culture plates in 1 ml of media. Acetylated azidosugars were administered to the cells by adding the appropriate volume of a 5 mM stock solution of the compound (dissolved in EtOH and filter sterilized) to the wells, followed by evaporation of the EtOH before addition of media and cells. After 2–3 d, cells were harvested and subjected to Staudinger ligation and flow cytometry analysis. Cells were washed once with 1 ml of PBS buffer (PBS, pH 7.4/0.1% FCS/0.1% NaN₃ wt/vol) and reacted with 100 μ l of phosphine-FLAG (0.5 mM in PBS, pH 7.4) (29) for 1 h at room temperature. After Staudinger ligation, cells were washed twice with PBS buffer and incubated with 200 μ l of mouse anti-FLAG monoclonal antibody (5 μ g/ml in PBS buffer, Sigma) for 30 min at room temperature. The cells were rinsed twice with PBS buffer and incubated with FITC-labeled goat anti-mouse secondary antibody (5.0 μ g/ml in PBS buffer) in the dark for 30 min at 4°C. Adherent cells were washed once with PBS buffer and trypsinized with 100 μ l of Trypsin-EDTA (PBS, pH 7.4) for 5 min at room temperature. Trypsinization was stopped by the addition of 400 μ l of 10% FCS serum-rich media, and the cells were transferred into 96-well v-bottom plates and centrifuged at 3,000 \times g for 3 min at 4°C. Pelleted cells were washed twice with 200 μ l of PBS buffer and resuspended in 300 μ l of PBS buffer for flow cytometry analysis using a FacsCaliber instrument (BD Instruments). Jurkat cells grown in suspension were washed by centrifugation and resuspended in PBS buffer; treatment with trypsin was omitted.

Competitive Metabolism of GalNAz with GalNAc, GlcNAc, and Galactose (Gal). Cells were treated with 50 μ M Ac₄GalNAz and the appropriate volume of GalNAc, Gal, or GlcNAc from 50 mM stock solutions to give final concentrations of Gal, GlcNAc, or GalNAc

between 0.5 and 5.0 mM in a total volume of 1 ml. Cells were harvested and subjected to flow cytometry analysis of cell surface azides as described above.

Western Blot Analysis of HexNAz Incorporation into Glycoproteins. Cells were grown on 100 \times 20 mm polystyrene tissue culture plates seeded at a density of $\approx 1.25 \times 10^6$ cells per plate in 10 ml of media (cell quantification was by Coulter counter). Cells were supplemented with acetylated azidosugar by adding the appropriate volume of a 5 mM stock solution (dissolved in EtOH and filter sterilized) to the plates, followed by evaporation of the EtOH before addition of media and cells. After 2–3 d, cells were washed twice with PBS and removed from the plates with a cell scraper. Cells were pelleted by centrifugation at 3,000 \times g for 10 min and the supernatant was discarded. The cell pellet was resuspended in 1 ml of homogenization buffer (1% SDS/100 mM Tris-HCl, pH 7.4) and lysed with a cell homogenizer with four to six bursts at 4°C. Homogenized lysates were incubated at 4°C for 1 h to further solubilize proteins and then centrifuged for 10 min at 3,000 \times g to pellet insoluble debris. The supernatant (total soluble protein fraction) was collected, and the remaining pellet was discarded. Total protein concentration was determined by dextran-coated charcoal (DCC) assay (Bio-Rad) to be ≈ 5 mg/ml. Staudinger ligation was performed by incubation of 20 μ l of the lysate (5 mg/ml protein) with 2 μ l of phosphine-FLAG (5 mM in PBS, pH 7.4) for 6 h at 37°C. An aliquot of the reaction mixture was then incubated with an equal volume of protein loading buffer and run on a SDS/PAGE (10% gel). Proteins were transferred to a nitrocellulose membrane and the membrane was blocked with 5% nonfat dried milk in TBST (50 mM Tris-HCl/150 mM NaCl/0.05% Tween 20, pH 7.4) overnight at 4°C. The membrane was washed with TBST (three times, 25 ml) and incubated with α -FLAG-HRP (Sigma A-8592, diluted 1:3,000 in TBST) for 1 h. The membrane was washed with TBST (three times, 25 ml) and developed by using SuperSignal West Pico Chemiluminescent Substrate (Pierce).

Lectin Staining of Cells. *ldld* cells were incubated with acetylated sugars as described above and stained with FITC-labeled lectins (EY Laboratories): HPA (1 μ g/ml) or jacalin (5 μ g/ml) dissolved in PBS (pH 7.4) in the dark for 1 h at 4°C. Cells were washed and trypsinized as described above before analysis by flow cytometry.

Incorporation of GalNAz into Recombinant GlyCAM-Ig. COS-7 cells in 100 \times 20 mm tissue culture plates ($\approx 85\%$ confluent) were trans-

ected with cDNA encoding GlyCAM-Ig fusion (10 μ g) by using Lipofectamine Plus (Invitrogen) reagent in Opti-MEM as recommended by the manufacturer. After 5 h, the cells were trypsinized and replated on new 100 \times 20 mm tissue culture plates in 10 ml of Opti-MEM with or without 50 μ M Ac₄GalNAz. Conditioned media from transfections were harvested after 4 d and centrifuged for 10 min at 3,000 \times g to remove any cell debris. The supernatant was incubated with 500 μ l of Protein A Sepharose 4B beads (Zymed) in 100 mM Tris-HCl, pH 8.0, for 14 h. The GlyCAM-Ig was eluted with 900 μ l of 100 mM Glycine, pH 3.5, and neutralized with 100 μ l of 1 M Tris-HCl, pH 8. The concentration of GlyCAM-Ig was determined by DCC assay to be \approx 100 μ g/ml. Staudinger ligation was performed by reacting 50 μ l of GlyCAM-Ig samples with 5 μ l of phosphine-FLAG (5 mM in PBS, pH 7.4) for 6 h at 37°C. Reactions were quenched with 20 μ l of standard protein loading buffer and run on a SDS/PAGE (10% gel). Western blot analysis of GlyCAM-Ig samples with α -FLAG-HRP or human α -Ig-HRP (The Jackson Laboratory, diluted 1:10,000 in TBST) were performed as described above. Treatment of GlyCAM-Ig and fetuin samples with pNGase-F (Calbiochem) was performed according to the manufacturer's instructions.

Results and Discussion

GalNAz Is a Substrate for the GalNAc Salvage Pathway in Cells. The common substrate of the ppGalNAcTs is UDP-GalNAc, which is derived primarily from UDP-GlcNAc via UDP-Glc/GlcNAc C₄-epimerase but can also be generated from GalNAc via the salvage pathway (Fig. 2) (30). UDP-GlcNAc, in turn, can be obtained from glucose via *de novo* hexosamine biosynthesis or from free GlcNAc via the salvage pathway. Subsequent transport into the Golgi lumen makes UDP-GalNAc available to the ppGalNAcTs. The characterization of a CHO cell line deficient in UDP-Glc/GlcNAc C₄-epimerase, termed *ldlD* (31), suggested that the salvage pathway can be an efficient vehicle for delivery of exogenous GalNAc into O-linked glycans.

For our strategy, we sought to exploit the salvage pathway for delivery of azido GalNAc analogs into mucin-type O-linked glycoproteins. This requires that the analogs act as substrates for GalNAc 1-kinase, UDP-GalNAc pyrophosphorylase, the Golgi UDP-GalNAc transporter, and finally, the ppGalNAcTs (Fig. 2). To identify an optimal azido GalNAc analog, we synthesized three compounds: 1,3,4,6-tetra-*O*-acetyl-*N*-azidoacetyl-galactosamine (Ac₄GalNAz), 1,3,4,6-tetra-*O*-acetyl-2-azido-2-deoxygalactose (Ac₄2AzGal), and 1,3,4-tri-*O*-acetyl-6-azido-6-deoxy-*N*-acetyl-galactosamine (Ac₃6AzGalNAc) (Fig. 3A and Scheme 1). The azidosugars were administered to CHO cells in acetylated form based on previous work showing that acetylated sugars have superior cellular uptake compared with free sugars and are rapidly deacetylated in the cytosol by nonspecific esterases (32). The metabolic incorporation of azidosugars into cell surface glycoproteins was evaluated by using an established flow cytometry assay (33). In this assay, cells are grown in the presence of the azidosugar for 2 days, reacted with a phosphine-FLAG (DYKDDDDK) probe, stained with FITC-labeled α -FLAG monoclonal antibody and analyzed by flow cytometry.

As shown in Fig. 3B, CHO cells treated with Ac₄GalNAz demonstrated \approx 30-fold higher fluorescence than untreated cells. In contrast, Ac₄2AzGal- and Ac₃6AzGalNAc-treated cells did not exhibit fluorescence above the level of untreated cells. These results suggest that 2AzGal and 6AzGalNAc are not effective substrates for the GalNAc salvage pathway, whereas GalNAz is used as a substrate. Ac₄GalNAz-treated cells showed a dose-dependent increase in the number of cell-surface azides, with saturation occurring at \approx 50 μ M (Fig. 3C). To establish that GalNAz is engaging the GalNAc salvage pathway, monosaccharide competition experiments were performed (Fig. 3D). Gal did not suppress the cell surface azide signal produced by Ac₄GalNAz (50 μ M), whereas GalNAc completely inhibited the expression of cell surface azides

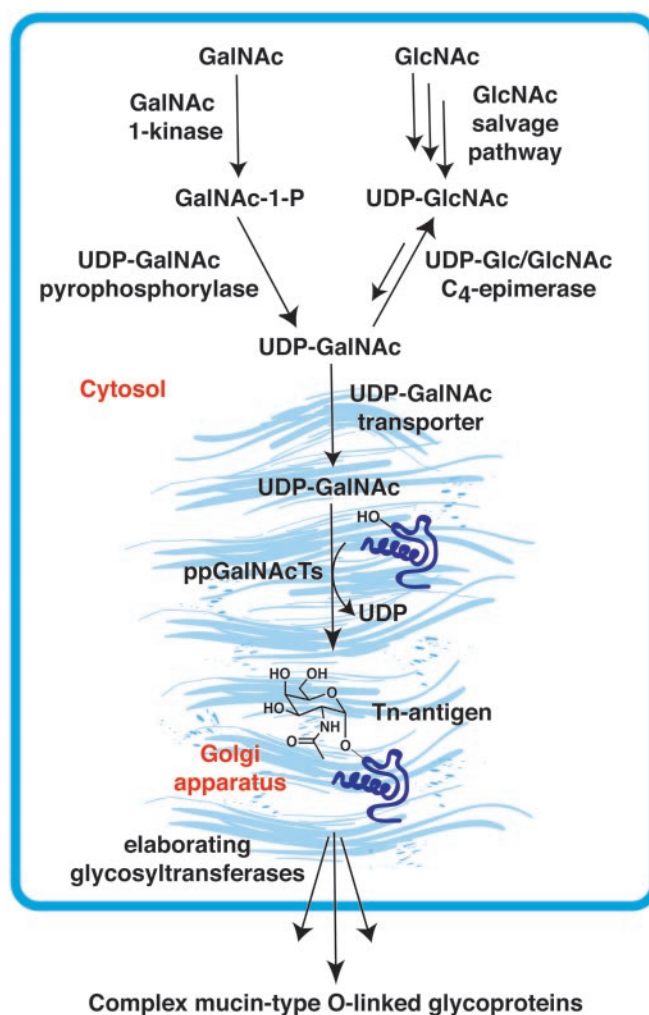


Fig. 2. Biosynthesis of mucin-type O-linked glycoproteins. UDP-GalNAc is produced endogenously from UDP-GlcNAc, which can be generated from GlcNAc via a salvage pathway. Alternatively, UDP-GalNAc can be generated from GalNAc by the action of GalNAc 1-kinase and UDP-GalNAc pyrophosphorylase enzymes of the salvage pathway. Transport of UDP-GalNAc into the Golgi lumen provides the nucleotide sugar donor for the ppGalNAcTs, which modify Ser or Thr residues with α -GalNAc. Further elaboration of this "Tn-antigen" (α -GalNAc-Thr/Ser) by downstream glycosyltransferases generates more complex mucin-type O-linked glycans.

at a concentration of 5 mM. GlcNAc partially suppressed the GalNAz-induced signal, as expected because of metabolic flux through the GlcNAc salvage pathway and conversion to UDP-GalNAz by UDP-Glc/GlcNAc C₄-epimerase (Fig. 2). These results suggest that GalNAz is converted into UDP-GalNAz through the GalNAc salvage pathway in CHO cells. The observations are also consistent with previous experiments using a ketone-isotere of GalNAc, which demonstrated that the salvage pathway enzymes are tolerant structural perturbations of the 2-acetamido position of GalNAc (34).

UDP-GalNAz Is Not Converted to UDP-GlcNAz in Cells. The reaction catalyzed by UDP-Glc/GlcNAc C₄-epimerase is reversible (35, 36), which raises the possibility that UDP-GalNAz generated via the salvage pathway could be converted into UDP-GlcNAz. Subsequent transport of UDP-GlcNAz into the Golgi compartment and incorporation of GlcNAz into glycoconjugates could contribute to the observed cell surface signal. We have previously shown that Ac₄GlcNAz can be converted by cells to UDP-GlcNAz and incor-

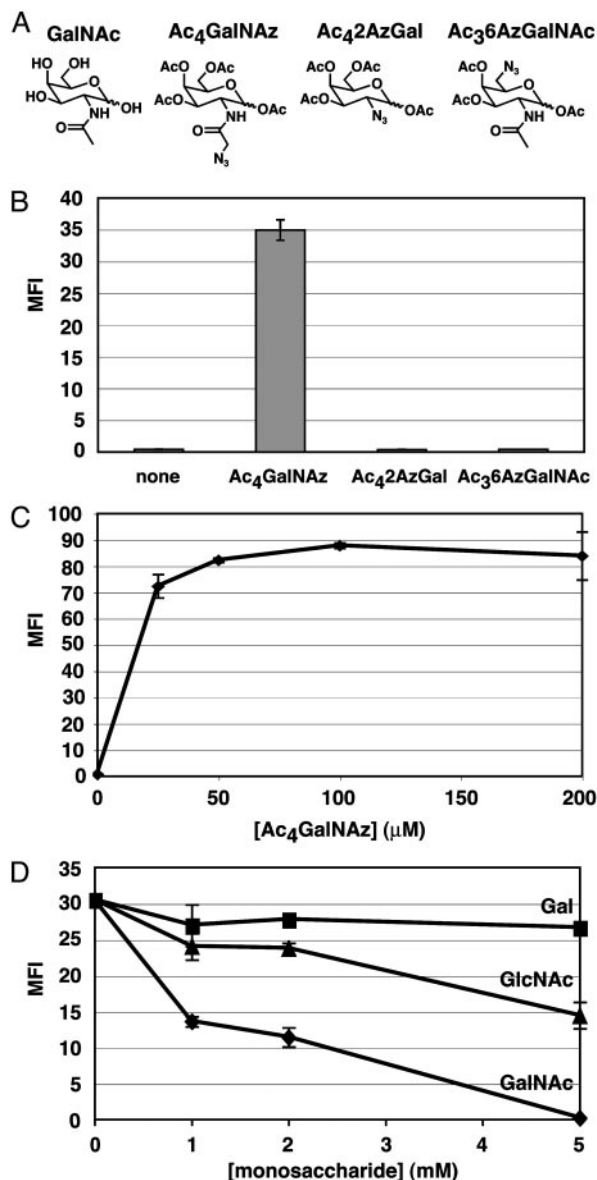


Fig. 3. (A) GalNAc and acetylated azido GalNAc analogs: Ac₄GalNAz, Ac₄2AzGal, and Ac₃6AzGalNAz. (B) Evaluation of acetylated azido GalNAc analogs (50 μM) for metabolic incorporation into CHO cell surface glycoproteins. Cells were incubated with the compounds, stained with phosphine-FLAG followed by a FITC-α-FLAG antibody, and analyzed by flow cytometry. (C) Dose-dependent incorporation of GalNAz into CHO cell surface glycoproteins. Cells were analyzed for cell surface azides as in B. (D) Competitive metabolism of 50 μM Ac₄GalNAz and various concentrations of Gal, GlcNAc, or GalNAc. Cells were analyzed for cell surface azides as in B. MFI, mean fluorescence intensity in arbitrary units. Data points represent the average of duplicate experiments. Bars indicate high and low values.

porated into β-O-GlcNAc-modified proteins in mammalian cells (37). To address the possible interconversion of UDP-GalNAz and UDP-GlcNAz, we incubated CHO cells with 50 μM Ac₄GlcNAz or Ac₄GalNAz and analyzed them by flow cytometry. As shown in Fig. 4A, CHO cells incubated with Ac₄GlcNAz showed significantly lower levels of cell surface azides than cells treated with Ac₄GalNAz. This dramatic difference suggests that UDP-GalNAz and UDP-GlcNAz are not interconverted to a significant extent relative to their direct incorporation into glycoconjugates. Based on previous studies of unnatural sialic acid biosynthesis, it is likely that the low level of cell surface azides produced by GlcNAz reflects

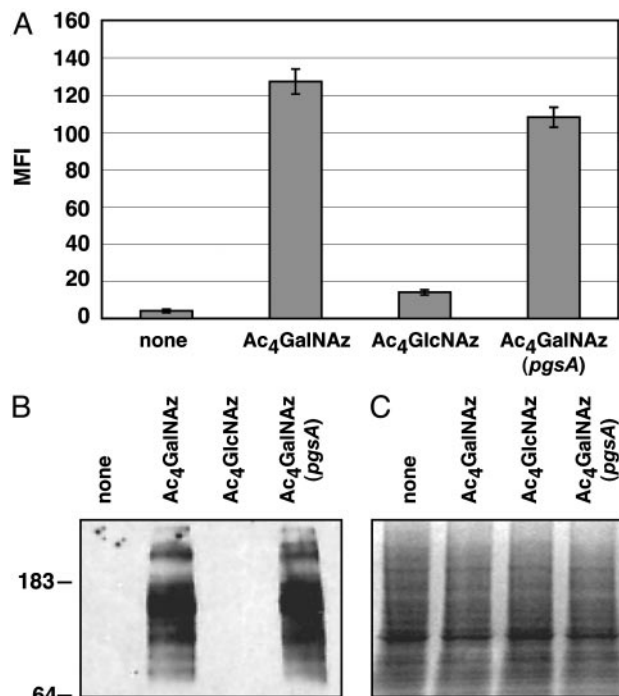


Fig. 4. Evaluation of the metabolic fates of Ac₄GalNAz and Ac₄GlcNAz in CHO cells. (A) Cells were incubated with 50 μM Ac₄GalNAz or Ac₄GlcNAz and analyzed by flow cytometry as in Fig. 3. MFI, mean fluorescence intensity in arbitrary units. Data points represent the average of duplicate experiments. Bars indicate high and low values. (B) α-FLAG Western blot analysis of cell lysates. After longer exposure times, faint labeling of glycoproteins is also observed from Ac₄GlcNAz-treated cells. (C) Coomassie staining of the same gel as in B, demonstrating comparable levels of protein loading (~50 μg per lane).

from its conversion to the corresponding sialic acid analog (33). Furthermore, the differential incorporation of GlcNAz and GalNAz into cellular glycoproteins was confirmed by Western blot analysis of cell lysates. Glycoproteins from cells incubated with Ac₄GalNAz were efficiently labeled with phosphine-FLAG, whereas glycoproteins from Ac₄GlcNAz-treated cells were not (Fig. 4B and C).

GalNAz Resides at the Core Positions of Mucin-Type O-Linked Glycoproteins in CHO Cells. In CHO cells, glycoconjugate-bound GalNAc residues are found in mucin-type O-linked glycoproteins and chondroitin sulfate (CS) proteoglycans. To establish the distribution of GalNAz among these glycoconjugate types, we used a mutant CHO cell line, termed *pgsA*, which is deficient in xylosyltransferase (XylT) activity responsible for initiating glycosaminoglycan biosynthesis (38). As shown in Fig. 4A, *pgsA* cells treated with Ac₄GalNAz showed levels of cell surface azides comparable to wild-type CHO cells, as measured by flow cytometry. Similar levels of azide-modified glycoproteins were also observed by Western blot (Fig. 4B and C). Thus, CHO cells incapable of generating CS (*pgsA*) can still incorporate GalNAz into cell surface glycoproteins at wild-type levels, suggesting that GalNAz resides primarily in mucin-type O-linked glycoproteins.

To demonstrate that GalNAz is incorporated at the core position of mucin-type O-linked glycoproteins, we took advantage of the *lldD* mutant CHO cell line that is deficient in UDP-Glc/GlcNAc C₄-epimerase activity (32). *lldD* CHO cells allow the specific reconstitution of N- and/or O-linked glycan biosynthesis by exogenous addition of Gal and/or GalNAc, respectively, under low serum conditions. Restoration of O-linked glycan biosynthesis can be monitored by staining the cells with FITC-labeled *helix pomatia*

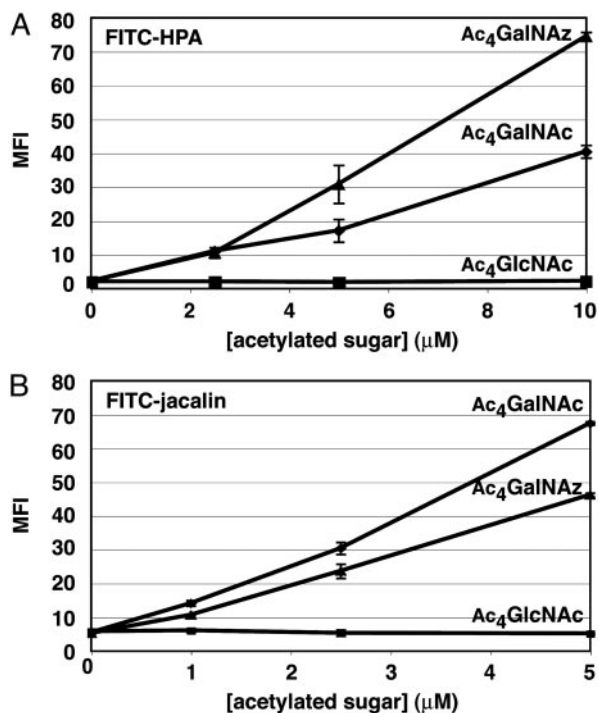


Fig. 5. Incorporation of GalNAz into the core positions of mucin-type O-linked glycoproteins on *ldlD* CHO cells. (A) Detection of Tn-antigen by FITC-HPA staining and flow cytometry analysis. (B) Detection of core 1 O-linked glycans by FITC-jacalin staining and flow cytometry analysis. MFI, mean fluorescence intensity in arbitrary units. Data points represent the average of duplicate experiments. Bars indicate high and low values.

agglutinin (HPA) (39), a lectin specific for the Tn-antigen (GalNAc- α -Thr/Ser, Fig. 2), or with FITC-labeled jacalin (40), a lectin that binds elaborated core 1 O-linked glycans (Sia α 2-3Gal β 1-3GalNAc- α -Thr/Ser). As shown in Fig. 5, GalNAz can reconstitute both HPA and jacalin staining in a dose-dependent manner, similar to GalNAc. As a negative control, GlcNAc failed to restore O-linked glycan biosynthesis (Fig. 5). The molecular basis of the different levels of HPA and jacalin binding to GalNAc- and GalNAz-treated *ldlD* cells is unknown. It is possible that the lectins bind GalNAz-modified O-glycans and natural O-glycans with different affinities. Alternatively, GalNAz and GalNAc may be metabolized and elaborated at different levels. Nonetheless, these results demonstrate that GalNAz is efficiently incorporated into the core position of mucin-type O-linked glycans in cells and can be elaborated by downstream glycosyltransferases to generate more complex epitopes.

Consistent with these cellular data suggesting that GalNAz has access to the core positions of O-linked glycans, we have shown that synthetic UDP-GalNAz is a substrate for several members of the ppGalNAcT family (1–5, 7, 10, 11) *in vitro* at approximately one-third the efficiency of UDP-GalNAc under saturating conditions (41). Thus, we expect labeling of mucin-type O-linked glycoproteins with GalNAz to be independent of the ppGalNAcT isoforms expressed by the cell. To determine the generality of GalNAz metabolism, we treated a variety of mammalian cell lines with Ac₄GalNAz and analyzed their glycoproteins by flow cytometry and Western blot. As shown in Fig. 6, all mammalian cell lines tested demonstrated GalNAz-dependent glycoprotein labeling.

The efficiency of incorporation of GalNAz into O-linked glycoproteins in the presence of endogenous UDP-GalNAc is critical for proteomic analysis. Unfortunately, quantitative analysis of GalNAz-labeled O-linked glycans cannot be performed by using conventional approaches, which require reducing agents and pro-

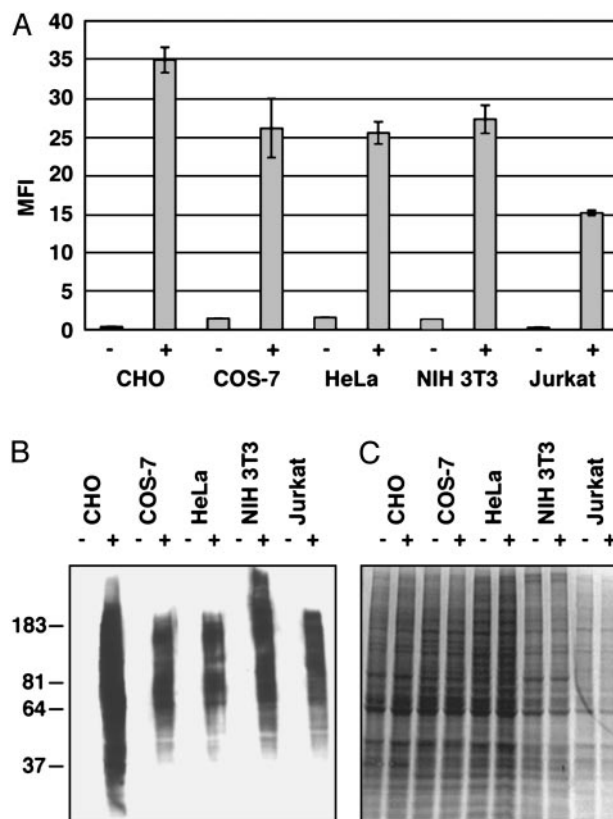


Fig. 6. Metabolic incorporation of GalNAz into glycoproteins of various mammalian cell lines. (A) Flow cytometry analysis of cell surface azides with (+) or without (-) 50 μ M Ac₄GalNAz. (B) α -FLAG Western blot analysis of total glycoprotein in cell lysates. (C) Coomassie staining of identical cell lysates indicating comparable levels of protein loading (\approx 50 μ g per lane). COS-7, green monkey kidney cells; HeLa, human cervical epithelial tumor cells; NIH 3T3, human fibroblasts; Jurkat, human T cell lymphoma; MFI, mean fluorescence intensity in arbitrary units. Data points represent the average of duplicate experiments. Bars indicate high and low values.

longed exposure to harsh acidic/basic conditions that are not compatible with azides (6, 21). This precluded the direct comparison of GalNAc and GalNAz levels in glycoprotein samples. Nonetheless, one can estimate the relative levels of the unnatural and natural sugars by using the data presented above. The concentrations of acetylated GalNAz and GalNAc required to reconstitute O-linked glycan biosynthesis in *ldlD* cells were similar (Fig. 5), suggesting that their fluxes through the GalNAc salvage pathway were also comparable. The final step in the incorporation process involves the ppGalNAcTs, which as mentioned above, use UDP-GalNAz with one-third the efficiency of UDP-GalNAc *in vitro* (41). Collectively, these data suggest that GalNAz replaces an upper limit of one-third of the natural GalNAc residues in cells that make endogenous UDP-GalNAc.

Incorporation of GalNAz into a Recombinant Mucin-Type O-Linked Glycoprotein. To further validate the approach, we analyzed a specific O-linked glycoprotein, GlyCAM-1, for metabolic incorporation of GalNAz. GlyCAM-1 serves as a ligand for the leukocyte homing receptor L-selectin and is known to be modified with clusters of O-linked glycans (42). We expressed GlyCAM-1 as an N-terminal Ig fusion (43) in COS-7 cells treated with or without Ac₄GalNAz. Immunoprecipitation followed by Western blot analysis showed GalNAz-dependent labeling of GlyCAM-Ig with phosphine-FLAG (Fig. 7A). To rule out the possibility that GalNAz was incorporated into the N-linked glycan of the Ig domain,

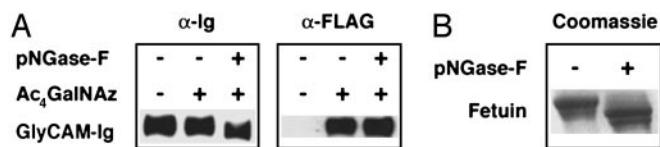


Fig. 7. Metabolic incorporation of GalNAz into a secreted recombinant mucin-type O-linked glycoprotein, GlyCAM-Ig. (A) GlyCAM-Ig from transient transfection of COS-7 cells with (+) or without (-) 50 μ M Ac₄GalNAz was purified with protein A-agarose beads and subjected to Staudinger ligation with phosphine-FLAG. Deglycosylation of N-linked glycans on GlyCAM-Ig samples (2 μ g per lane) with pNGase-F was performed before Staudinger ligation with phosphine-FLAG labeling. (B) Fetuin (20 μ g per lane) was used a positive control for pNGase-F activity as indicated by a shift in molecular weight.

GlyCAM-Ig was treated with pNGase-F. No significant loss in azide labeling was observed for GlyCAM-Ig after N-linked glycan cleavage, suggesting the GalNAz-dependent signal was not derived from the N-linked glycan (Fig. 7A).

Conclusion

In summary, we demonstrate that GalNAz, an azido analog of GalNAc, is metabolized by mammalian cells and incorporated into the core positions of both recombinant and endogenous mucin-type O-linked glycoproteins. Although these glycoproteins are a major reservoir of glycoconjugate-bound GalNAc residues, other glycoconjugate subtypes possess GalNAc as well. For example, certain

glycosphingolipids (GSLs) (44) and chondroitin sulfate proteoglycans (45) may also be labeled with GalNAz, but can be readily separated from glycoproteins based on their disparate physical properties (46). GalNAc residues can also occupy terminal positions on N- and O-linked glycoproteins in neuromuscular junctions (47) and cytotoxic T cells (48) as well as N-linked glycans on pituitary hormones (49). However, these are highly cell-type specific and, in most cell types, mucin-type O-linked glycans should be the primary targets of GalNAz metabolism.

We exploited the azide as a bioorthogonal functional group that can be specifically reacted with phosphine probes by the Staudinger ligation. This chemical strategy targets the conserved core GalNAc-residue of mucin-type O-linked glycoproteins, which allows specific labeling in a manner that is independent of the complex terminal glycan structures generated from various cell types. Thus, the generality of this method should enable enrichment of mucin-type O-linked glycoproteins for proteomic analysis, complementary to existing methods for N-linked glycoproteins.

We thank Drs. Jeff Esko for *pgsA* cells, Monty Krieger for *ldID* cells, and Steven Rosen for GlyCAM-Ig cDNA. The Center for New Directions in Organic Synthesis is supported by Bristol-Myers Squibb as sponsoring member and Novartis as supporting member. H.C.H. acknowledges Dupont Pharmaceuticals and the Organic Division of the American Chemical Society for a graduate fellowship. C.Y. acknowledges the Beckman Scholars Program for an undergraduate fellowship. This research was supported by National Institutes of Health Grant GM66047 (to C.R.B.).

- Mann, M. & Jensen, O. N. (2003) *Nat. Biotechnol.* **21**, 255–261.
- Bill, R. M., Revers, L. & Wilson, I. A. (1998) *Protein Glycosylation* (Kluwer Academic, Boston).
- Rudd, P. M. & Dwek, R. A. (1997) *Crit. Rev. Biochem. Mol. Biol.* **32**, 1–100.
- Lowe, J. B. & Marth, J. D. (2003) *Annu. Rev. Biochem.* **72**, 643–691.
- Hirabayashi, J. (2003) *Trends Biotechnol.* **21**, 141–143.
- Dell, A. & Morris, H. R. (2001) *Science* **291**, 2351–2356.
- Harvey, D. J. (2001) *Proteomics* **1**, 311–328.
- Cooper, C. A., Gasteiger, E. & Packer, N. H. (2001) *Proteomics* **1**, 340–349.
- Kornfeld, R. & Kornfeld, S. (1985) *Annu. Rev. Biochem.* **54**, 631–664.
- Helenius, A. & Aebi, M. (2001) *Science* **291**, 2364–2369.
- Van den Steen, P., Rudd, P. M., Dwek, R. A. & Opdenakker, G. (1998) *Crit. Rev. Biochem. Mol. Biol.* **33**, 151–208.
- Dempski, R. E., Jr., & Imperiali, B. (2002) *Curr. Opin. Chem. Biol.* **6**, 844–850.
- Hanisch, F. G. (2001) *Biol. Chem.* **382**, 143–149.
- Ten Hagen, K. G., Fritz, T. A. & Tabak, L. A. (2003) *Glycobiology* **13**, 1–16.
- Jung, E., Veuthey, A. L., Gasteiger, E. & Bairoch, A. (2001) *Proteomics* **1**, 262–268.
- Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. & Brunak, S. (1998) *Glycoconj. J.* **15**, 115–130.
- Zhang, H., Li, X. J., Martin, D. B. & Aebersold, R. (2003) *Nat. Biotechnol.* **21**, 660–666.
- Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N. & Isobe, T. (2003) *Nat. Biotechnol.* **21**, 667–672.
- Xiong, L. & Regnier, F. E. (2002) *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **782**, 405–418.
- Plummer, T. H., Jr., Elder, J. H., Alexander, S., Phelan, A. W. & Tarentino, A. L. (1984) *J. Biol. Chem.* **259**, 10700–10704.
- Piller, F. & Piller, V. (1993) in *Glycobiology: A Practical Approach*, eds. Fukuda, M. & Kobata, A. (Oxford Univ. Press, Oxford), pp. 291–326.
- Knight, Z. A., Schilling, B., Row, R. H., Henski, D. M., Gibson, B. W. & Shokat, K. M. (2003) *Nat. Biotechnol.* **21**, 1047–1054.
- Jaffe, H., Veeranna & Pant, H. C. (1998) *Biochemistry* **37**, 16211–16224.
- Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J. & Hart, G. W. (2002) *Mol. Cell. Proteomics* **1**, 791–804.
- Herbert, B., Hopwood, F., Oxley, D., McCarthy, J., Laver, M., Grinyer, J., Goodall, A., Williams, K., Castagna, A. & Righetti, P. G. (2003) *Proteomics* **3**, 826–831.
- Greis, K. D., Hayes, B. K., Comer, F. I., Kirk, M., Barnes, S., Lowary, T. L. & Hart, G. W. (1996) *Anal. Biochem.* **234**, 38–49.
- Lis, H. & Sharon, N. (1998) *Chem. Rev.* **98**, 637–674.
- Saxon, E. & Bertozzi, C. R. (2000) *Science* **287**, 2007–2010.
- Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 19–24.
- Freeze, H. H. (2000) in *Carbohydrates in Chemistry and Biology*, eds. Ernst, B., Hart, G. W. & Sinay, P. (Wiley, New York), Vol. 3, pp. 3–14.
- Kingsley, D. M., Kozarsky, K. F., Hobbie, L. & Krieger, M. (1986) *Cell* **44**, 749–759.
- Jacobs, C. L., Yarema, K. J., Mahal, L. K., Nauman, D. A., Charters, N. W. & Bertozzi, C. R. (2000) *Methods Enzymol.* **327**, 260–275.
- Saxon, E., Luchansky, S. J., Hang, H. C., Yu, C., Lee, S. C. & Bertozzi, C. R. (2002) *J. Am. Chem. Soc.* **124**, 14893–14902.
- Hang, H. C. & Bertozzi, C. R. (2001) *J. Am. Chem. Soc.* **123**, 1242–1243.
- Piller, F., Hanlon, M. H. & Hill, R. L. (1983) *J. Biol. Chem.* **258**, 10774–10778.
- Thoden, J. B., Wohlers, T. M., Friedovich-Keil, J. L. & Holden, H. M. (2001) *J. Biol. Chem.* **276**, 15131–15136.
- Vocadlo, D. J., Hang, H. C., Kim, E. J., Hanover, J. A. & Bertozzi, C. R. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 9116–9121.
- Esko, J. D., Stewart, T. E. & Taylor, W. H. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3197–3201.
- Hammarstrom, S., Murphy, L. A., Goldstein, I. J. & Etzler, M. E. (1977) *Biochemistry* **16**, 2750–2755.
- Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. & Vijayan, M. (1996) *Nat. Struct. Biol.* **3**, 596–603.
- Hang, H. C., Yu, C., Pratt, M. R. & Bertozzi, C. R. (2004) *J. Am. Chem. Soc.*, in press.
- Dowbenko, D., Andalibi, A., Young, P. E., Lusic, A. J. & Lasky, L. A. (1993) *J. Biol. Chem.* **268**, 4525–4529.
- Tangemann, K., Bistrup, A., Hemmerich, S. & Rosen, S. D. (1999) *J. Exp. Med.* **190**, 935–942.
- Hakomori, S. (2003) *Curr. Opin. Hematol.* **10**, 16–24.
- Mizuguchi, S., Uyama, T., Kitagawa, H., Nomura, K. H., Dejima, K., Gengyo-Ando, K., Mitani, S., Sugahara, K. & Nomura, K. (2003) *Nature* **423**, 443–448.
- Manzi, A. E., Norgard-Sumnicht, K., Argade, S., Marth, J. D., van Halbeek, H. & Varki, A. (2000) *Glycobiology* **10**, 669–689.
- Martin, P. T., Scott, L. J., Porter, B. E. & Sanes, J. R. (1999) *Mol. Cell. Neurosci.* **13**, 105–118.
- Smith, P. L. & Lowe, J. B. (1994) *J. Biol. Chem.* **269**, 15162–15171.
- Manzella, S. M., Hooper, L. V. & Baenziger, J. U. (1996) *J. Biol. Chem.* **271**, 12117–12120.