Cell

Chemical Technologies for Probing Glycans

Jennifer A. Prescher[1] and Carolyn R. Bertozzi[1,*]
[1]Departments of Chemistry and Molecular and Cell Biology and Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA
*Contact: crb@berkeley.edu
DOI 10.1016/j.cell.2006.08.017

Glycans are central to many biological processes, but efforts to define their functions at the molecular level have been frustrated by a lack of suitable technologies. Here we highlight chemical tools that are beginning to address this need.

Like proteins and nucleic acids, glycans are vital biopolymers found in organisms across all domains of life. In eukaryotes, glycans decorate cell-surface proteins and secreted proteins, where they are poised to mediate molecular recognition events (Varki et al., 1999). Glycans serve as points of attachment for viruses, bacteria, and other cells and participate in many facets of the vertebrate immune system. Inside the cell, they can direct protein trafficking and serve as regulatory switches for protein function. Specific changes in glycan profiles correlate with certain disease states such as cancer and inflammation, suggesting that glycans could be used in clinical diagnostics and perhaps as targets for developing therapeutics.

These observations should stimulate an explosion of interest in glycobiology. But many researchers still express frustration when glycans are implicated at the nexus of their system of study. One fundamental problem is that glycans have complex, branched structures and are intrinsically heterogeneous. Thus, the vast majority of glycoproteins, which are estimated to comprise 50% of eukaryotic proteomes, have not been well-characterized at a molecular level. In cases where the structural details of protein-associated glycans are defined, their functions are still mostly unknown. Our current view of glycobiology therefore remains largely descriptive and focused at the cellular, rather than the molecular, level.

In other areas of biology, the gap between descriptive analysis and mechanistic inquiry has been bridged by experimental tools for perturbing, visualizing, and profiling biopolymers. For example, molecular biology techniques, such as gene disruption, overexpression, and silencing, allow the perturbation of specific proteins and their subsequent functional assignment. Fluorescent proteins such as GFP can serve as genetically encoded tags for protein visualization. GFP fusion proteins have been used in numerous studies of protein localization, dynamics, and association in both cells and living organisms. Finally, DNA microarray technologies have driven the field of genomics, enabling the system-wide profiling of transcripts associated with physiological states.

Unfortunately, these powerful technologies cannot be directly extended to the study of glycans. Glycans are not primary gene products, making their genetic manipulation far more complicated than for proteins or nucleic acids. Disruption of genes encoding glycan biosynthetic enzymes can effect changes in glycan structure and, in some studies, the phenotypes of these mutants have been instructive. But often, redundancies among related glycosyltransferases or embryonic lethality render mutant phenotypes impossible to interpret (Lowe and Marth, 2003). Furthermore, glycans are not amenable to being genetically tagged for visualization in living systems, nor can they be amplified for profiling. Lectins (receptor proteins that bind to glycans) and antibodies can be used to probe glycosylation at a global level, but these reagents cannot reveal the intricacies of glycan structures, their points of attachment to proteins, or more subtle alterations in glycan profiles that accompany physiological changes. Mechanistic studies of glycan function will benefit tremendously from tools for perturbing, visualizing, and profiling glycans in a more refined manner. Chemical tools have much to contribute in this regard and are the subject of this Minireview.

Perturbing Glycan Structures with Small Molecules

Small molecule inhibitors enable temporally controlled perturbation of a system in a reversible and tunable manner. For studies of glycans, the most obvious targets for small molecule inhibition are glycosyltransferases, the enzymes that assemble glycans in the secretory pathway using nucleotide-sugar building blocks (Varki et al., 1999). Glycosyltransferase inhibitors have been identified from the pharmacopeia of natural products, most notably the *Streptomyces* metabolite tunicamycin. This uridine analog blocks the biosynthesis of N-linked glycans by disrupting the assembly of their common pentasaccharide core (Figure 1A). Tunicamycin has revealed numerous aspects of N-glycan function, including pivotal roles in protein folding and trafficking. However, global disruption of N-glycosylation causes a variety of effects that typically masks the function of a specific glycoprotein or glycan.

Recently, global inhibitors of mucin-type glycosylation, the most common form of O-linked glycosylation, have been developed via substrate-based design. Mucin-type O-linked glycosylation (Figure 1B) is initiated by the polypeptide *N*-acetylgalactosaminyltransferases (ppGalNAcTs), a family of enzymes (~24 isoforms in humans) that use UDP-GalNAc as a common substrate. A synthetic compound library of uridine derivatives was screened

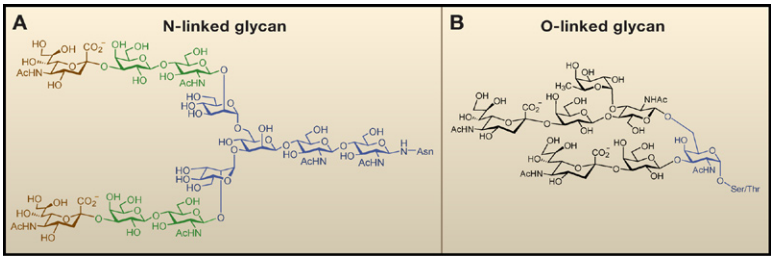

Figure 1. Common Cell Surface Glycans

(A) An N-linked glycan. The $Man_3GlcNAc_2$ pentasaccharide core (blue) is conserved across all N-glycan structures and can be elaborated with *N*-acetyllactosamine units (green), sialic acid residues (brown), and other epitopes.

(B) A mucin-type O-linked glycan. The core GalNAc moiety (blue), installed on a serine/threonine residue by a member of the ppGalNAcT family, is common to all mucin-type O-linked glycans.

against members of the ppGalNAcT family, revealing two structurally related compounds that disrupt mucin-type O-linked glycosylation in cultured cells and organs (Hang et al., 2004). Like tunicamycin, these compounds suppress O-linked glycosylation across the proteome, without specificity for discrete glycoprotein targets. This broad-spectrum inhibitory activity may account for the apoptotic effects observed when certain cell types are exposed to these reagents.

Small molecules that disrupt peripheral glycan "epitopes" are more likely to provide functional data without causing global cellular effects. This goal has been achieved using chemical "primers" of glycosylation (Brown et al., 2003). These cell-permeable small molecules—typically comprising hydrophobic glycosides of simple sugars—act as decoys for glycosyltransferases and thus compete with endogenous substrates for their activity. When supplied to cells, the primers suppress the elaboration of endogenous glycans, thereby perturbing the structures observed on cell-surface glycoproteins (Figure 2A). Concomitantly, the elaborated glycans can be observed on the primers, which are eventually secreted from the cell.

Among the most popular of these reagents is α-benzyl GalNAc, a primer that competes for elaboration of the core GalNAc residue found in mucin-type O-linked glycans (Figure 2A). Studies using α-benzyl GalNAc to truncate O-linked glycans suggest a link between this form of glycosylation and cancer metastasis. Primers have also been used to perturb glycosaminoglycans (e.g., heparan sulfates) and complex cancer-associated glycans.

In addition to glycosyltransferases, other glycan-processing enzymes can be targeted for inhibition with small molecules. Glycosidases, enzymes that remove sugars from glycans, are one example. Members of this enzyme family degrade glycans within lysosomes and processing N-linked glycans in the endoplasmic reticulum and Golgi (Varki et al., 1999). The natural products deoxynojirimycin and swainsonine inhibit glycosidases involved in N-linked glycan maturation and can be used to disrupt the elaboration of terminal epitopes on N-linked glycans. In the presence of these inhibitors, N-linked glycans are installed on proteins, but their peripheral epitopes lack structures typically involved in molecular recognition. Thus, deoxynojirimycin and swainsonine can perturb N-linked glycan structures in a more refined manner than tunicamycin.

The above small molecule tools target the enzymes involved in glycan assembly as either inhibitors or competitive substrates. An alternative strategy for chemical perturbation involves the metabolic incorporation of chain terminators into the glycan. Similar to dideoxy nucleotides in a DNA polymerase reaction, deoxy or otherwise modified monosaccharides can be incorporated into glycans by the cell's metabolic machinery, resulting in truncation of the glycan at that site (Figure 2B). For the strategy to succeed, the modified monosaccharide must be taken up by cells, processed into a nucleotide sugar substrate, and transferred by the relevant glycosyltransferases onto the appropriate glycan targets. Thus, substrate promiscuity among metabolic enzymes is a strict prerequisite. Hsieh-Wilson and coworkers recently treated neurons with 2-deoxygalactose, which was integrated into glycans but refractory to further elaboration with fucose at the 2-position (Figure 2B). The observed stunted neurite outgrowth and delayed synapse formation implicated the fucosylated glycans in proper neuronal development (Murrey et al., 2006).

Visualizing Glycans with Chemical Reporters

GFP and its relatives have revolutionized biology by enabling the visualization of proteins within authentic environments. Analogous tools for visualizing glycans could propel major advances in the field of glycobiology. At the cellular level, the ability to image glycans could enable studies of localization, trafficking, and dynamics. At the organismal level, glycan visualization would facilitate the real-time analysis of changes in glycosylation associated with development and disease. To date, these changes have been studied primarily by lectin and antibody staining of tissues ex vivo or inferred by expression profiling of glycosyltransferases. The direct monitoring of glycans within living animals using lectin or antibody conjugates is a promising direction, but these protein-based reagents have limited tissue access and may be restricted to imaging vascular targets.

An emerging approach toward glycan visualization exploits their underlying biosynthetic machinery for the incorporation of detectable probes. We have termed this the "bio-orthogonal chemical reporter" strategy (Figure 2C) (Prescher and Bertozzi, 2005). First, a monosaccharide substrate is modified with a functional group (the reporter) that is chemically inert in biological systems (hence the term "bio-orthogonal"). Upon administration to cells, the modified sugar is processed similarly to its native counterpart and integrated into cellular glycans. Finally, the labeled glycans are reacted with a detectable

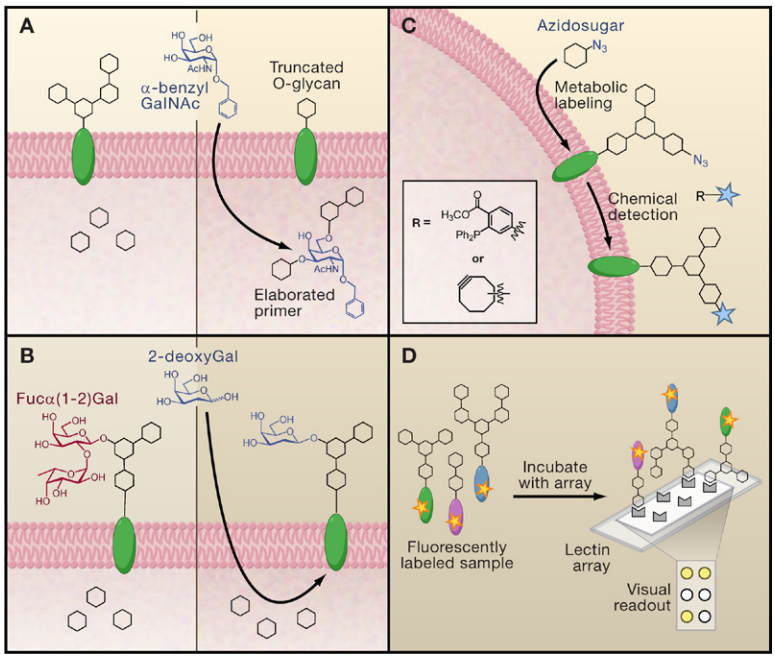

Figure 2. Strategies for Perturbing, Visualizing, and Profiling Glycans

(A) Suppressing the elaboration of cell-surface glycans with chemical primers. The primer α-benzyl GalNAc (blue) competes with endogenous substrates for elaboration of the core GalNAc residue. Cells treated with this compound express truncated mucin-type O-linked glycans (monosaccharide units shown in black).

(B) Blocking the extension of glycan chains with deoxy sugars. Neurons incubated with 2-deoxyGal (blue) incorporate the sugar into cell-surface glycans but are unable to elaborate the epitope at the 2-position. The Fucα(1-2)Gal-deficient neurons display abnormal growth characteristics.

(C) Visualizing and profiling glycans via the bio-orthogonal chemical reporter strategy. Synthetic azidosugars are incorporated into cellular glycans by the cell's metabolic machinery. In a second step, the azidosugars are detected. Depending on the choice of probe (e.g., fluorophore or affinity tag), this strategy can be used to image or enrich glycoconjugate subtypes.

(D) Profiling glycans with lectin microarrays (Hsu et al., 2006). Chips arrayed with lectins are incubated with fluorescently labeled biomolecules, cells, or tissue lysates. The pattern of bound glycoconjugates provides a readout of specific classes of glycan structures that are present in the mixture.

probe linked to a complementary bio-orthogonal functional group. The mutually selective chemical reactivity of the two functional groups ensures that only the metabolically labeled glycans are targeted for detection.

This approach hinges on the selection of a reporter group that can be readily installed on monosaccharide substrates with minimal perturbation to their structures. This requirement precludes the use of conventional fluorophores or affinity reagents (e.g., biotin), which are large relative to the sugar and likely to impede enzymatic processing within cells. We have found that organic azides ($R\text{-}N_3$) are ideal chemical reporters. Azides are small functional groups that are metabolically stable, essentially inert in biological systems, and selectively reactive with phosphines via the Staudinger ligation or alkynes via a [3+2] cycloaddition (Figure 2C). We and others have visualized numerous glycan types on living cells using the bio-orthogonal chemical reporter strategy (Dube and Bertozzi, 2003; Keppler et al., 2001). For example, azido analogs of ManNAc (i.e., *N*-azidoacetylmannosamine, ManNAz) and GalNAc (i.e., *N*-azidoacetylgalactosamine, GalNAz) have been used to tag sialylated and mucin-type O-linked glycoproteins, respectively, with detectable probes. Importantly, both azidosugar metabolism and subsequent chemical reactions are well tolerated by cultured cells.

Extension of the bio-orthogonal chemical reporter system to the visualization of glycans in living animals appears possible. In an important precedent, both the metabolic labeling of sialylated glycans with ManNAz and subsequent chemical reaction with a phosphine reagent were achieved in mice (Prescher and Bertozzi, 2005). Similarly, mucin-type O-linked glycans were labeled with GalNAz and then reacted with a phosphine probe in vivo (Dube et al., 2006). This work is now being applied to the noninvasive imaging of glycans in living animals, with the goal of monitoring changes in glycosylation associated with disease in real-time.

Profiling Glycans—The Emerging Field of Glycomics

System-wide analyses of transcripts and proteins, the heart of genomics and proteomics, have vastly increased our understanding of a cell's biochemical pathways. The revelation that many cellular systems, such as primary metabolism and transcription, are intimately connected has generated renewed appreciation for biological complexity at the systems level. The availability of complete genome sequences for many organisms provides a blueprint from which to infer their chemical makeup. Already, great strides have been made in relating an organism's genome to its proteome. A major frontier in systems biology is to further relate the genome and proteome to an organism's complete repertoire of glycans—the glycome (Murrell et al., 2004). Taking inventory of an organism's glycome is the first step in this process.

Like the other "omics" efforts, glycomics is being driven by new technologies for high-throughput profiling, several of which are rooted in chemistry. Glycan profiling can be performed at various hierarchical levels of structural information content. A first level of inquiry might be the identification of all glycoproteins bearing certain monosaccharide components and fluctuations in sugar modifications that are associated with physiological changes. As an example, numerous cytosolic and nuclear proteins are modified in a dynamic fashion with the monosaccharide β-O-GlcNAc linked to serine or threonine residues. This form of glycosylation is thought to play a role in regulating protein function (similar to phosphorylation), but the complete inventory of proteins modified with β-O-GlcNAc

is unknown. A convenient approach to profiling these proteins involves the bio-orthogonal chemical reporter strategy mentioned above. Zhao and coworkers (Nandi et al., 2006) metabolically labeled cultured cells with *N*-azidoacetylglucosamine (GlcNAz) and then tagged the labeled proteins with a phosphine-biotin conjugate. Subsequent affinity capture and mass spectrometry analysis revealed many previously unidentified proteins bearing the β-O-GlcNAc modification. A similar approach can be applied to profiling sialylated or mucin-type O-linked glycoproteins using the appropriate azidosugar substrates.

A second level of inquiry might involve the profiling of higher-order glycan structures comprising specific linkage patterns. Subtle changes in the connectivity of monosaccharide units can have profound biological consequences, many of which have been correlated with stages of development or pathogenesis. The ability to profile diverse glycan structures has been aided by a battery of commercial antibodies and naturally occurring lectins. These reagents can be immobilized on microarrays, which are then incubated with fluorescently labeled tissue or cell lysates. The pattern of bound glycoconjugates provides a profile of specific classes of glycan structures that are present in the sample (Figure 2D).

Lectin/antibody microarrays provide high-throughput analysis with moderate structural information content but are limited by the number of discrete glycan epitopes for which complementary lectins or antibodies are available. Thus, this profiling method will not detect all glycans in a complex sample, nor will it provide complete structural assignments for most glycans. To satisfy this level of inquiry, mass spectrometry has been artfully employed (Dell and Morris, 2001; Morelle and Michalski, 2005). This technique can provide detailed structural information on glycans, including composition, connectivity of monosaccharides, and the identity of underlying protein or lipid scaffolds. Furthermore, high-resolution mass spectrometry techniques can simultaneously analyze many glycan structures from complex cell or tissue samples. Combined with chromatographic separations, mass spectrometry has the potential to fully define an organism's glycome. However, considerable expertise in instrumentation and sample preparation is required for this level of glycan analysis (Park and Lebrilla, 2005).

In addition to glycan profiling, there is parallel interest in characterizing the glycan binding activities of isolated proteins and intact cells. Homogeneous samples of glycans are required for this purpose, underscoring the importance of advances in carbohydrate chemistry (Hanson et al., 2004). Already, synthetic glycans have been arrayed on chips and integrated into water-soluble probes for high-throughput profiling of living cells (Kiessling et al., 2000; Paulson et al., 2006).

Future Outlook

The chemical tools highlighted above for perturbing, visualizing, and profiling glycans have reduced the barriers to experimental pursuit that have historically plagued these biopolymers. Still, the experimental playing field is far from level with that for proteins or nucleic acids, and expansion of the chemical toolkit should therefore be a continuous effort. Additionally, for the expanding chemical toolkit to impact the field of glycobiology, new technologies must be transferred from chemistry to biology laboratories and, when possible, used in combination with prevailing genetic and biochemical tools. Biological researchers need access to small molecule inhibitors, metabolic labels, and synthetic glycans, as well as instrumentation and expertise for microarray and mass spectrometry profiling. Similar needs for genomic and proteomic technologies prompted the development of core facilities in academic centers and commercial services in the private sector. Likewise, we anticipate that biologists interested in studying glycans will benefit tremendously from recently formed core facilities that provide reagents, databases, and analytical services to the community at large (Raman et al., 2005).

REFERENCES

Brown, J. R., Fuster, M. M., and Esko, J. D. (2003). Glycoside Primers and Inhibitors of Glycosylation. In Carbohydrate-Based Drug Discovery, Wong, C.-H., ed. (Weinheim: Wiley-VCH), pp. 883–898.

Dell, A., and Morris, H.R. (2001). Science *291*, 2351–2356.

Dube, D.H., and Bertozzi, C.R. (2003). Curr. Opin. Chem. Biol. *7*, 616–625.

Dube, D.H., Prescher, J.A., Quang, C.N., and Bertozzi, C.R. (2006). Proc. Natl. Acad. Sci. USA *103*, 4819–4824.

Hang, H.C., Yu, C., Ten Hagen, K.G., Tian, E., Winans, K.A., Tabak, L.A., and Bertozzi, C.R. (2004). Chem. Biol. *11*, 337–345.

Hanson, S., Best, M., Bryan, M.C., and Wong, C.H. (2004). Trends Biochem. Sci. *29*, 656–663.

Hsu, K.L., Pilobello, K.T., and Mahal, L.K. (2006). Nat. Chem. Biol. *2*, 153–157.

Keppler, O.T., Horstkorte, R., Pawlita, M., Schmidt, C., and Reutter, W. (2001). Glycobiology *11*, 11R–18R.

Kiessling, L.L., Gestwicki, J.E., and Strong, L.E. (2000). Curr. Opin. Chem. Biol. *4*, 696–703.

Lowe, J., and Marth, J. (2003). Annu. Rev. Biochem. *72*, 643–691.

Morelle, W., and Michalski, J.C. (2005). Curr. Pharm. Des. *11*, 2615–2645.

Murrell, M.P., Yarema, K.J., and Levchenko, A. (2004). ChemBioChem *5*, 1334–1347.

Murrey, H.E., Gama, C.I., Kalovidouris, S.A., Luo, W.I., Driggers, E.M., Porton, B., and Hsieh-Wilson, L.C. (2006). Proc. Natl. Acad. Sci. USA *103*, 21–26.

Nandi, A., Sprung, R., Barma, D.K., Zhao, Y., Kim, S.C., Falck, J.R., and Zhao, Y. (2006). Anal. Chem. *78*, 452–458.

Park, Y., and Lebrilla, C.B. (2005). Mass Spectrom. Rev. *24*, 232–264.

Paulson, J.C., Blixt, O., and Collins, B.E. (2006). Nat. Chem. Biol. *2*, 238–248.

Prescher, J.A., and Bertozzi, C.R. (2005). Nat. Chem. Biol. *1*, 13–21.

Raman, R., Raguram, S., Venkataraman, G., Paulson, J.C., and Sasisekharan, R. (2005). Nat. Methods *2*, 817–824.

Varki, A., Cummings, R., Esko, J.D., Freeze, H., Hart, G.W., and Marth, J., eds. (1999). Essentials of Glycobiology (Cold Spring Harbor, NY: Cold Spring Harbor Press), pp. 1–653.