Glycoproteomics: protein modifications for versatile functions

Meeting on Glycoproteomics

Reiko T. Lee, Gordan Lauc & Yuan C. Lee[+]

Johns Hopkins University, Baltimore, Maryland, USA

This satellite meeting of the Thirtieth FEBS Congress and the Ninth IUBMB Conference was held in Dubrovnik, Croatia, between 28 and 30 June 2005. The conference was organized by G. Lauc, J. Dumic and M. Flögel from the University of Zagreb, Croatia.

Introduction

Glycosylation is perhaps the most common and versatile post-translational modification of proteins. It is estimated that well over one-half of all mammalian proteins are glycosylated (Apweiler *et al*, 1999). The roles of protein glycans, whether *N*- or *O*-linked forms, are varied. Glycan is required for the biological function of certain proteins, such as the Fc-effector function of immunoglobulin G (IgG), the regulated clearance of the glycohormone lutropin (LH), the targeting of lysosomal enzymes and controlling the circulation lifetime of glycoprotein drugs (for example, erythropoietin). Glycans also serve as recognition targets for their complementary binding proteins, the lectins. This family includes self proteins that are involved in such diverse functions as fertilization,

Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA
[+]Corresponding author. Tel: +1 410 516 7041; Fax: +1 410 516 8716;
E-mail: yclee@jhu.edu

development, differentiation and lymphocyte circulation, as well as in the development of pathological conditions. Glycans are also used as sites of attachment by microbes and toxins, and often constitute the first contact point between pathogen and host. In addition, glycans serve subtle functions, such as protecting proteins from proteolytic enzymes, promoting protein folding into proper tertiary and quaternary structures, and modulating the biological activities of proteins. Other forms of glycans, such as glycolipids and proteoglycans, also have important structural roles, as well as acting as recognition targets.

Glycomics is the investigation and documentation of the large collection of glycan structures. The function of a glycan can vary depending on the protein to which it is attached and the precise site of attachment. It is imperative, therefore, to document the structure and function of the entire glycoprotein. This integrated approach of compiling glycoprotein structure and function is defined as glycoproteomics. In this meeting, various aspects of the structure–function relationship of glycoproteins were discussed, with an emphasis on the roles of glycans in the development of numerous pathological conditions, as well as in pathogen–host interactions.

Glycans in pathogenesis

The biosynthesis of *N*- and *O*-glycans involves many glycosyltransferases. Several of the early glycan-transfer steps are obligatory, but there is often competition between two or more transferases during later processing steps. Some well-known changes in glycan structures during neoplastic transformation seem to result from changes in the balance of transferases. H. Narimatsu (Tsukuba, Japan) presented an example of marked downregulation of an *N*-acetylglucosaminyl transferase (GnT), β3Gn-T6, which is involved in *O*-glycan synthesis in gastric and colon cancers. It is well known that T (Galβ3GalNAcα-Ser/Thr) and Tn (GalNAcα-Ser/Thr) antigens are expressed in some cancer cells. As seen in Fig 1, two transferases compete for the same substrate, the Tn antigen. A severe downregulation of β3Gn-T6, which is responsible for the formation of the Core 3 structure, might be directly responsible for the abundance of T and Tn antigens in these carcinomas. Another example is the competition between GnT-III and GnT-V in *N*-glycan biosynthesis. GnT-III transfers a
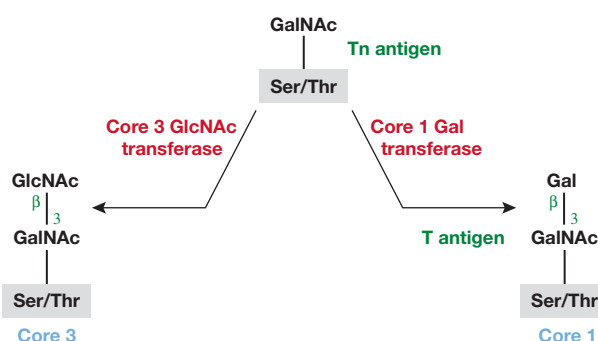
Fig 1 | The transferases for the formation of Core 1 and Core 3 structures compete for the same attachment site on the Tn antigen.
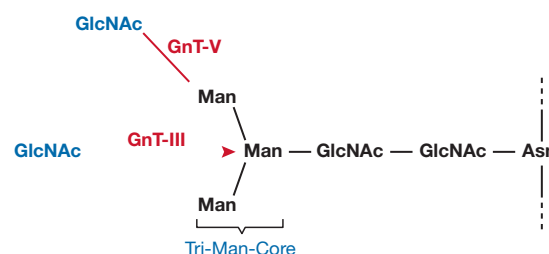
Fig 2 | Trimannosyl core structure of the *N*-glycan and the GlcNAc-attachment sites by GnT-III and GnT-V.

GlcNAc residue (bisecting GlcNAc) to the central mannose residue of the trimannosyl core structure of *N*-glycan, whereas GnT-V transfers a β6-linked GlcNAc to one of the outer mannose residues of the mannosyl core structure (Fig 2). Once the bisecting GlcNAc is attached by the action of GnT-III, the transfer of another GlcNAc residue by GnT-V becomes greatly hindered. The β6-linked GlcNAc residue has the highest potential for elongation, and the presence of this chain is often implicated in metastatic potential. This phenomenon can be explained by the finding of N. Taniguchi (Osaka, Japan) that matriptase, which is a membrane-bound serine protease, becomes constitutively active when its glycan is modified to contain the β6 branch. The relationship between matriptase activity and metastatic potential is being actively investigated. A rare form of *O*-glycan, in which an *O*-mannosyl group is attached directly to the peptide, is found in limited numbers in the brain, nervous system and muscle. T. Endo (Tokyo, Japan) reported that the protein *O*-mannosyl transferase 1 (POMT1) is responsible for adding *O*-mannose to α-dystroglycan, and mutations in *POMT1* that result in aberrant glycosylation of α-dystroglycan cause some forms of congenital muscular dystrophy. An interesting case of the modulation of glycosylation in a parasite was presented by S. Turco (Lexington, KY, USA). A dominant surface glycoconjugate of *Leishmania major* is a lipophosphoglycan, the backbone of which consists of repeats of a phosphodiester-linked disaccharide. Modification of the backbone with galactose side chains enables *L. major* to attach to the midgut of the sandfly vector through a midgut lectin, whereas subsequent modification with D-arabinose allows the parasite to detach from the midgut as it progresses through its life cycle. A large family of genes is involved in side-chain galactose attachment and two families are involved in arabinose attachment. How the parasite controls the expression of galactosyl and arabinosyl transferases is an interesting subject for future studies.

Many mammalian lectins, such as galectins, siglecs and C-type lectins, are involved in homeostasis as well as in defence-orientated functions. R. Schnaar (Baltimore, MD, USA) described the roles of myelin-associated glycoprotein (MAG), which belongs to the siglec family that recognizes sialic acid, and is present as a minor constituent in central and peripheral nervous system myelin. MAG interacts strongly with two gangliosides, GD1a and GT1b, which are present on the axon surface, and this interaction is required for the long-term stability of axons. Interestingly, although these particular gangliosides are prominent in the brain, they are hardly detectable in the liver, again highlighting the importance of the differential expression and control of glycosyl transferases within tissue type, development and neoplastic transformation. MAG is also an axon-regeneration inhibitor. A promising observation is that the MAG-mediated inhibition of axon regeneration was reversed when the interaction between MAG and its ligands was inhibited (Vyas *et al*, 2005).

An unusual glyco-polymer, polysialic acid, is present prominently on the embryonic form of neural cell-adhesion molecule (NCAM) and is postulated to support neuronal plasticity. Two key enzymes, polysialyltransferases ST8-Sia-II and ST8-Sia-IV, are involved in the polysialylation of NCAM. R. Gerardy-Schahn (Hanover, Germany) reported that the knockout of individual polysialyltransferases produced mice with distinctive but mild phenotypes, whereas knocking out both transferases caused a lethal phenotype characterized by severe brain-wiring defects. Interestingly, the further inactivation of the *NCAM* gene rescued the defect. These data show the main role of polysialic acid in controlling NCAM function during mouse development.

Congenital disorders of glycosylation (CDGs) are a diverse collection of genetic disorders that result from proteins being aberrantly glycosylated. H. Freeze (La Jolla, CA, USA) reported that in several types of CDG, the proteoglycan heparan sulphate (HS) is lost from intestinal epithelial cells during environmental stress, and this correlates with the onset of protein-losing enteropathy (PLE). The leakage of plasma proteins through the intestine caused by HS loss is amplified by the inflammatory cytokines, interferon-γ and tumour necrosis factor-α, as well as hydrostatic pressure. The loss of HS might be due to faulty processing of a major glycoprotein carrier of HS, syndecan-1 (Bode *et al*, 2005). Protein glycosylation could also be involved in the initiation of autoimmune diseases. The remnant epitope-generated autoimmunity (REGA) model previously put forward by G. Opdenakker (Leuven, Belgium) and colleagues postulates that glycopeptides represent important remnant epitopes, which might activate T cells to initiate autoimmune diseases (Opdenakker *et al*, 1993). Opdenakker presented results on the proteolysis of type II collagen by gelatinase B, which support this hypothesis. Glycosylation might also protect certain protein domains from proteolysis, modulate the activity of the proteases involved and influence their interactions with naturally occurring protease inhibitors.
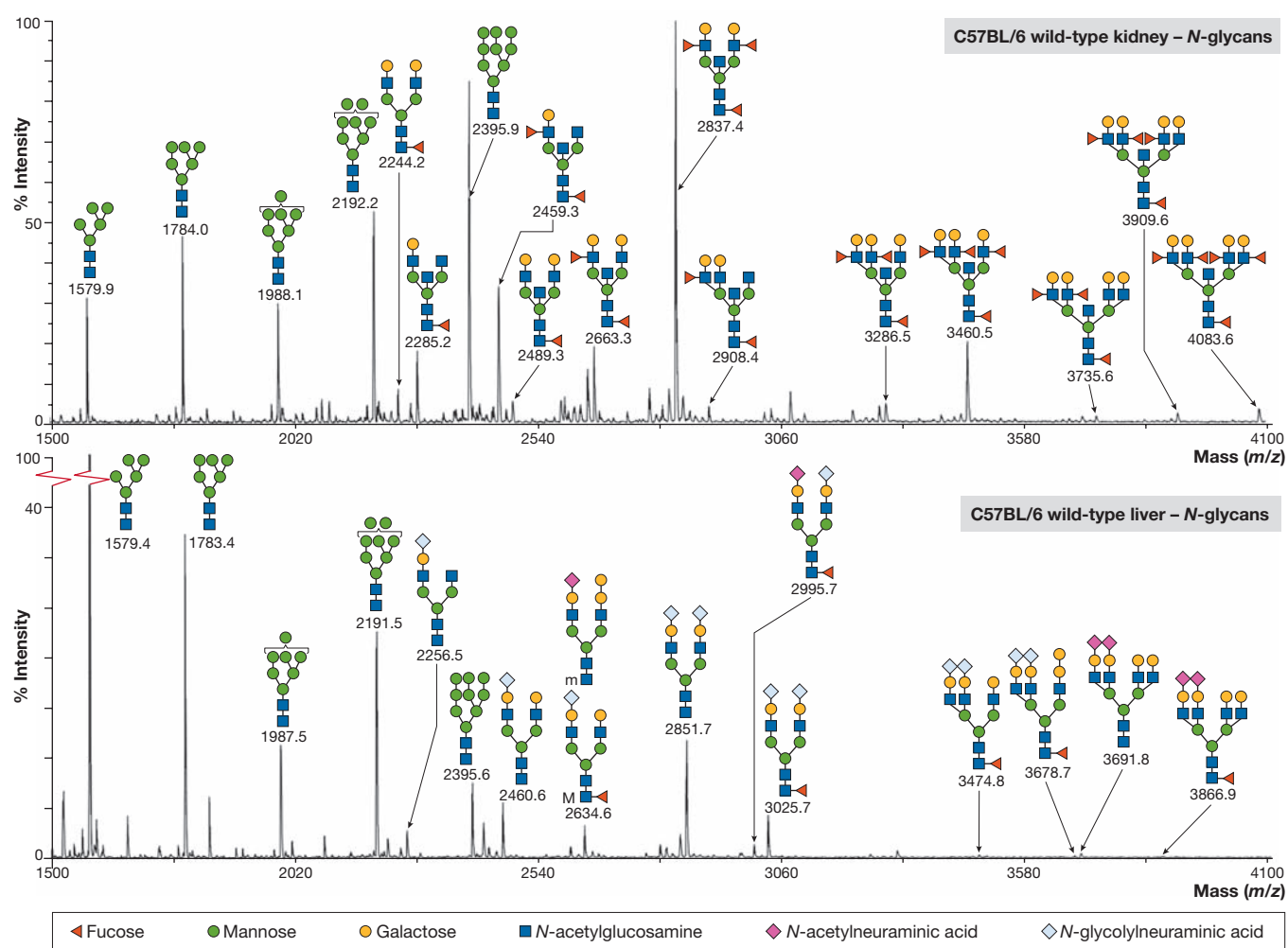
Fig 3 | Matrix-assisted laser desorption/ionization mapping of *N*-linked glycans from tissue samples of kidney and liver. Figure courtesy of A. Dell, London, UK.

Glycans in disease and therapeutics

The surface of human immunodeficiency virus (HIV) is covered with gp120/gp41 glycoprotein complexes. The surface-exposed gp120 contains numerous *N*-glycans, many of which are of the high-mannose type. P. Rudd (Oxford, UK) discussed an unusual antibody, 2G12, which can neutralize infection by many HIV isolates through recognizing an unusually tight cluster of mannose residues present on gp120. The antibody 2G12 contains crucial mutations that result in a much narrower distance between binding pockets, owing to domain swapping between two antennae of the IgG molecule (Calarese *et al*, 2003). Interestingly, H. Umeyama (Tokyo, Japan) described an antiviral protein, actinohivin, which is a lectin that contains three homologous mannose-binding sites, and is active against HIV and simian immunodeficiency virus (SIV). It is postulated that actinohivin also binds to a cluster of three high-mannose-type glycans of gp120, which fixes the gp120 conformation in such a way as to block its interaction with host-cell CD4. *N*-glycans of gp120 can also be exploited as targets for therapeutics. The biosynthesis of gp120 requires glyco-mediated chaperone activity for proper folding of the protein in the

endoplasmic reticulum. α-Glucosidases are needed to produce glycan structures that are recognized by the chaperones. Rudd reported that *N*-butyldeoxynojirimycin (*N*B-DNJ), an inhibitor of α-glucosidases I and II, effectively reduced HIV infectivity, particularly when the inhibitor was delivered in liposomes.

Many important glycoprotein therapeutics have been expressed using various production vehicles; for example, cell lines, transgenic animals and plants (Jefferis, 2005). However, nonhuman glycosylation might affect their bioactivity, pharmacokinetics, stability, immunogenicity, allergenicity and so on. For these reasons, 'human-type' glycans are essential for therapeutics to be administered systemically. R. Jefferis (Birmingham, UK) discussed recombinant antibody therapeutics and illustrated how IgG–Fc-mediated biological activities can be profoundly influenced by the glycoforms present at the conserved glycosylation site. Interactions between IgG–Fc and its ligands depend crucially on the conformation of the IgG–Fc that is determined by numerous noncovalent interactions between the glycan and the protein surface. H. Kamerling (Utrecht, The Netherlands) described a detailed structural analysis of glycans that are present on protein

therapeutics produced in the milk of transgenic rabbits. The glyco-form profiles differed significantly from those of the natural human glycoprotein. Interestingly, a therapeutic lysosomal enzyme, α-glucosidase, contains masked mannose-6-phosphate residues (that is, the phosphate group was capped with α-GlcNAc), which require unmasking for efficient lysosomal targeting.

Glycans in pathogen–host interactions

Often the first step in a bacterial infection is the recognition of host glycans by bacterial lectins or vice versa. N. Sharon (Rehovot, Israel) discussed the adhesion of fimbriated bacteria, which is mediated by an adhesin (lectin) present at the tip of fimbria. A common urinary tract infection by *Escherichia coli* is mediated by the mannose-specific adhesin, FimH. Neoglycoproteins and large dendrimers that contain many residues of α-mannoside with phenyl or other hydrophobic aglycons were the most effective inhibitors of FimH interaction with host ligands. Other examples of bacterial lectins involved in infection include the adhesin on uropathogenic P-fimbriated *E. coli* that recognizes galabiose (Galα4Gal). Various secreted bacterial toxins, such as Shiga toxin and Shiga-like toxin (verotoxin), also recognize galabiose. Animal studies have shown conclusively that carbohydrate inhibitors can prevent infection by lectin-carrying bacteria. Anti-adhesion therapy of infectious diseases with powerful inhibitors, although not yet successful in humans, is clearly an important alternative to antibiotics. In particular, it is conceivable that the emergence of resistant strains would be much slower compared with antibiotic administration.

The frequent occurrence of galabiose as the target of pathogen or toxin recognition is interesting, as galabiose is present in humans as a minor component of glycolipids. No galabiose-containing glycoprotein has been reported in humans. By contrast, pigeons and related birds express galabiose abundantly as *N*-glycans. Y. Lee (Baltimore, MD, USA) reported that such glycoproteins immobilized on gels can effectively capture Shiga-like toxin. Interestingly, most of the birds that are familiar to us as food sources, such as ducks and chickens, belong to a minor branch in the phylogenetic tree that was separated from the main branch (representing ~95% of bird species) some 65–100 million years ago. None of the birds belonging to the minor branch express galabiose, whereas many do in the main branch (Suzuki *et al*, 2004).

Analytical methods for glycoproteomic studies

Glycoproteomics (that is, the structural investigation of glycoproteins) should be carried out at various levels. Initially, it is desirable to assess the types of glycan present and their position of attachment on a particular protein backbone. For this, mass spectrometry (MS) strategies are often most suitable. A. Dell (London, UK) described a sensitive multi-MS strategy that allows the analysis of simple or complex mixtures from biological extracts (Fig 3). For a detailed glycan-structure analysis, several fluorescence-based methods, as well as direct analysis by a Dionex system (Dionex Corporation, Sunnyvale, CA, USA), are available. Recent innovations on the Dionex analytical system were presented by J. Weiss (Idstein, Germany) from Dionex Corporation. K. Kato (Nagoya, Japan) reported on the use of stable-isotope-assisted nuclear magnetic resonance (Fig 4) for obtaining atomic-contact information on the interaction between carbohydrate and protein.
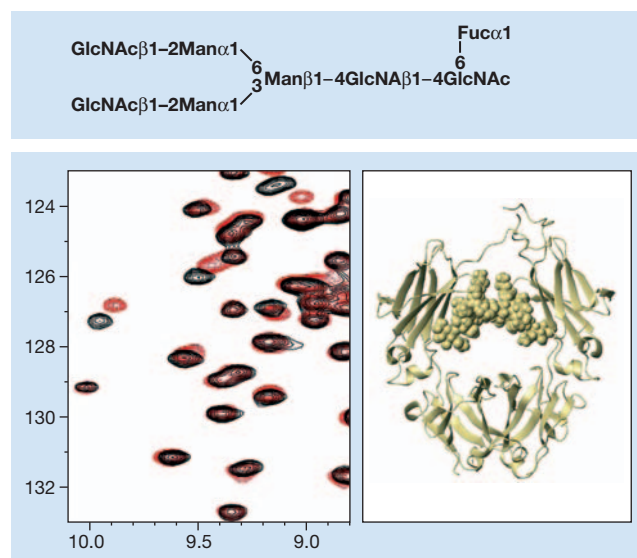
Fig 4 | Structure of IgG –Fc fragment with attached oligosaccharide chains determined using stable-isotope-assisted nuclear magnetic resonance. The protein backbone is shown as a ribbon and the two glycan chains are shown in space-filling mode. Figure courtesy of K. Kato, Nagoya, Japan.

For a survey of potential ligand structures of a lectin, many carbohydrate arrays on plastic or glass surfaces are now available. N. Bovin (Moscow, Russia) described a new array system in which carbohydrate–lectin interactions are performed in micro-drops of a high-permeation gel rather than on a flat plastic or glass surface. The gel system seems to be more sensitive than those that use two-dimensional surface interactions, and might provide a more natural environment for glycan–protein interactions. In a more physiological system, M. Heffer-Lauc (Osijek, Croatia) reported the pitfalls of gangliosides 'jumping' between tissues and cells under the normal immunostaining conditions. Jumping also seems to occur with glycosylphosphatidylinositol (GPI)-anchored proteins, and distinguishing the actual physiological locations of gangliosides and GPI-anchored proteins from their immunostaining artefacts presents a great challenge (Heffer-Lauc *et al*, 2005).

Glycomics is inherently more complex than proteomics, even at the level of a single organism. This is because of the tremendous structural complexity and diversity that results from different anomeric configurations, the presence of several substitution sites, the possibility of double or triple substitutions (that is, branching), and the presence of non-carbohydrate substituents, such as sulphate, phosphate and acetate, which might modify any number of hydroxyl groups. In addition, each animal group is likely to present different structures from those found in mammals. For this reason, it is imperative for the glycobiology community to pool and share available knowledge and resources. Indeed, this is already taking place in different regions of the world. C. von der Lieth (Heidelberg, Germany) discussed some of the more prominent open-access consortia with respect to their methods of operation and particular strengths (von der Lieth *et al*, 2004); these include the Kyoto Encyclopedia of Genes and Genomes (KEGG) in Japan,

the Consortium for Functional Glycomics (CFG) in the USA and EurocarbDB (including glycoSCIENCES.de in Germany). Strong interactions among these three, as well as further integration of other groups and resources, are actively being pursued. These efforts are likely to be the vanguard for future advances in integrated glycoproteomics.

REFERENCES

Apweiler R, Hermjakob H, Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* 1**473:** 4–8

Bode L, Eklund EA, Murch S, Freeze HH (2005) Heparan sulfate depletion amplifies TNF-α-induced protein leakage in an *in vitro* model of protein-losing enteropathy. *Am J Physiol Gastrointest Liver Physiol* **288:** G1015–G1023

Calarese DA *et al* (2003) Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300:** 2065–2071

Heffer-Lauc M, Lauc G, Nimrichter L, Fromholt SE, Schnaar RL (2005) Membrane redistribution of gangliosides and glycosylphosphatidylinositol-anchored proteins in brain tissue sections under conditions of lipid raft isolation. *Biochim Biophys Acta* **1686:** 200–208

Jefferis R (2005) Glycosylation of recombinant antibody therapeutics. *Biotechnol Prog* **21:** 11–16

Opdenakker G, Rudd PM, Ponting CP, Dwek RA (1993) Concepts and principles of glycobiology. *FASEB J* **7:** 1330–1337

Suzuki N, Laskowski M Jr, Lee YC (2004) Phylogenetic expression of Galα1–4Gal on avian glycoproteins: glycan differentiation inscribed in the early history of modern birds. *Proc Natl Acad Sci USA* **101:** 9023–9028

von der Lieth CW, Bohne-Lang A, Lohmann KK, Frank M (2004) Bioinformatics for glycomics: status, methods, requirements and perspectives. *Brief Bioinform* **5:** 164–178

Vyas AA, Blixt O, Paulson JC, Schnaar RL (2005) Potent glycan inhibitors of myelin-associated glycoprotein enhance axon outgrowth in vitro. *J Biol Chem* **280:** 16305–16310

Reiko T. Lee   *Gordan Lauc*   *Yuan C. Lee*