REVIEW

Glycoprotein Structure Determination by Mass Spectrometry

Anne Dell* and Howard R. Morris

The human genome encodes 30,000 to 40,000 proteins, and a major challenge is to understand how posttranslational events, such as glycosylation, affect the activities and functions of these proteins in health and disease. Glycosylated proteins are ubiquitous components of extracellular matrices and cellular surfaces where their oligosaccharide moieties are implicated in a wide range of cell-cell and cell-matrix recognition events. The power of ultrahigh-sensitivity mass spectrometric strategies for defining the primary structures of highly complex mixtures of glycoprotein glycoforms is set to revolutionize structural glycobiology in the coming postgenomic era.

Some 25 years ago, the first glycoprotein primary structure was determined successfully by mass spectrometry (MS) (*1*). These early studies of so-called antifreeze proteins led to the discovery of multiple disaccharides O-linked through threonines on a heterogeneous polypeptide backbone sequence. The methodologies that were available at the time, electron impact (EI) and chemical ionization (CI) MS, required the initial production of volatile derivatives. The derivatization chosen was permethylation, which served the dual purpose of conferring volatility and of releasing the saccharide portions of the molecule by beta elimination while "labeling" the sites of attachment to the polypeptide by the concomitant creation of dehydrated residues from the substituted threonines. The permethylation strategy survives to this day for some aspects of mass spectrometric carbohydrate analysis, although the analysis of the intact polymer units, whether oligosaccharide, polypeptide, or glycopeptide, is now conducted by the advanced "soft ionization" methods of fast atom bombardment (FAB), electrospray ionization (ES), or matrix-assisted laser desorption ionization (MALDI). Considerable technological advances in soft ionization MS have been made over the past two decades in particular. However, full structural characterization of a glycopeptide or glycoprotein also requires definition of branching, linkages, configurations, and the identification of same-mass sugar isomers. A complete analysis will therefore use not only the modern soft ionization MS approaches, but also gas chromatography EI-MS of hydrolysates for composition and linkage analysis, together with biochemical methods such as exoglycosidase digestion to define terminal sequences. In addition, many structural studies are facilitated by knowledge of the biosynthetic pathways of oligosaccharides, and therefore we first summarize the key principles of N- and O-oligosaccharide biosynthesis, focusing on animal glycosylation. We then describe the different soft ionization MS methods that are currently in use in glycobiology, and highlight some areas of mammalian structural glycobiology in which MS currently plays a key role.

Biosynthetic Pathways Define Key Structural Features of Glycoprotein Oligosaccharides

There are two main types of protein glycosylation: N-glycosylation, in which the oligosaccharide is attached to an asparagine residue, and O-glycosylation, in which the oligosaccharide is attached to a serine or threonine residue. Glycoproteins usually exist as complex mixtures of glycosylated variants (glycoforms). Glycosylation occurs in the endoplasmic reticulum (ER) and Golgi compartments of the cell and involves a complex series of reactions catalyzed by membrane-bound glycosyltransferases and glycosidases. Many of these enzymes are exquisitely sensitive to other events taking place within the cell in which the glycoprotein is expressed. The populations of sugars attached to an individual protein will therefore depend on the cell type in which the glycoprotein is expressed and on the physiological status of the cell, and may be developmentally and disease regulated.

All mammalian N-linked oligosaccharides share a common trimannosyl-chitobiose core, which is derived from a biosynthetic precursor $Glc_3Man_9GlcNAc_2$ (Figs. 1 and 2) that is added cotranslationally to polypeptides in the ER. The three glucosyl (Glc) residues are removed by ER-resident glucosidases, and the resulting $Man_9GlcNAc_2$ oligosaccharide is further processed in the ER and Golgi via pathways that are shared at least partially by all plant and animal cell types (*2*, *3*). Processing involves stepwise trimming by exoglycosidases and stepwise addition of new sugar residues catalyzed by glycosyltransferases.

Trimming by α-mannosidases, without any subsequent glycosyl addition to the periphery, results in oligosaccharides of composition $Man_xGlcNAc_2$ called "high mannose" structures (legend to Fig. 2A). Trimming by α-mannosidases plus glycosyl addition to the distal side of the core gives the most abundant class of mammalian oligosaccharides, the so-called complex-type structures (Fig. 2B). They are characterized by the presence of variable numbers of antennae (most commonly two to four) whose biosynthesis is initiated by the addition of GlcNAc "stubs" to the two α-mannoses of the core. GlcNAc may also be attached to the β-mannose; a GlcNAc residue attached at C-4 of this mannose is referred to as a "bisecting" residue. Unlike the GlcNAc "stubs," the bisecting GlcNAc is never further substituted with glycosyl residues. In plants and some invertebrates (but not mammals), the β-mannose can be substituted with xylose. The core is often modified by the addition of fucose to the proximal GlcNAc (*4*).

Further processing converts the small pool of "core plus stubs" into an extensive array of mature oligosaccharides. In mammals, the antennae stubs are usually elongated by the addition of β-galactose (β-Gal) to form the ubiquitous antenna building block Galβ1-4GlcNAc (LacNAc). Antennae can be lengthened by the sequential addition of GlcNAc and Gal residues, resulting in tandem repeats of LacNAc, i.e., "polylactosamine." In some mammalian glycoproteins, and not uncommonly in invertebrates (*5*), β-GalNAc is added to the GlcNAc stubs in place of β-Gal, yielding antennae with a LacDiNAc backbone (GalNAcβ1-4GlcNAc). Biosynthesis of complex-type structures is completed by a variety of "capping" or "decorating" reactions, the most important being the addition of sialic acid, fucose, and sulfate. Capping sugars are usually alpha-linked and therefore protrude away from the beta-linked ribbonlike antennae backbones, thus facilitating their presentation to lectins and antibodies. Many of the most abundant antennae in mammalian glycoproteins (Fig. 2B) were first identified as blood-group substances and are now known to play key roles in cell-cell and cell-matrix recognition events. For example, early events in lymphocyte homing are mediated by L-selectin molecules, constitu-

Department of Biochemistry, Imperial College of Science, Technology and Medicine, London SW7 2AY, UK.

*To whom correspondence should be addressed. E-mail: a.dell@ic.ac.uk

tively expressed on lymphocytes, which recognize sulfated sialylated Le^x (Lewis) structures (Fig. 2B) carried on glycoproteins on the surfaces of specialized endothelial cells (*6*).

In addition to high-mannose and complex-type N-linked oligosaccharides, there is a third, less common, family referred to as "hybrid." Members of this family share structural features of the other two families. They usually retain two mannoses on the 6-arm of the trimannosyl core, while complex-type antennae are elaborated on the 3-arm. Hybrid structures are frequently bisected and core-fucosylated.

O-Oligosaccharide biosynthesis is initiated in the Golgi by the addition of a single sugar to serine or threonine. In mammals, the initiating sugar is normally N-acetylgalactosamine (GalNAc), although alternative pathways exist [for example, mannose-linked O-oligosaccharides are common in brain glycoproteins (*7*)]. O-Oligosaccharides can vary in size from a single GalNAc residue (referred to as the Tn antigen) to oligosaccharides comparable in size to complex-type N-oligosaccharides. As with the latter, it is convenient to divide the glycan into two domains, namely the core and the antennae. The antennae are biosynthesized in a similar manner to N-oligosaccharides and consequently N- and O-oligosaccharides often carry the same terminal structures. The core domains are, however, very different. There are at least seven O-oligosaccharide core structures, four of which (core types 1, 2, 3, and 4) are particularly widespread in mammalian glycoproteins (Fig. 2C) (*3*).

FAB-MS, ES-MS, and MALDI-MS

The introduction of high-field mass spectrometry in the mid-1970s and fast atom bombardment ionization in the early 1980s, leading to HF-FAB-MS (*8*, *9*), revolutionized the structure determination of a wide range of carbohydrate-containing biopolymers. This revolution has continued with the newer techniques of ES-MS (*10*) and MALDI-MS (*11*). All three technologies permit the direct ionization and desorption of nonvolatile substances and are applicable to intact glycoconjugates or fragments thereof.

In a FAB-MS experiment, an accelerated beam of atoms, or more usually 30-kV ions, is fired from a gun toward a small metal target attached to the end of a probe. The target is loaded with a viscous liquid, called the matrix (usually glycerol or thioglycerol), in which the sample of interest is dissolved. FAB is particularly useful for analyzing the permethyl derivatives of oligosaccharides released from glycoproteins by chemical or enzymatic methods. When the atom or ion beam collides with the matrix, sample molecules are sputtered out of the matrix into the high vacuum of the ion source. A substantial number of these molecules are ionized during the sputtering process, producing positively charged species called quasimolecular ions $[M + H]^+$ and $[M + Na]^+$. Furthermore, during ionization, internal energy is imparted to the molecules, resulting in fragmentation of labile bonds. A particular strength of FAB-MS is its ability to provide both compositional (by way of the molecular ions) and sequence information [by way of the fragment ions, in particular nonreducing-end glycosidic bond cleavage ions (*12*)] from complex mixtures of oligosaccharides in a single experiment. The FAB-MS experiment is normally conducted on a magnetic sector double-focusing mass spectrometer, providing good sensitivity, mass accuracy, and resolving power in the mass range through and beyond the largest expected N-linked structures.

In ES-MS, a stream of liquid containing the sample of interest is introduced into the atmospheric-pressure ion source of a mass spectrometer by means of a glass- or metal-coated capillary. By applying appropriate voltages to the capillary, counterelectrode, and focusing optics, an aerosol of highly charged microdroplets is generated in the source, which traverses a series of skimmers where it encounters a drying gas. The net effect is the creation of gaseous ions, devoid of solvent, whose charge distribution is proportional to the number of ionizable groups in the molecule. The ionization process in ES-MS is very gentle, resulting in little or no fragmentation, although this can be engineered to some extent by varying source voltages. To overcome this problem, many ES instruments have tandem MS analyzers that allow the detection of fragment ions produced by collisional activation [collision-induced dissociation (CID)] of molecular ions in so-called MS-MS experiments in which the parent ion is selected by the first mass analyzer and the fragment ions are screened by the second analyzer. The most widely used, until relatively recently, were the tandem four-sector magnetic instruments for high-energy CID (*13*–*15*) and the so-called triple-quadrupole instruments for low-energy CID MS-MS (*16*). The most powerful current technology for low-energy MS-MS is the quadrupole orthogonal acceleration time of flight (Q-TOF) mass spectrometer (*17*, *18*). This instrument, which was conceived in the early 1990s in our laboratory as potentially the best tandem MS-MS configuration using an orthogonal TOF (*19*) component, has a novel geometry that offers significant advantages of sensitivity coupled with mass accuracy and resolution. It has had a major impact on proteomics research and is playing an increasingly important role in structural glycobiology. The instrument is compatible with a variety of ionization sources, including ES and MALDI. An important technical capability of ES-MS on Q-TOF–type instruments is

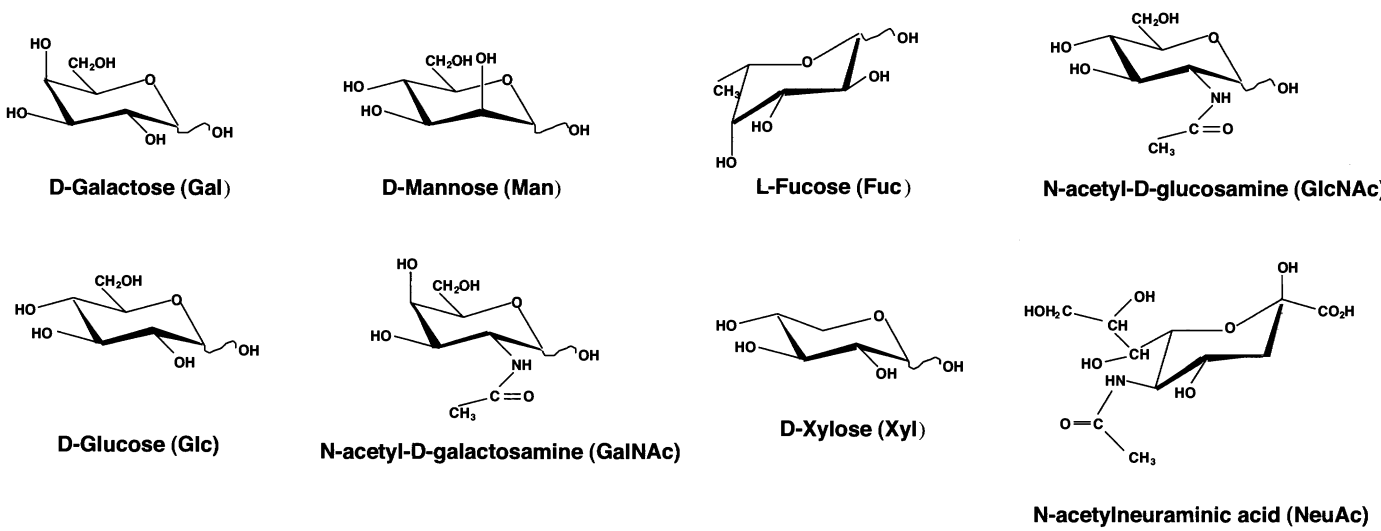

Fig. 1. Structures and names of the monosaccharides that are referred to in this article.

the possibility of carrying out high-sensitivity MS-MS experiments online, for example, on peaks eluting from an LC (liquid chromatography) system, in real time. Sample fractions for other experiments can be collected by stream splitting. MS-MS experiments may then be optimized by direct injection, or by nanospray experiments on small volumes of collected fractions (*17*).

MALDI-MS is arguably the most sensitive of the three ionization technologies, at least for simple mass analysis. In a MALDI experiment, the sample is embedded in a low molecular weight ultraviolet-absorbing matrix. Ionization is effected by a pulsed laser. The matrix absorbs the laser pulse, and enough energy is transferred to the sample, by mechanisms that are not fully understood but are akin to flash thermal evaporation, to enable the predominant formation of singly charged quasimolecular ions. A MALDI experiment yields very few fragment ions, and MALDI-MS is therefore the preeminent technique for screening molecular ions (mass mapping), especially when high-throughput and high sensitivity is required. The Q-TOF technique was originally developed with ES to take advantage of the MS/MS of doubly charged ions often found in peptide and glycopeptide research, but other sources can be fitted. MALDI–Q-TOF–type instruments have recently been developed (*20*, *21*), and their use will undoubtedly become more widespread. However, the inherent advantages in the low-energy CID of doubly charged ions may require the preparation of special derivatives for some MALDI studies, and in carbohydrate and glycopeptide work both ES-MS/MS on Q-TOF–type instruments and traditional FAB-MS will continue to play a major role in sequence determination.

In general, devising strategies to optimize fragment ion information is always a priority for glycopolymer mass spectrometrists. Three approaches are now widely used: inducing fragmentation by collisional activation, monitoring natural ionization-induced fragmentation, and selecting derivatives that enhance and direct fragmentation.

Collisional activation MS/MS. As discussed above, perhaps the most powerful current example of this approach for glycoproteins comes from ES work on the Q-TOF. When monitoring an LC elution or collected fractions from, say, an enzymatic digest of a glycoprotein, initially in the MS mode, it is a simple matter to identify the peaks or fractions containing glycopeptides from the presence of sugar-fragment ions in the individual mass spectra produced. Switching to the MS/MS mode of a doubly or triply charged ion of interest will, with careful tuning of the collision energy, give a composite spectrum containing fragmentation data on both the saccharide and peptide portions of the molecule. The glycosidic bonds are generally weaker than peptide bonds and thus fragment more easily, providing basic oligosaccharide sequence information. Peptide fragmentation also takes place, and the crucial signals to interpret in the MS/MS spectrum are those fragments on either side of the amino acid carrying the sugar(s). An example is our identification of an unusual cytoplasmic glycosylation of Skp1, an F-box binding protein, with West and co-workers at the University of Florida in Gainesville. Q-TOF ES-MS/MS analysis (*22*) of a Skp1-derived glycopeptide yielded a set of product ions, which identified the peptide and showed that it was modified at Pro^{143}. A second series of product ions showed that Pro^{143} was hydroxylated and derivatized with an unusual linear pentasaccharide, Hex-Hex-Fuc-Hex-HexNAc. On the basis of this work, the Q-TOF has been used in a proteomic study, in which de novo sequencing of the fucosyl and GlcNAc transferases produced primers leading to the full sequencing of the corresponding genes (*23*).

Natural fragmentation. Without the luxury of tandem or multiple sector mass spec-

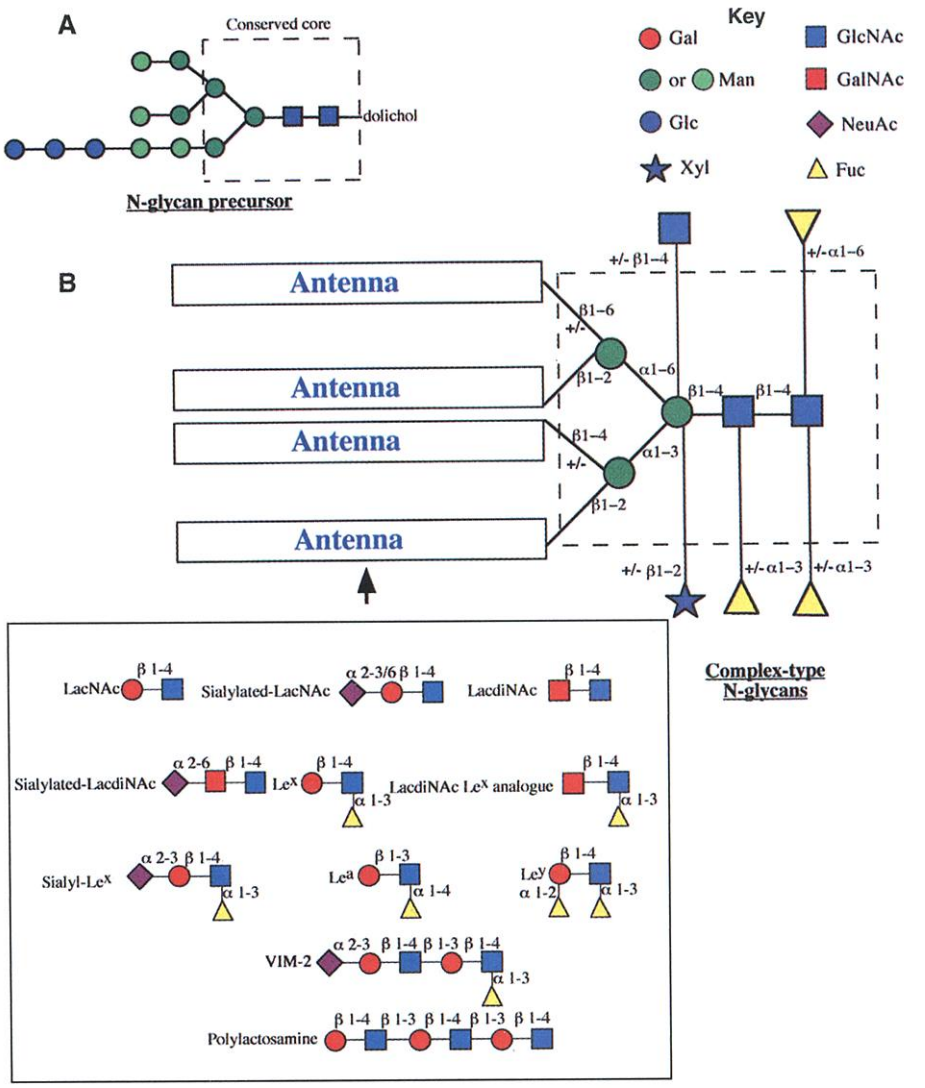

Fig. 2. (**A**) Structure of the dolichol-linked oligosaccharide, the precursor for all N-oligosaccharides. The conserved trimannosyl-chitobiose core is indicated by the dashed rectangle. High-mannose structures retain all the dark green mannoses but may lack one, two, three, or four of the light green mannoses. (**B**) Complex-type N-oligosaccharide structures. A composite diagram showing how the oligosaccharides are composed of two domains, namely, the core defined by the dashed rectangle and the antennae that are attached to the alpha-linked mannoses of the core; sequences frequently found in the antennae of mammalian N-oligosaccharides are depicted within the solid rectangle; the types of substitution that can occur on the core are shown outside the dashed rectangle. (**C**) The four most common O-oligosaccharide core types.

trometers, one option is to rely on natural fragmentation created by internal energy transfer to the ion during the ionization process. A good example is the abundant nonreducing-end fragmentation of permethyl oligosaccharides seen in FAB-MS spectra (*12*). Natural fragmentation can also be observed in MALDI-TOF instruments, where the metastable decomposition taking place after acceleration in the source region can be observed with a stepping reflectron mirror by means of a technique called postsource decay analysis (*24*). Although useful in many circumstances, this technique is not as powerful as full MS/MS analysis on a multisector or tandem instrument.

Derivatization methods. These can be broadly divided into two categories: "tagging" of reducing ends, and protection of most or all of the functional groups. Commonly used tagging reagents include *p*-aminobenzoic acid ethyl ester, 2-aminopyridine, 2-aminobenzamide, and aminolipids. This type of derivatization facilitates chromatographic purification and enhances the formation of reducing-end fragment ions in MS and MS/MS experiments. Protection of functional groups by permethylation is by far the most important type of full derivatization used in oligosaccharide MS. Permethylated derivatives form abundant fragment ions arising from cleavage on the reducing side of each HexNAc residue. This yields ions [usually referred to as A-type ions (*12*)] whose masses define important structural features of N- and O-antennae, including the types of capping sugars and the presence or absence of polylactosamine sequences (*12*). These ions are a valuable source of information in "mapping" experiments (see below).

Carbohydrate-Deficient Glycoprotein Syndromes

Carbohydrate deficient glycoprotein syndromes (CDGSs), now called congenital disorders of glycosylation, are rare, multisystemic diseases that are typically associated with major nervous system impairment. During the past decade, ES-MS has played a crucial role in defining the causes of these debilitating disorders. The serum transferrin of affected individuals serves as a convenient marker of glycosylation abnormalities. Normal transferrin has two fully occupied N-glycosylation sites and is a relatively homogeneous glycoprotein with more than 80% of its oligosaccharides being sialylated structures with two antennae (Fig. 3). This high level of homogeneity permits meaningful ES-MS analysis on the intact glycoprotein. Alterations in the oligosaccharide structures are readily revealed by shifts in molecular weight. Initial ES-MS experiments on transferrin from CDGS patients (*25*, *26*) revealed three major molecular species. The first had the correct mass for normal transferrin [mass-to-charge ratio (*m/z*) 79,570], whereas the others (*m/z* 77,364 and 75,157) were smaller by the masses of one and two biantennary, sialylated structures, respectively. These data indicated that the defect manifests itself in the occupancy of the N-glycosylation sites rather than in the oligosaccharide structure itself (Fig. 3). Subsequently, CDGS was shown to be caused by a defect in the gene encoding phosphomannomutase, which leads to insufficient biosynthesis of the oligosaccharide precursor required for N-oligosaccharide biosynthesis. At least six types of CDGS disorders are known, the majority of which are caused by deficiencies in enzymes required for synthesizing this precursor. In addition, one variant results from a deficiency of the Golgi resident N-acetylglucosaminyltransferase II, responsible for initiating the second antenna in N-oligosaccharide biosynthesis (*27*). This was revealed by a major molecular ion at *m/z* 78,247 in the serum transferrin ES spectrum, consistent with two sites being occupied with structures having only one antenna (Fig. 3).

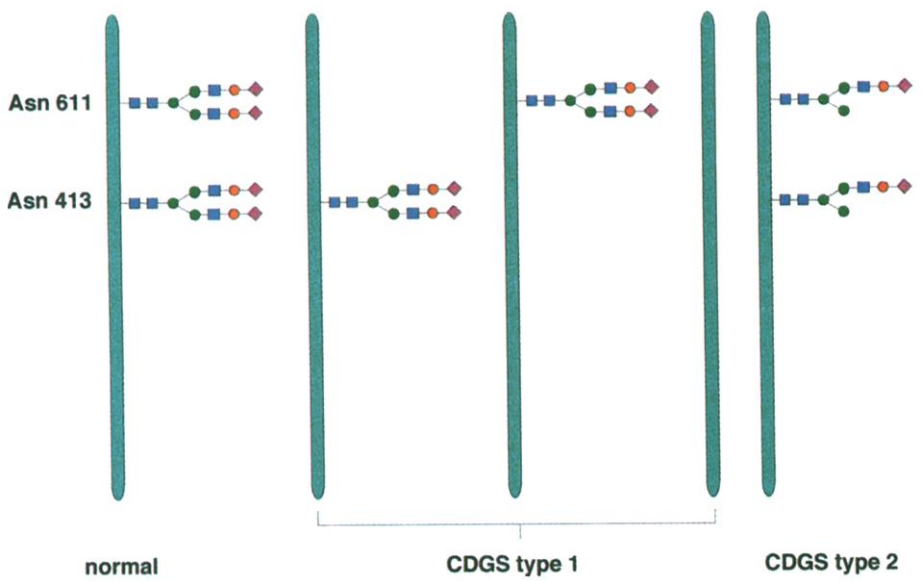

Fig. 3. Diagram of the major structures present in normal (left), CDGS type 1 (middle), and CDGS type 2 serum transferrin (right).

Prion Glycosylation

Spongiform encephalopathies involve conversion of the normal prion glycoprotein (PrP^c) into a proteolytically resistant, disease-causing conformer (PrP^{Sc}). PrP contains two conserved N-glycosylation sites, which are partially occupied by a highly heterogeneous array of complex-type oligosaccharides carrying sialylated and fucosylated antennae (*28*). It has been proposed that the occupancy of the N-glycosylation sites can be correlated with different strains of prions (*29*), but this hypothesis remains controversial (*30–33*). Deletions of one or both glycosylation sites in transgenic mice result in unusual patterns of PrP^{Sc} deposition (*34*), suggesting that glycosylation might play a role in the disease process, but the functions of the sugars remain enigmatic. Two recent studies incorporating high-sensitivity MS have provided new structural information that should facilitate future work aimed at addressing possible roles of sugars in prion diseases. In the first study (*35*), MALDI-MS molecular weight profiling of N-oligosaccharides released from hamster PrP and PrP^{Sc} revealed decreased amounts of oligosaccharides with bisecting GlcNAc residues, and increased amounts of tri- and tetraantennary sugars, in the PrP^{Sc} oligosaccharides compared with those from normal PrP. The second study (*36*) used ES methodology, including ultrahigh-sensitivity MS/MS experiments on a Q-TOF instrument to provide detailed information on the occupancy of the second glycosylation site (Asn^{196}) of murine PrP, thus complementing a previous MS investigation of glycosylation at Asn^{181} in the hamster (*37*). The two sites are substituted with more than 60 bi-, tri-, and tetraantennary core-fucosylated structures, many of which are bisected. The oligosaccharides carry a limited number of antennae types. The diversity and difference in site occupancy arise largely from the arrangements of the antennae types in individual oligosaccharides.

Mucins in Health and Disease

Mucins are the main component of the mucus protecting the internal epithelial layers of our body. They are characterized by dense O-glycosylation in tandem repeat domains that are rich in serine, threonine, and proline. Mucins carry a large repertoire of O-oligosaccharides, which are believed to provide

attachment sites for microorganisms, as well as being implicated in events such as leukocyte migration from the bloodstream. In diseases such as cystic fibrosis and cancer, altered mucin O-glycosylation has been observed (*38*, *39*). Despite a large body of research on mucin oligosaccharides (*40*–*43*), defining mucin glycoforms is an enormously challenging task but is nevertheless an essential prerequisite for addressing fundamental questions of mucin function in health and disease. Recent reports suggest that the Q-TOF MS promises to have a major impact in this area. For example, online LC-ES-MS/MS has been used to sequence sulfated mucin oligosaccharides purified from porcine large intestine (*44*). Information on core sequences and terminal blood-group antigenic determinants was obtained for 28 different oligosaccharides at a sensitivity appropriate for biological studies. A combination of Q-TOF and Edman sequencing of partially deglycosylated glycopeptides, prepared from the tandem repeat region of MUC1 derived from a breast cancer cell line, has provided evidence that all five potential glycosylation sites in the tandem repeat are O-glycosylation targets (*45*).

Immunosuppressive Glycoproteins in Pregnancy

Human development begins when a sperm binds to the extracellular matrix of the egg (the zona pellucida) and initiates a complex series of events leading to fertilization. The developing embryo is well tolerated in utero regardless of its genetic composition (fetal-maternal tolerance). Both of these phenomena are believed to involve carbohydrate-lectin recognition, but little is known about the putative recognition molecules (*46*, *47*). MS analyses of reproduction-associated glycoproteins are providing some of the first structural clues to the types of oligosaccharides that might play a part in these processes. Glycodelin-A (GdA) is a human glycoprotein found in amniotic fluid that has potent contraceptive and immunosuppressive activities (*48*). During the menstrual cycle, GdA is not expressed in the proliferative endometrium (the lining of the uterus) but increases substantially from the fourth postovulatory day, peaking around the 12th day. After implantation of the embryo, GdA synthesis is induced to very high levels (4 to 10% of total protein). GdA is also secreted into the amniotic fluid in concentrations sufficient to manifest immunosuppressive effects in vitro. The human embryo/fetus is thus surrounded by and bathed in a glycoprotein with potent immunosuppressive activities. The structure of GdA was determined by a combination of LC-ES-MS, FAB-MS, exoglycosidase digestion, and linkage analysis (*48*). GdA has two occupied sites for N-linked glycosylation, which carry substantially different ensembles of oligosaccharides. Among these are complex-type biantennary structures whose antennae are composed of fucosylated or sialylated LacDiNAc (GalNAcβ1-4GlcNAc) sequences, which are rare in other human glycoproteins (Fig. 4). Oligosaccharides with at least one fucosylated LacDiNAc sequence have been previously shown to be potent inhibitors of selectin-mediated adhesions (*49*). It is thus possible that GdA may mani-

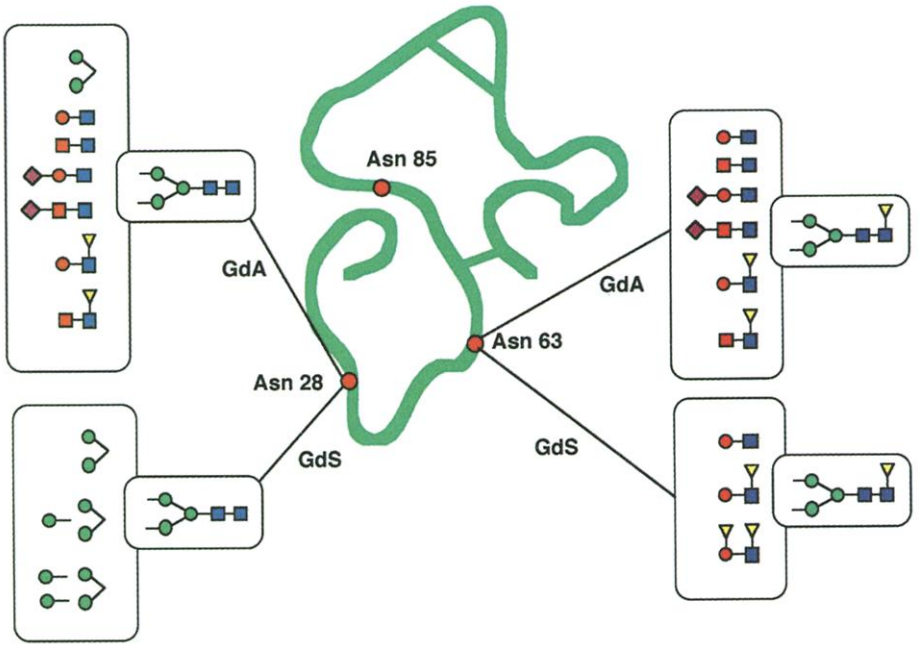

Fig. 4. Diagram of the glycodelin molecule showing the major antennae and core sequences found on the GdA and GdS oligosaccharides at the two occupied sites. Symbols are as in Fig. 2.

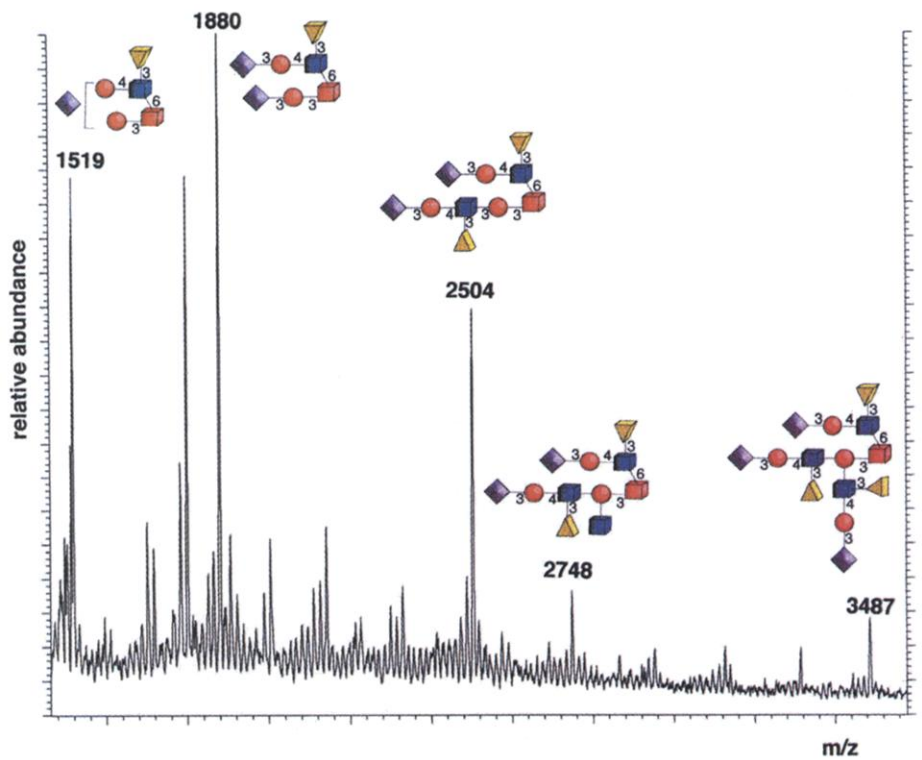

Fig. 5. Partial FAB mass spectrum of permethylated O-oligosaccharides from uromodulin showing molecular ions for core 2–type structures carrying up to three sialyl Le^x moieties (*52*). Symbols are as in Fig. 2.

fest some of its immunosuppressive effects by blocking selectin-mediated events. Immunological and molecular biological analyses have suggested that GdA is expressed in tissues other than the endometrium. Furthermore, a glycoprotein that cross-reacted with antibodies to GdA was detected in human seminal plasma over 15 years ago (*50*). The role of this seminal plasma form of glycodelin (GdS) remains unknown, but the expression of a potential contraceptive glycoprotein in the seminal plasma of the human male did not make physiological sense. Comparative structure-function studies of GdA and GdS (*51*) have helped to address this anomaly by revealing that GdS is glycosylated quite differently from GdA. None of the LacDiNAc structures in GdA is present in GdS, and the latter carries abundant, highly fucosylated antigenic determinants, which are absent in GdA (Fig. 4).

Tamm-Horsfall glycoprotein (THP), the major glycoprotein present in human urine, is an immunosuppressive molecule whose ability to inhibit T cell proliferation is increased 13-fold during pregnancy. Pregnancy-associated THP is called uromodulin. Mass spectrometric strategies have recently uncovered pregnancy-associated changes in the O-glycosylation of THP, which could account for the enhanced immunomodulatory effects of uromodulin (*52*). THP from nonpregnant females and males expresses primarily core 1–type O-oligosaccharides terminated with either sialic acid or fucose but not the sialyl Le^x epitope. In contrast, the O-oligosaccharides linked to uromodulin include unusual core 2–type oligosaccharides terminated with up to three sialyl Le^x sequences (Fig. 5).

Mapping Strategies

In recent years, our laboratory has been actively involved in developing rapid, high-sensitivity MS strategies for screening the types of oligosaccharides present in a diverse range of biological materials, including body fluids, secretions, organs, cultured cell lines, and whole parasites (*4*, *53*–*58*). These "mapping" strategies are invaluable for investigating the glycome (i.e., the full carbohydrate repertoire) of cells, tissues, organs, and so forth. They are based on the analysis of permethylated derivatives that, as described earlier, yield molecular ions at high sensitivity irrespective of the type of ionization and that, in the FAB-MS experiment, reliably afford characteristic fragment ions resulting from cleavage at HexNAc residues. The "map" of ions generated from a mixture of oligosaccharides defines all the terminal sequences present in the sample. Putative structures are assigned to each molecular ion on the basis of the usually unique oligosaccharide composition for a given mass, the terminal sequences defined by the fragment ions, and previous knowledge of N- and O-oligosaccharide biosynthesis. Assignments can then be confirmed by MS experiments on chemical and enzymatic degradations, the choice of which is guided by the sequence information provided by the screening experiments. Particularly interesting data have been obtained from experiments on α-mannosidase II null mice produced in Marth's laboratory at the Howard Hughes Medical Institute at the University of California in San Diego. We have identified a family of previously unknown oligosaccharides in kidney extracts from these mice (*59*). α-Mannosidase II regulates the biosynthetic branching pattern of N-oligosaccharides. Our data show that abundant hybrid oligosaccharides with unusual bisected structures carrying Le^x antennae and core fucosylation are present on kidney glycoproteins from the knockout mice. The Marth group has provided evidence that the mice exhibit an autoimmune disease that is similar to human systemic lupus erythematosus, and that anti-self antibodies with reactivity toward the N-oligosaccharides are induced in these mice (*59*).

Conclusion and Outlook

Structural glycobiology is undergoing rapid expansion owing to the application of modern ultrahigh-sensitivity mass spectrometric methods, including MALDI- and ES-MS on Q-TOF–type instruments capable of low-femtomole MS/MS analysis, which provide detailed sequence and site-of-attachment data for glycosylated proteins. The methods and strategies being developed are compatible with the problems of microheterogeneity commonly found, allowing characterization of even very complex minor components. The power of the MS approach in the solution of biological problems at the primary structural level is set to revolutionize protein and glycoprotein biology in the coming postgenomic era.

References and Notes

1. H. R. Morris *et al.*, *J. Biol. Chem.* **253**, 5155 (1978).
2. R. Kornfeld, S. Kornfeld *Annu. Rev. Biochem.* **54**, 631 (1985).
3. H. Schachter, in *New Comprehensive Biochemistry*, vol. 29a, *Glycoproteins*, J. Montreuil, J. F. G. Vliegenthart, H. Schachter, Eds. (Elsevier, Amsterdam, 1995), pp. 123–454.
4. In mammals this core fucose is linked to the 6-position of the GlcNAc, whereas in plants it occurs at the 3-position. Invertebrates are capable of fucosylating the core at either or both positions, and in rare instances a third fucose can be attached to the distal GlcNAc of the core (*60*).
5. A. Dell, H. R. Morris, S. M. Haslam, K.-H. Khoo, *Biochim. Biophys. Acta* **1455**, 353 (1999)
6. S. D. Rosen *Am. J. Pathol.* **155**, 1013 (1999).
7. W. Chai *et al.*, *Eur. J. Biochem.* **263**, 879 (1999).
8. H. R. Morris, *Nature* **286**, 447 (1980).
9. M. Barber, R. S. Bordoli, R. D. Sedgwick, A. N. Tyler, *J. Chem. Soc. Commun.* 325 (1981).
10. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, *Mass Spectrom. Rev.* **9**, 37 (1990).
11. M. Karas, A. Ingendoh, U. Bahr, F. Hillenkamp, *Biomed. Environ. Mass Spectrom.* **18**, 841 (1989).
12. A. Dell, *Adv. Carbohydr. Chem. Biochem.* **45**, 19 (1987).
13. K. Biemann, *Methods Enzymol.* **193**, 455 (1990).
14. A. L. Burlingame, R. K. Boyd, S. J. Gaskell, *Anal. Chem.* **66**, R634 (1994).
15. D. Chatterjee *et al.*, *Glycobiology* **3**, 497 (1993).
16. D. F. Hunt, J. R. Yates, J. Shabanowitz, S. Winston, C. R. Hauer, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6233 (1986).
17. H. R. Morris *et al.*, *Rapid Commun. Mass Spectrom.* **10**, 889 (1996).
18. H. R. Morris, T. Paxton, M. Panico, R. A. McDowell, A. Dell, in *Mass Spectrometry of Biological Materials*, B. Larsen, C. McEwen, Eds. (Dekker, New York, 1998), vol. 2, pp. 53–80.
19. J. H. Dawson, M. Guilhaus, *Rapid Commun. Mass Spectrom.* **3**, 155 (1989).
20. A. Shevchenko, A. Loboda, A. Shevchenko, W. Ens, K. G. Standing, *Anal. Chem.* **72**, 2132 (2000).
21. D. J. Harvey, R. H. Bateman, R. S. Bordoli, R. Tyldesley, *Rapid Commun. Mass Spectrom.* **14**, 2135 (2000).
22. P. Teng-umnuay *et al.*, *J. Biol. Chem.* **273**, 18242 (1998).
23. H. van der Wel *et al.*, unpublished data.
24. X. Tang, W. Ens, K. G. Standing, J. B. Westmore, *Anal. Chem.* **60**, 1791 (1988).
25. Y. Wada *et al.*, *Biochem. Biophys. Res. Commun.* **189**, 832 (1992)
26. K. Yamashita, T. Ohkura, H. Ideo, K. Ohno, M. Kanai, *J. Biochem. (Tokyo)* **114**, 766 (1993).
27. B. Coddeville, H. Carchon, J. Jaeken, G. Briand, G. Spik, *Glycoconjugate J.* **15**, 265 (1998).
28. T. Endo, D. Groth, S. B. Prusiner, A. Kobata *Biochemistry* **28**, 8380 (1989).
29. J. Collinge, K. C. Sidle, J. Meads, J. Ironside, A. F. Hill, *Nature* **383**, 685 (1996).
30. S. B. Prusiner *Science* **278**, 245 (1997).
31. R. A. Somerville *et al.*, *Nature* **386**, 564 (1997).
32. J. Safar *et al.*, *Nature Med.* **4**, 1157 (1998).
33. J. A. Mastrianni *et al.*, *N. Engl. J. Med.* **340**, 1630 (1999).
34. S. J. DeArmond *et al.*, *Neurology* **50**, 1271 (1998).
35. P. M. Rudd *et al.*, *Proc. Natl Acad. Sci. U.S.A.* **96**, 13044 (1999).
36. E. Stimson, J. Hope, A. Chong, A. L. Burlingame, *Biochemistry* **38**, 4885 (1999).
37. N. Stahl *et al.*, *Biochemistry* **32**, 1991 (1993).
38. T. F. Scanlin, M. C. Glick, *Biochim. Biophys. Acta* **1455**, 241 (1999).
39. J. Taylor Papadimitriou, O. J. Finn, *Immunol. Today* **18**, 105 (1997).
40. E. F. Hounsell, M. J. Davies, D. V. Renouf, *Glycoconjugate J.* **13**, 19 (1996).
41. A. Klein *et al.*, *Eur. J. Biochem.* **211**, 491 (1993).
42. D. L. Chance, T. P. Mawhinney *Carbohydr. Res.* **295**, 157 (1996).
43. H. van Halbeek *et al.*, *Glycobiology* **4**, 203 (1994)
44. K. A. Thomsson, H. Karlsson, G. C. Hansson, *Anal. Chem* **72**, 4543 (2000).
45. S. Müller *et al.*, *J. Biol. Chem.* **274**, 18165 (1999).
46. G. F. Clark *et al.*, *Hum. Reprod.* **11**, 467 (1996).
47. M. S. Patankar *et al.*, *Mol. Hum. Reprod.* **3**, 101 (1997).
48. A. Dell *et al.*, *J. Biol. Chem.* **270**, 24116 (1995).
49. B. W. Grinnell, R. B. Hermann, S. B. Yan, *Glycobiology* **4**, 221 (1994).
50. M. Julkunen *et al.*, *Arch. Androl.* **12** (suppl.), 59 (1984).
51. H. R. Morris *et al.*, *J. Biol. Chem.* **271**, 32159 (1996).
52. R. L. Easton, M. S. Patankar, G. F. Clark, H. R. Morris, A. Dell, *J. Biol. Chem.* **275**, 21928 (2000).
53. K.-H. Khoo, D. Chatterjee, J. P. Caulfield, H. R. Morris, A. Dell, *Glycobiology* **7**, 663 (1997).
54. S. M. Haslam *et al.*, *J. Biol. Chem.* **274**, 20953 (1999).
55. S. M. Haslam, G. C. Coles, H. R. Morris, A. Dell, *Glycobiology* **10**, 223 (2000).
56. R. L. Easton *et al.*, *J. Biol. Chem.* **275**, 7731 (2000).
57. M. Sutton-Smith, H. R. Morris, A. Dell, *Tetrahedron Assymetry* **11**, 363 (2000).
58. M. Morelle, S. M. Haslam, H. R. Morris, A. Dell, *Mol. Biochem. Parasitol.* **109**, 171 (2000).
59. D. Chui *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1142 (2001).
60. S. M. Haslam *et al.*, *J. Biol. Chem.* **271**, 30561 (1996).

61. We are grateful to R. Easton and S. Haslam for assistance in preparing the figures.