Proteomics in Functional Genomics
ed. by P. Jollès and H. Jörnvall

Analysis of glycoprotein-derived glycopeptides

Gerrit J. Gerwig and Johannes F. G. Vliegenthart*

Bijvoet Center for Biomolecular Research, Department of Bio-Organic Chemistry, Utrecht University, P.O. Box 80075, NL-3508 TB Utrecht, The Netherlands

Summary. Glycosylation of proteins represents one of the most important post-(co-)translational events in view of the ubiquity of the phenomenon. In most cases, the covalently linked glycans are involved in the functioning of these biomolecules in biological systems. Detailed information on the carbohydrate moieties including monosaccharide composition, anomeric configurations, type of glycosidic linkages and attachment sites at the protein is indispensable in describing the ultimate structure of a specific glycoprotein. This chapter presents a general strategy for the structural characterization of glycoproteins/glycopeptides focussed on the glycan part. Some of the techniques commonly used, like enzyme treatments, separation methods, chemical analyses, mass spectrometry and nuclear magnetic resonance spectroscopy are briefly reviewed.

1. Glycoproteins and their glycans

In vivo glycosylation of proteins, that is the covalent attachment of one or more carbohydrate chains (glycans) at specific sites to a protein backbone, is a common phenomenon shared by most organisms [1, 2]. The glycans are linked to side chains of the constituting amino acid residues and are distinguished by the type of linkage atom involved as N-, O-, C-, or S-glycans. For the functioning *in vivo* of these biomacromolecules the glycans are essential. Some significant biological roles of these carbohydrate units include location of a protein within the cell, protection of the protein against proteolytic attack, control of the lifetime of circulating cells and glycoproteins, induction and maintenance of the spatial conformation in a biologically active form, facilitation of the extracellular secretion and ultimate fate, direction and modulation of the immune response, and the provision of ligand structures for cell recognition/interaction [3].

Glycosylation of a protein is a post-(co-)translational nontemplate event and takes place in the endoplasmic reticulum and Golgi complex of the cell [4]. The glycans of glycoproteins are displayed as secondary gene products.

The biosynthesis of N-glycans consists of two stages: first, an *en bloc* transfer of a lipid-linked preassembled oligosaccharide to the nascent polypeptide chain and, in most cases, its processing/trimming to a small oligosaccharide core unit, and secondly, a final glycosylation by sequential addition of monosaccharide residues. These monosaccharides are transferred

* To whom correspondence should be sent.

by specific glycosyltransferases mainly from nucleotide sugar donors [5]. Multiple factors (some still unknown) play a role in directing and regulating glycosylation. For instance, to obtain N-glycosylation, the acceptor amino acid Asn must be located in an Asn-X-Ser/Thr/(Cys) sequence within the polypeptide (X can be any amino acid except Pro) [6]. However, Asn in this sequence is not always glycosylated [7], and in case of effective glycosylation, there may be a preference for the Asn-X-Thr sequon.

The biosynthesis of O-glycans consists exclusively of sequential transfers of single monosaccharides from nucleotide sugar donors. A specific polypeptide sequence around the acceptor amino acids Thr or Ser (present in a β-turn) does not appear to be required for glycosylation, although proline (Pro) residues are often abundantly present in combination with O-glycosylation [8]. Furthermore, it appears that charge distribution (but not charge density) of flanking amino acids within the primary structure of a protein plays a role in determining the extent to which solitary Thr/Ser residues are glycosylated [9].

The degree of glycosylation and ultimate structures of the mature N,O-glycans are mostly species-, tissue-, organ- and even cell-specific [10] and, in some cases, dependent on disease states. Finally, glycoproteins may contain from 0.4% to more than 80% (by mass) carbohydrate.

In general, four classes of monosaccharides are present as constituents in glycoprotein glycans of higher organisms: (i) neutral sugars: D-mannose (Man), D-galactose (Gal), D-glucose (Glc), L-fucose (Fuc), L-arabinose (Ara), D-xylose (Xyl); (ii) amino sugars: *N*-acetyl-D-glucosamine (GlcNAc), *N*-acetyl-D-galactosamine (GalNAc): (iii) uronic acids: D-glucuronic acid (GlcA), D-galacturonic acid (GalA), L-iduronic acid (IdoA); and (iv) sialic acids: *N*-acetylneuraminic acid (Neu5Ac), *N*-glycoloylneuraminic acid (Neu5Gc), 2-keto-3-deoxynonulosonic acid (Kdn). In addition, noncarbohydrate substituents like methyl, acetyl, lactoyl, pyruvyl, sulfate or phosphate groups occur. Furthermore, some rare monosaccharides can be present, in particular, in micro- and lower organisms [1].

The glycosylation of glycoproteins can be divided into three main classes: (i) N-glycosylation, the linkage between carbohydrate and the amide nitrogen of Asn; (ii) O-glycosylation, the linkage between carbohydrate and the hydroxyl oxygen of Thr or Ser; and (iii) the glycosylphosphatidylinositol (GPI) anchor, linked to the C-terminal amino acid of a protein via ethanolamine phosphate. However, other types of linkages have been found recently, including S- and C-glycosylation. Most of the known carbohydrate-protein linkage types are depicted in Table 1.

In nature, glycoprotein glycan structures are not randomly synthesized and can be ordered into families with similar structures and common oligosaccharide sequences. The major types of N-linked carbohydrate chains are summarized in Figure 1. Likewise, the mucin-type of O-linked carbohydrate chains are generally constructed from a limited number of core structures as shown in Figure 2. From these basic structures, complicated

Table 1. Carbohydrate-amino acid linkage types found in glycoproteins (the three most often occurring linkages are indicted in boldface)

Class	Linkage	Class	Linkage
N-linked		*Other O-linked*	
	GlcNAc → Asn		Glc → Tyr
	GalNAc → Asn		Gal → OH-Lys
	Glc → Asn		L-Ara → OH-Lys
	L-Rha → Asn		L-Ara*f* → OH-Pro
O-linked			Gal → OH-Pro
	GalNAc → Ser/Thr		Gal → OH-His
	GlcNAc → Ser/Thr		GlcA → OH-Trp
	Man → Ser/Thr		GlcA → OH-Phe
	L-Fuc → Ser/Thr		GlcA → OH-Ser
	Gal → Ser/Thr	*S-linked*	
	Glc → Ser		Gal → Cys
	Xyl → Ser		Glc → Cys
ADP ribosylation		*C-linked*	
	ADP-Rib → N^{ε}-Arg		Man → Trp
	ADP-Rib → N^{δ}-Asn	*Glycation*	
	ADP-Rib → N^{1}-His		Glc → Lys
	ADP-Rib → *O*-*C*(*O*)Glu		Rib → Lys
	ADP-Rib → *S*-Cys	*Amide Bond*	
Phosphoglycosylation			GlcA/GalA(6 → N^{α})Lys
	GlcNAc1-PO_4 → Ser		GlcA/GalA(6 → N^{α})Thr/Ser
	Man1-PO_4 → Ser		GlcA/GalA(6 → N^{α})Ala
	Xyl1-PO_4 → Ser/Thr		MurNAc(3 → N^{α})Ala
Glypiation			
	Protein-C(O)-NH-$(CH_2)_2$-PO_4 → 6Man (GPI)		

glycans can be built up in glycoproteins due to elongation and branching. A glycoprotein often contains more than one glycosylation site, and on a single glycoprotein the glycan structures at each of the sites can comprise a family of glycans. This phenomenon is known as microheterogeneity. However, the ensemble of glycoforms has a fairly constant composition for a given protein from a specific source. Nowadays, commercial kits are available for sensitive detection of glycosylated proteins.

The occurrence of glycopeptides in nature, for instance in urine, is predominantly due to degradation of glycoproteins. However, in insects, some of the antibacterial proline-rich peptides (2–3 kDa) are O-glycosylated [11]. Several antibiotics (e.g. vancomycin, teicoplanin, ristocetin, eremomycin), isolated from microorganisms, have proved to be complex glycopeptides containing rare amino sugars [12].

The term "glycopeptide" should not be confused with peptidoglycan. Peptidoglycan, found particularly in bacterial cell walls, has a polysaccharide backbone of alternating GlcNAc and *N*-acetylmuramic acid

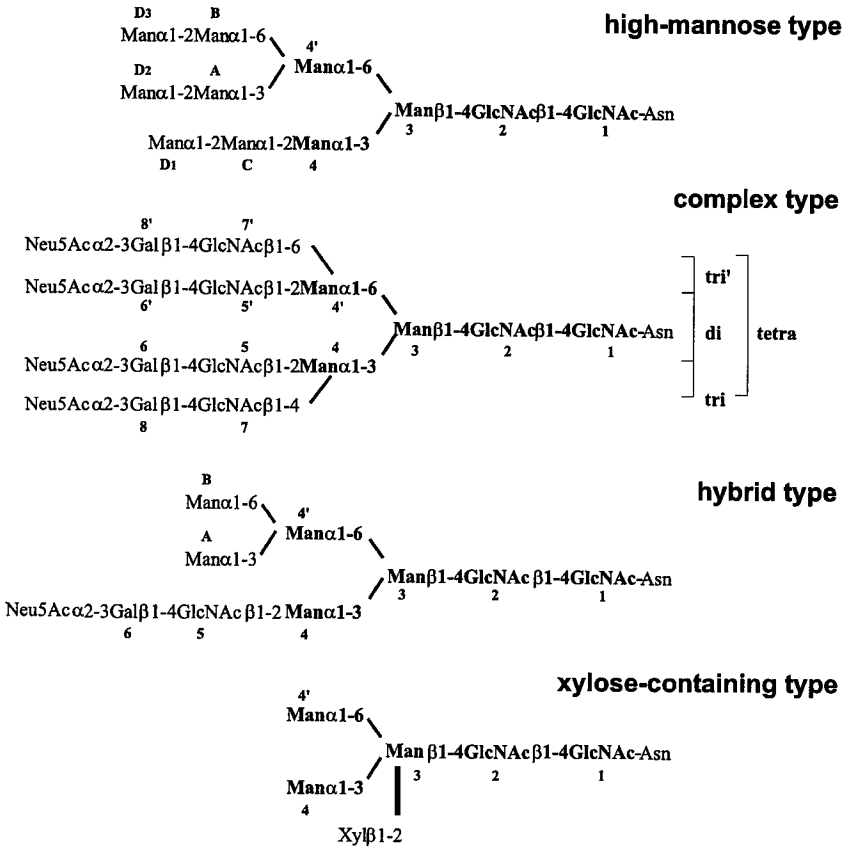

Figure 1. Examples of the four types of N-linked glycans of glycoproteins. The common pentasaccharide core is shown in boldface. For the complex type the di-, tri-, tri'- and tetra-antennary subtypes are indicated. The standard notation of the monosaccharide residues has been included.

(MurNAc). Some of the MurNAc residues contain a tetrapeptide side chain, and the polysaccharide is cross-linked with peptide bridges [10].

2. Glycoprotein analysis

In higher animals most of the proteins essential for life, like immunoglobulins, enzymes, cell-membrane receptors and hormones, are in fact glycoproteins, and their functioning greatly depends on the glycan part. To investigate the metabolism and molecular biology of these biomolecules, knowledge of their precise structures and in particular of the glycan struc-

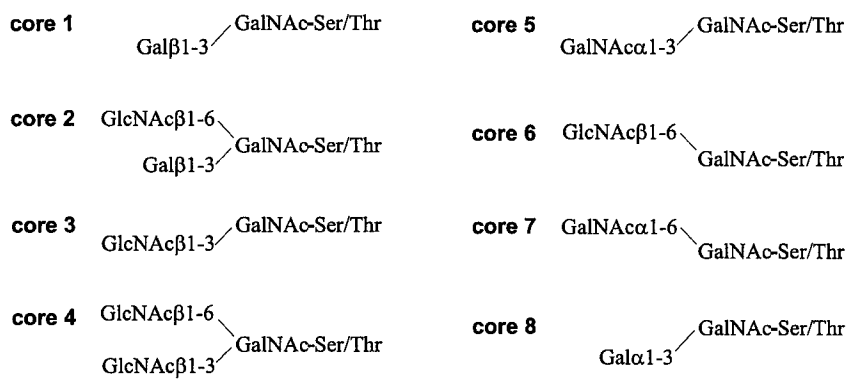

Figure 2. Core structures of mucin-type O-linked glycans of glycoproteins. The core structures can be extended with Gal, GalNAc, GlcNAc, Fuc and sialic acid residues.

tures is indispensable. Furthermore, the impressive increase of the number of biotechnologically produced glycoprotein-therapeutics demands accurate, fast and reliable analysis methods for determination of composition and structure [13–16].

Structural analysis of glycoproteins involves several levels of detail: (i) the amino acid sequence of the protein; (ii) the number, nature and position of glycosylated amino acids; (iii) the identity and quantity of monosaccharide residues present in the protein; (iv) the structural class(es) of oligosaccharides present in the protein; (v) the nature and structural heterogeneity of the glycans attached at specific sites in the protein; (vi) the monosaccharide sequence, branching and glycosidic linkage types of each glycan; and (vii) the secondary and spatial structure of the glycoprotein, in particular, the protein and the carbohydrate part, and their mutual influence. For these studies it is essential that adequate starting material is available, which means that much care has to be bestowed on the isolation and purification of the glycoprotein. A glycoprotein that is apparently homogeneous with respect to the protein backbone may nevertheless comprise a collection of glycoforms.

The structural characterization of glycoproteins is generally considered as a relatively difficult problem, in particular because it requires the use of highly sophisticated instrumentation. In view of the diversity and complexity of the glycans, their primary structure analysis has still not reached the level of routine analysis and remains a specialized and laborious endeavour depending on the combined use of several physical, chemical and biochemical techniques [17].

The discussion of the analysis of glycoproteins will be divided into three topics, namely, (i) the protein/peptide part, (ii) the glycan part and (iii) the

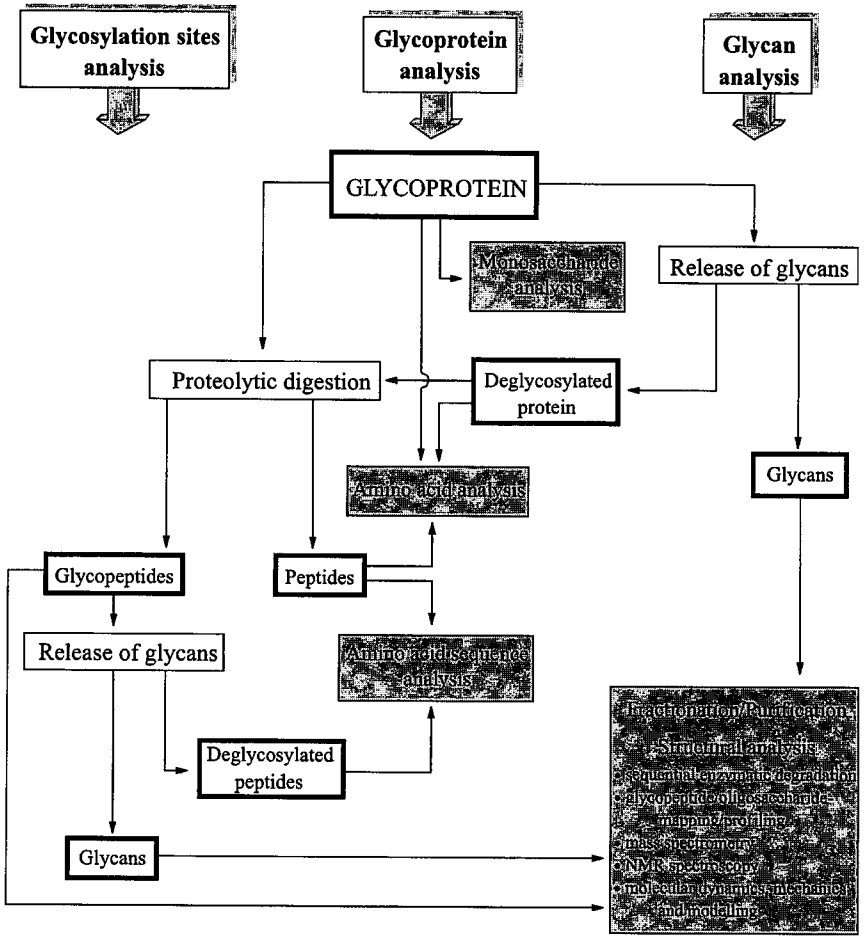

Figure 3. A general strategy for the analysis of glycoproteins/peptides.

determination of the glycosylation sites. Moreover, it will be focused on the preparation and analysis of glycopeptides. A general strategy for glycoprotein glycan analysis is presented in Figure 3.

2.1. Protein/peptide analysis

The presence of carbohydrate chains linked to the protein is no obstacle for the determination of the amino acid composition. Amino acid analysis is generally performed, after acid hydrolysis, on an Amino Acid Analyzer using ion-exchange chromatography. Absorbance is detected by post-column treatment with ninhydrin. However, it has to be noted that amino-

sugars, if present in large amounts, can give rise to peaks in the chromatogram coeluting with certain amino acids.

For the determination of the complete amino acid sequence, the protein is usually partially hydrolyzed into peptides by proteolytic enzymes [this book]. In this case, covalently linked carbohydrates, in particular when adjacently clustered at specific regions of the polypeptide backbone (e.g. in mucins), can seriously hamper the action of the proteases. Preceding excision of the carbohydrate chains without destroying the protein or peptide may be achieved chemically (e.g. O-glycans by mild alkaline borohydride treatment) and/or enzymatically (e.g. N-glycans by PNGase) (*vide infra*). Also, the use of anhydrous trifluoromethanesulfonic acid (TFMSA) has been found to be successful [18]. In this case, N- and O-glycans are cleaved nonselectively, leaving the primary structure of the protein intact for sequence analysis, but the carbohydrate chains are destroyed. The best results to deglycosylate heavily O-glycosylated proteins have been obtained by a combined approach using mild TFMSA treatment, followed by periodate oxidation and alkaline treatment [19].

2.2 Glycan analysis

In order to get an overview of the carbohydrate portion of a protein, the intact glycoprotein is usually first subjected to a monosaccharide analysis. The determination of the monosaccharide composition and carbohydrate content is currently established after methanolysis and analysis of volatile sugar derivatives by gas chromatography(-mass spectrometry) (GC-MS) [20]. Recently, high-pH anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) has been introduced for monosaccharide analysis using underivatized monosaccharides obtained by trifluoroacetic acid (TFA) or HCl hydrolysis [21], but different hydrolysis conditions are needed for quantitative determination of different types of monosaccharides.

Once the monosaccharide compostion of the glycosylation is known, putative structural classes of glycans present on the glycoprotein can be inferred. The detection of GlcNAc, Man, Fuc and Neu5Ac indicates the presence of N-glycans, whereas detection of relatively large amounts of GalNAc indicates possibly O-glycans. These initial clues are beneficial for further tackling the structural analysis of the glycans.

Since glycoproteins contain mostly multiple glycosylation sites, and the oligosaccharide chains usually exhibit considerable structural heterogeneity, it is hardly possible to analyze complete glycan structures at the level of the intact glycoprotein. The exhaustive nonselective proteolytic cleavage of the protein and subsequent analysis of the isolated glycopeptides is a good approach to establish the glycan structures [22]. In this case, it is important that the number of glycosylation sites comprised in the structure is known.

An alternative practical approach is to liberate the carbohydrate chains, then to separate the oligosaccharide pool from the deglycosylated protein, and to fractionate the oligosaccharides. Glycans can be chemically and/or enzymatically released from the protein backbone (Tab. 2). Hydrazinolysis [23] is frequently applied to liberate N-glycans, and depending on graded reaction conditions, sequential release of O-glycans followed by N-glycans may be possible [24]. Other chemical methods are reductive deamination [25] and trifluoroacetolysis [26]. It must be noted that chemical treatments may lead to some degradation of the carbohydrate portion and (partially) destroy the polypeptide backbone. The release of O-glycans can be realized with alkaline borohydride treatment (β-elimination reaction) [27].

In the enzymatic approach two types of enzymes are currently used, namely, the endo-β-*N*-acetylglucosaminidases (endoglycosidases) and the peptide-N^4-(*N*-acetyl-β-D-glucosaminyl) asparagine amidases (PNGases) to liberate N-glycans. For both types of enzymes, the efficiency of release depends on their specificity and the type of oligosaccharide structures [28]. For application to glycopeptides, introduction of restrictions by the length of the peptide backbone must be taken into account (Tab. 2). The enzymatic release of O-glycans by endo-*N*-galactosaminidases (O-glycanases) is very restricted because of their limited specificities and therefore rarely used [29]. Occasionally, it may be necessary first to denaturate the glycoprotein to make possible complete oligosaccharide release. Care should always be

Table 2. Chemical and enzymatic release of glycans from glycoproteins/peptides

Chemical Procedure	Specificity/target	Comments
TFMSA treatment (trifluoromethane sulfonic acid)	N- and O-glycans	– glycans are destroyed
Hydrazinolysis (eventually with concomitant re *N*-acetylation and reduction)	N- and O-glycans (depending on conditions)	– protein is destroyed – loss of *N*- and *O*-acyl groups – partial loss of sulfate and phosphate
Reductive deamination (nitrous acid hydrolysis)	GPI anchor	– cleavage between $GlcNH_2$ and inositol
Alkaline borohydride treatment (β-elimination reaction) (often with concomitant reduction)	O-glycans (N-glycans to a certain extent)	– peeling reactions – loss of *O*-acetyl groups – not active on N-terminal Thr/Ser glycan – possible destruction of protein
Trifluoroacetolysis (mixture TFA:TFAA 1:100)	glycosylamine linkage is cleaved (peptide bonds)	– oligosaccharides are released as TFA derivatives – Fuc attached to GlcNAc1 is destroyed – low recovery

Table 2 (continued)

Enzymatic Enzyme	Abbreviation	Source	Specificity
Endoglycosidase D (EC 3.2.1.96) Endo-β-*N*-acetyl glucosaminidase D	Endo D	*Diplococcus pneumoniae*	– N-glycans – cleavage between GlcNAc residues of chitobiose-Asn – high-mannose ($Man_5GlcNAc_2$) types – not active on intact complex and hybrid types
Endoglycosidase F (EC 3.2.1.96) (mixture of endo-glycosidases F_1, F_2, F_3) Endo-β-*N*-acetyl glucosaminidase F	Endo F	*Flavo-bacterium meningo-septicum*	– N-glycans – cleavage between GlcNAc residues of chitobiose-Asn – high-mannose, diantennary complex and hybrid types
Endoglycosidase H (EC 3.2.1.96) Endo-β-*N*-acetyl-glucosaminidase H	Endo H	*Streptomyces griseus* (*S. plicatus*) (recombinant in *E. coli*/*S. lividans*)	– N-glycans – cleavage between GlcNAc residues of chitobiose-Asn – high-mannose (not sulfated) types – some hybrid types – not active on complex types
Peptide-N-Glycosidase F (EC 3.2.2.18) (EC 3.5.1.52) peptide-N_4-(*N*-acetyl-β-D-glucosaminyl) asparagine amidase Glycopeptidase F *N*-glycosidase *N*-glycanase	PNGase F	*Flavo-bacterium meningo-septicum* (recombinant in *E. coli*)	– N-glycans – cleavage between GlcNAc-Asn (Asn converted into Asp) – high-mannose, complex and hybrid types – not from a single or N/C-terminal Asn – hardly active when core Fucα1-3 present – substituents remain
Peptide-*N*-Glycosidase A (EC 3.5.1.52) peptide-N_4-(*N*-acetyl-β-D-glucosaminyl) asparagine amidase Glycopeptidase A *N*-glycosidase	PNGase A	Almond emulsine	– N-glycans – cleavage between GlcNAc-Asn – high-mannose, complex and hybrid types – not from a single or N/C terminal Asn – peptide backbone restrictions
Endo-α-*N*-acetyl-galactosaminidase (EC 3.2.1.97) (EC 3.2.1.110) O-Glycosidase O-glycan peptide hydrolase O-Glycanase	–	*Streptococcus* (*Diplococcus*) *pneumoniae* (*Alcaligenes* sp.)	– O-glycans – only cleavage of disaccharide Galβ1-3GalNAcα1- from Thr/Ser

taken to avoid glycan modifications, including loss of sialic acid residues, Fuc, and noncarbohydrate substituents.

A convenient strategy for the study of glycans of N,O-glycoproteins is based on cleavage of the N-glycans with PNGase-F, followed by alkaline borohydride release of the O-glycans from the remaining purified O-glycoprotein [30]. In case of minimal amounts, labeling [e.g. with 2-aminobenzamide (2-AB), aminonapthalenesulfonic acid (ANTS) or tritium] of the released glycans is useful to enable their detection in subsequent procedures.

2.3 Glycosylation site analysis

A suitable approach to characterize the glycosylation sites in a glycoprotein consists of degrading the protein backbone. Selective mild proteolytic digestion followed by fractionation of the formed glycopeptides is applied for this purpose. The location of the glycosylation sites in the polypeptide chain is determined by analysis of the amino acid sequence around each glycan attachment site of the glycopeptide and comparison of this sequence with the known amino acid sequence of the protein. Identification of the glycan structure and of the peptide chain are performed before and after removal of the glycan.

3. Glycopeptide analysis

In the framework of this review glycopeptides are defined as glycoconjugates that contain oligosaccharides still attached to a portion of the original peptide sequence (oligopeptide) of the glycoprotein. Preparation of glycopeptides from glycoproteins represents an important step in the determination of the glycan structures, but moreover it is essential for the determination of the glycosylation sites and of the site-specific (micro)heterogeneity. For the study of cell surface glycosylation, the preparation and analysis of glycopeptides is a general procedure [31]. Glycopeptides can chemically be prepared by cyanogen bromide (CNBr) cleavage of glycoproteins at the carboxylic side of Met [32]. Sometimes this treatment is necessary prior to proteolytic digestion of glycoproteins, which have a tendency to form aggregates [33].

3.1 Proteolytic digestion

Proteolytic digestion may be performed either on native or on denaturated glycoproteins and should be optimized for each glycoprotein. The size and type of glycopeptides formed are dependent on the amino acid sequence of

the protein, the glycan structures present in the protein and the specificity of the proteases being used. Some frequently used proteolytic enzymes are listed in Table 3.

Denaturation of proteins is often achieved by heat treatment and/or reduction of cystine bridges followed by carboxymethylation of the free SH groups [31].

Nonselective proteolysis with pronase, which is a mixture of endopeptidases, is often used to obtain glycans linked to one amino acid or glycopeptides with a very short (two or three amino acids) peptide chain. Proteases with restricted selectivity, such as trypsin or chymotrypsin, are used when a longer peptide chain is required to assess the different glycan structures at each of the glycosylation sites. Today, endoproteinase Glu-C hydrolysis is frequently used for glycosylation site analysis. However, care must be taken because often incomplete and aspecific proteolytic cleavages give rise to complex mixtures of peptides and glycopeptides, including the possibility that one glycosylation site can be represented by more than one glycopeptide. In O- and N,O-glycoproteins, multiple O-glycosylation sites may occur that are clustered in certain regions of the polypeptide backbone, thereby impeding the preparation of glycopeptides containing a single O-glycosylation site. Consequently, in heavily O-glycosylated proteins (e.g. mucins), the assignment of a particular glycan to a specific amino acid is difficult to achieve.

Table 3. Some proteolytic enzymes frequently used in glycoprotein studies

Enzyme	Source	pH optimum	Specificity	Comments
Carboxypeptidase A (EC 3.4.17.1) Carboxypoly-peptidase	bovine pancreas	7.0–8.0	– successive cleavage from C-terminus – slow for Gly, Asp, Glu, Cys	Zn metallo-protease – does not act at Arg, Lys, Pro(OH)
Carboxypeptidase B (EC 3.4.17.2) Protaminase	porcine/hog pancreas	7.0–9.0	– successive cleavage from C-terminus of basic amino acids (Lys, Arg)	Zn metallo-protease
Chymotrypsin (α) (EC 3.4.21.1)	porcine/ bovine pancreas	7.5–8.5	– C-terminal bonds of Tyr, Phe, Trp – slow for Met, Leu, Ala, Asp, Glu	Serine endopeptidase
Endoproteinase Arg-C (EC 3.4.21.40)	Mouse sub-maxillaris glands *Clostridium histolyticum*	8.0–8.5	– carboxylic side of Arg	Serine protease
Endoproteinase Asp-N (EC 3.4.24.33)	*Pseudomonas fragi* (mutant) *Clostridium histolyticum*	7.0–8.0	– amino side of Asp (Glu) and cysteic acid	Metalloprotease

Table 3 (continued)

Enzyme	Source	pH optimum	Specificity	Comments
Endoproteinase Glu-C (EC 3.4.21.19) V-8 protease proteinase V8	*Staphylococcus aureus* V8	4.0 and 7.8	– carboxylic side of Glu (and Asp)	Serine protease
Endoproteinase Lys-C (EC 3.4.99.30) (EC 3.4.21.50)	*Lysobacter enzymogenes*	8.5–8.8	– carboxylic side of Lys	Serine protease
Papain (EC 3.4.22.2)	*Carica papaya*	6.0–7.0	– Arg, Lys, Glu, His, Gly, Tyr – total hydrolysis on prolonged incubation – does not act at acidic residues	Cysteine endopeptidase (thiol protease)
Pepsin (EC 3.4.23.1)	pig gastric mucosa	2.0–4.0	– preferentially carboxylic side of Phe, Met, Leu, Trp – Tyr-X and X-Val/Ala/Gly are relatively resistant	Aspartic endopeptidase (carboxyl protease) (acid protease)
Pronase (EC 3.4.24.4) Actinase E	*Streptomyces griseus*	7.5–8.0	– total hydrolysis	Mixture of several unspecific endo- and exoproteases
Subtilisin (EC 3.4.21.14) (EC 3.4.21.62) Alcalase/Nargase	*Bacillus subtilis* *B. licheniformis*	7.0–8.0	– total hydrolysis – preferentially Asp, Glu, Ala, Gly, Val	Serine endopeptidase
Thermolysin (EC 3.4.24.4) (EC 3.4.24.27)	*Bacillus thermoproteolyticus*	7.0–9.0	– low cleavage specificity Ile, Leu, Met, Phe, Trp, Val, Ala	Zn metalloendopeptidase thermostable 4–80°C
Trypsin (EC 3.4.21.4)	bovine/hog pancreas	7.5–8.5	– carboxylic side of Arg, Lys	Serine endopeptidase

3.2 Isolation and purification of glycopeptides

A large variety of high-resolution separation techniques are available today for the isolation, fractionation and purification of glycopeptides (and oligosaccharides), as summarized in Table 4. A combination of these methods is advisable, and often necessary, to obtain homogeneous compounds. After nonspecific proteolytic treatment of the glycoprotein, the first step is the separation of the glycopeptides from the bulk of small peptides, amino acids and salts. This can be achieved by size exclusion (gel filtration) chromatography. To facilitate the isolation and analysis procedures,

Table 4. Separation techniques frequently used for glycoproteins/peptides and their glycans

Separation method	Abbreviation	Separation parameter	Target compounds (analysis)
Chromatography			
gel filtration size exclusion gel permeation	GFC SEC GPC	molecular size and shape (hydrodynamic volume)	– glycoproteins – glycopeptides – oligosaccharides (size profile)
ion-exchange anion-exchange cation-exchange (HPLC)	IEC AEC CEC	net charge	– glycoproteins – glycopeptides – sialyloligosaccharides (charge profile)
high-pH anion-exchange chromatography	HPAEC	charge/size (oxy-anions) hydrogen bonds dipole interaction	– (glycopeptides) – monosaccharides – oligosaccharides (oligosaccharide mapping)
normal-phase high-performance liquid chromatography	NP-HPLC	hydrogen bonding and/or partition (hydrophilic interaction) (size/polarity)	– neutral (acidic) oligosaccharides (high-mannose type N-glycans)
reversed-phase high-performance liquid chromatography	RP-HPLC	hydrophobic interaction	– glycopeptides – (permethylated) oligosaccharides (sialic acid analysis)
affinity chromatography on immobilized lectins	AC	biospecific interaction	– (glycoproteins) – glycopeptides – oligosaccharides
Electrophoresis			
preparative paper electrophoresis (high/low voltage)	PPE	charge	– (glyco)peptides – oligosaccharides
(sodium dodecyl-sulfate)-polyacrylamide gel electrophoresis	(SDS)/ PAGE FACE™	molecular mass/ charge	– glycoproteins – deglycosylated proteins (N-glycan analysis)
isoelectric focusing (polyacrylamide gel)	IEF	charge/pI (pH gradient)	– glycoproteins
capillary zone electrophoresis	CZE	mass/charge ratio (pH/ionic strength)	– glycopeptides – charged oligosaccharides
capillary gel electrophoresis (polyacrylamide)	CGE	molecular size charge/mass ratio constant	– glycoproteins – glycopeptides – oligosaccharides (glycoform mapping)
affinity capillary electrophoresis	ACE	charge/biomolecular interaction	– lectin-sugar binding

N-terminal amino acids in the glycopeptides can be labeled by *N*-(^{3}H)- or *N*-(^{14}C)-acetylation or *N*-dansylation, and elutions be performed with volatile buffers.

Reversed-phase high-performance liquid chromatography (RP-HPLC) is most commonly used for glycopeptide purification [34]. The separation is based on the hydrophobicity of the peptide part. Hydrophobicity can be

Table 5. Lectins frequently used for carbohydrate identification and separation in glycoprotein studies

Lectin	Abbreviation	Source	Specificity
Agaricus bisporus	ABA	mushroom	Galβ1-3GalNAc
Anguilla anguilla	AAA	fresh water eel	Fucα, Me-C2/C3 Fuc
Arachis hypogaea	PNA	peanut	Galβ, Galβ1-3GalNAc, Galβ1-4Glc
Canavalia ensiformis	Con A	Jack bean	Manα, Glcα, GlcNAcα, Me-αMan, branched mannoses
Datura stramonium	DSL	Jimson weed, thorn apple	GlcNAcβ, GlcNAcβ 1-4GlcNAc oligomers, Galβ1-4GlcNAc, *N*-acetyllactosamine repeats
Dolichos bifloris	DBA	horse gram	terminal GalNAcα, GalNAcα1-3GalNAc
Erythrina cristagalli	ECA	coral tree	Galα/β, GalNAcα/β, Galβ1-4GlcNAc
Galanthus nivalis	GNA	snowdrop bulb	Manα, Manα1-3Man, terminal Manα1-3
Glycine max	SBA	soybean	terminal GalNAcα/β, GalNAcα1-3Gal
Helix pomatia	HPA	Roman snail	GalNAcα/β
Lens culinaris	LcH	lentil	Manα, Glcα, GlcNAcα, branched mannoses with αFuc, Fucα1-6coreGlcNAc
Lotus tetragonolobus	LTA	asparagus pea	Fucα, Fucα1-2-Galβ1-4-[Fucα1-3]GlcNAc
Maackia amurensis	MAA	–	Sialic acid, Neu5Acα2-3Gal, (Galβ1-4Glc)
Maclura pomifera	MPA	osage orange	Galα, GalNAcα
Narcissus pseudonarcissus	NPA	daffodil	terminal and internal Manα
Pisum sativum	PSA	garden pea	Manα, Glcα
Sambucus nigra	SNA	elderberry	Galβ, Sialic acid, Neu5Acα2-6Gal(NAc)
Triticum vulgaris	WGA	wheat germ	GlcNAcβ, sialic acid, GlcNAcβ1-4GlcNAc
Ulex europaeus	UEA-1	gorse, furze	Fucα, Fucα1-2Galβ1-4GlcNAc

increased by modification of the amino-terminus with *tert*-butyloxycarbonyl tyrosine. Since peptide sequences surrounding each glycosylation site, in particular for N-glycosylation, are usually quite different, RP-HPLC often resolves glycopeptides from multiple glycosylation sites derived from the same glycoprotein. The glycosylation-site-specific glycopeptide mixture can be further fractionated on the basis of carbohydrates. Separation based simultaneously on peptide and oligosaccharide structures can also be accomplished with RP-HPLC [35]. Anion-exchange (AE) HPLC has been used to separate glycopeptides, but the method is highly dependent on the structure of the peptide portion. Elutions are usually monitored by ultraviolet (UV) absorbance.

For the detection of the glycopeptide fractions among the many peptide fractions in a chromatogram, different strategies are applied: (i) monosaccharide analysis or thin-layer chromatography (TLC)-spottest on each collected peak; (ii) amino acid sequence analysis, which usually indicates the presence of amino acid-linked oligosaccharides, on each collected peak; (iii) tandem MS/MS; (iv) enzymatic cleavage of the glycans in the (glyco)peptide pool causing a shift in the HPLC elution time; and (v) lectins conjugated to antibodies or enzymes to detect the presence of carbohydrates in collected fractions [36]. The properties of lectins to bind specifically to a certain sugar and/or sugar sequence in oligosaccharides and glycoconjugates (Tab. 5) are applied in affinity chromatography to fractionate and purify glycopeptides by use of immobilized lectin columns [37, 38], and more recently, by affinity capillary electrophoresis for microscale analyses [39]. Improvements in chromatographic techniques have been major factors in carbohydrate-research advances.

4. Structural analysis of glycans

The two main methods to determine the primary structure of carbohydrate chains are mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR). They are applied to purified oligosaccharides and glycopeptides, but when mixtures are not too complicated, sometimes the analysis of mixtures is possible. In some cases, additional evidence from chemical and/or enzymatic analytical methods is still necessary. This holds in particular for the unambiguous characterization of novel structures. The information from more than one analytical method (Tab. 6) is needed to arrive at conclusive evidence on the identity of a compound.

4.1 Chemical and enzymatic methods

It is evident that GC-MS monosaccharide composition analysis of the purified glycopeptides or oligosaccharides already provides information

Table 6. Methods of obtaining information about specific carbohydrate features of glycoprotein glycans

Information	Methods
Carbohydrate content, composition, D/L-configuration	colorimetric determinations, GLC-monosaccharide analysis, GLC-absolute configuration determination, NMR spectroscopy
Molecular mass of glycoprotein/glycan (presence of glycosylation)	gel filtration chromatography, mass profile by FAB/ES/MALDI-mass spectrometry, SDS/PAGE (before/after enzyme treatment)
Nature of carbohydrate-peptide linkage (N/O)	proteolytic digestion, amino acid analysis, examination of alkali lability, hydrazinolysis
Type of glycans (high-mannose, complex, hybrid), glycoforms	GLC-monosaccharide analysis, size/charge profile analysis, capillary electrophoresis
Number/proportions of glycans present	size/charge profile analyss, mapping by HPLC, HPAEC, FACE, MALDI-MS
Sequence of monosaccharide residues	digestion by exoglycosidases, partial hydrolysis, NMR spectroscopy, mass spectrometry
Positions of glycosidic linkages	methylation analysis/GLC-MS, FAB-MS, NMR spectroscopy
Anomeric configuration	digestion by exoglycosidases, NMR spectroscopy
Certain structural determinants	antibody responses, endo/exoglycosidases, affinity chromatography/electrophoresis (lectins)
Type of charged substituents	size/charge profile analysis, HPLC, HPAEC, NMR spectroscopy
Spatial structure of glycoprotein/peptide/glycan	X-ray analysis, 2D/3D NMR spectroscopy, molecular dynamics, mechanics and modelling

about the type of structures that can be expected. The absolute configuration (D or L) of the monosaccharide residues can be established by GC of the corresponding (–)2-butyl glycosides [40].

Methylation analysis is a reliable chemical method for the elucidation of the substitution pattern and of the ring size of the individual monosaccharide residues in an oligosaccharide. After methylation [41] of all free hydroxyl groups, the monosaccharides liberated by hydrolysis are reduced, acetylated, and subsequently analyzed by GC-MS as partially *O*-methylated alditol acetates. The positions of the *O*-acetyl and *O*-methyl groups, as deduced from specific MS fragmentation, indicate the substitution pattern in the monosaccharide [42].

Information about the monosaccharide sequence in a glycan chain can be obtained enzymatically by successive exoglycosidase digestion [43]. In this approach monosaccharide residues are stepwise cleaved from the nonreducing terminus of the oligosaccharide by hydrolases (e.g. sialidase, α-mannosidase, β-galactosidase, etc.) which are highly specific towards

their substrate including the anomeric configuration. By consequence, the change in effective size of the glycan can be analyzed by gel filtration chromatography under standard conditions before and after the enzyme treatment. The enzymatic sequence analysis of N-glycans has been automated into a RAAM (Reagent Array Analysis Method) GlycoSequencer [44].

Because in many instances only pico/nanomole amounts of material are available, much effort is focused on the development of profiling techniques based on high-performance separation procedures like capillary electrophoresis [45, 46], HPLC mapping [34, 47] and fluorophore assisted-carbohydrate electrophoresis (FACE) [48].

4.2 Mass spectrometry

Electron impact ionization mass spectrometry (EI-MS) in combination with GC-MS has already been used for a long time for the identification of monosaccharide derivatives in monosaccharide and methylation analysis [20]. This technique remains a standard for carbohydrate composition and linkage analysis. Chemical ionization mass spectrometry (CI-MS) can be used to increase sensitivity and detection of the molecular ion.

In recent years, several new mass spectrometric methods using soft ionization techniques like fast atom bombardment (FAB-MS), electrospray ionization (ES-MS) and matrix-assisted laser desorption ionization (MALDI-MS) have been applied for the study of glycoproteins, oligosaccharides and glycopeptides because of their high sensitivity. Structural information about branching pattern, number and length of branches and sequence in terms of hexoses, deoxyhexoses, *N*-acetylhexosamines and sialic acids, as well as the net molecular mass of the glycopeptide/protein, can be obtained.

FAB-MS is useful for defining the glycosylation sites at the polypeptide backbone and delineating the complete sequences of suitably derivatized individual glycans in glycopeptides from fragment ions [49, 50]. Glycopeptides and glycans have mostly been studied after permethylation. A typical example of FAB-mass ion fragmentation of an N- and O-glycan is given in Figure 4.

ES-MS has proven its power for the analysis of glycopeptides and glycoproteins, since the peptide/protein portion of the molecule usually provides the necessary multiple charge sites, ensuring mass-to-charge ratios within the range of the quadrupole analyzer. Interfacing ES mass spectrometers with liquid chromatography systems (e.g. HPLC-MS) allow rapid separation and analysis of mixtures of peptides and glycopeptides obtained after proteolytic digestion of a glycoprotein [51, 52]. Recent examples to highlight the applicability of MS technology are demonstrated by the structural analysis of the glycans of glycodelin-A, a human endometrial glycoprotein, glycodelin-S, a human seminal plasma glycoprotein [53], and horseradish

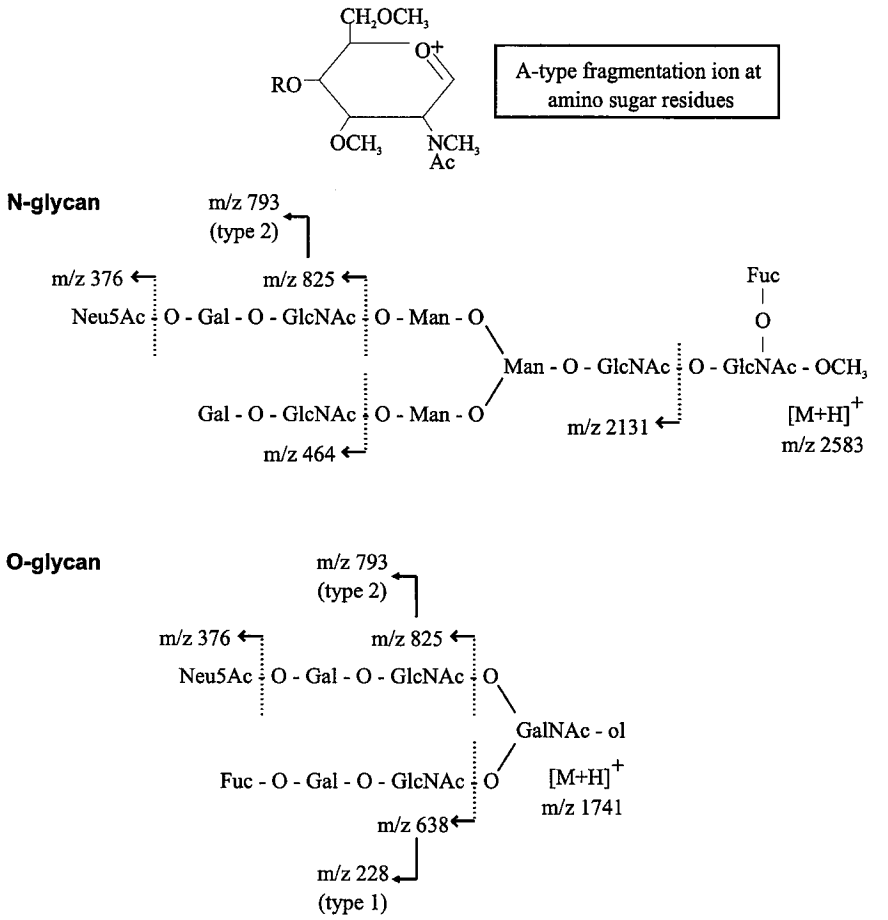

Figure 4. FAB-mass ion fragmentation of permethylated glycans derived from glycopeptides. A-type cleavage occurs reliably at each HexNAc residue, providing information about the composition of glycan moieties attached on the nonreducing sides of the cleavage sites.

peroxidase [54]. The introduction of array detectors and sector instruments increased the molecular ion detectability, thereby increasing the sensitivity to the picomole (femtomole) range.

The application of MALDI-TOF-MS (time-of-flight detection) is rapidly increasing in the glycoprotein/peptide field, due to its ease of operation [55] and its better tolerance of impurities (buffers, salts, additives and detergents) in comparison with other MS methods [56]. The recent introduction of delayed extraction technology for MALDI-MS has significantly improved mass accuracies [57, 58]. Producing only the molecular mass ion for each glycan, MALDI-MS is very useful for profiling glycan mixtures without derivatization. Furthermore, the new technique PSD (post-source decay)-MALDI-TOF-MS has been demonstrated to be a fast,

highly sensitive and reproducible method for the localization of O-glycosylation sites of polymorphic epithelial mucin using glycopeptides [59] and the isomeric differentiation of N-glycans [60].

A development of particular significance in MS technology is the construction of a novel quadrupole orthogonal acceleration time-of-flight tandem mass spectrometer (Q-TOF) for ultra high sensitivity low femtomole/attomole-range glycopeptide sequencing [61].

4.3 NMR spectroscopy

High-resolution NMR spectroscopy is the most powerful method for the unambiguous identification of N- as well as O-type carbohydrate chains. For the elucidation of the primary structure, it is the only (nondestructive) method that provides all details, comprising type of constituent monosaccharides, including ring size and anomeric configuration, position of glycosidic linkages as well as position of noncarbohydrate substituents. Studying glycopeptides containing small peptide parts, usually the signals from the peptide portion do not interfere with those from the oligosaccharide. As an illustrative example, a typical ^{1}H-NMR spectrum of a glycopeptide is shown in Figure 5.

Already at first sight, the one-dimensional (1D) ^{1}H-NMR spectrum can be used as a "fingerprint" containing many characteristic details to conclude whether or not compounds are pure and/or identical. The definitive interpretation of the proton spectrum in terms of primary structure is based on the structural-report-group concept [62]. Most of the skeleton protons of the monosaccharide residues resonate in a narrow region between (δ) 3.5 and 3.9 ppm. Proton signals at clearly distinguishable positions outside this bulk, denoted as structural-reporter-group signals (Tab. 7), reveal

Table 7. ^{1}H-NMR structural-reporter-group signals in glycoprotein glycan analysis

- Anomeric protons (H-1)
- Amide protons (in H_2O)
- Protons attached to carbon atoms in the direct vicinity of a substitution position
  - Man H-2 and H-3
  - Gal H-3 and H-4
- Protons at deoxy carbon atoms
  - sialic acid H-3 (equatorial and axial)
  - Fuc H-5 and H-6 (CH_3)
- Protons shifted (out of the bulk region) due to glycosylation shifts
- Alditol protons
  - GalNAc-ol H-2, H-3, H-4 and H-5
- Protons shifted (out of the bulk region) due to noncarbohydrate substituents like alkyl, acyl, phosphate or sulfate groups
- Protons belonging to substituents on carbohydrate residues like *O*- and/or *N*-alkyl or acyl groups

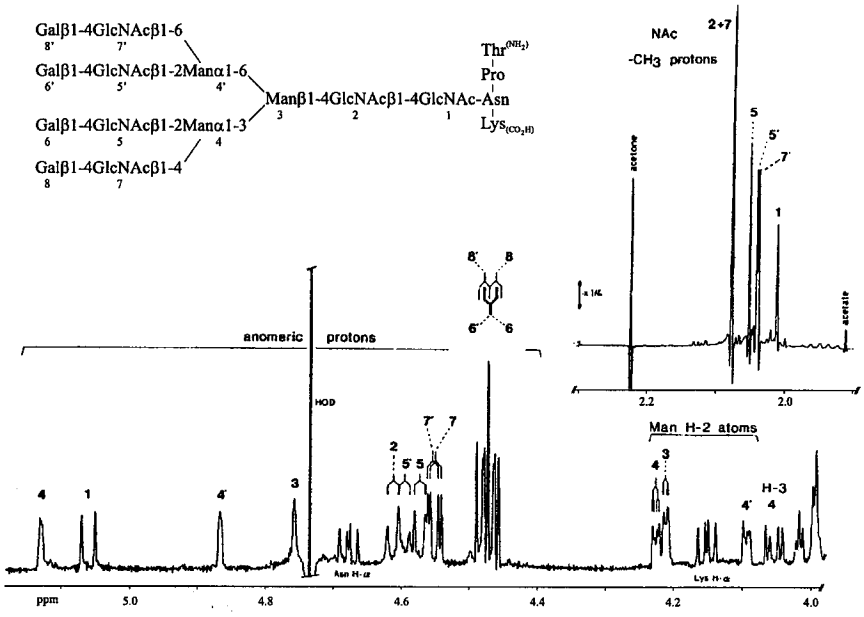

Figure 5. 500-MHz ^{1}H-NMR spectrum showing the structural-reporter-group regions of a desialylated complex-type glycan-peptide. The bulk region has been omitted. The boldface numbers in the spectrum refer to the corresponding residues in the structure. The relative-intensity scale of the NAc-proton region differs from that of the other part as indicated. The spectrum was recorded in D_2O at 300 K.

information on the primary structure depending on their chemical shifts (δ), intensities (area), line widths and coupling constants (J).

The application of various types of two-dimensional (2D)-NMR methods significantly improves the interpretation and assignments for the assessment of complex carbohydrate structures [63]. 2D ^{1}H-total correlation spectroscopy (TOCSY) is used to identify the spin systems in the constituent monosaccharides by means of the scalar coupled network starting from the anomeric protons. 2D ^{1}H-nuclear Overhauser effect spectroscopy (NOESY) is employed to obtain information on protons that are close in space, in particular the protons connected to the glycosidic linkage C atoms. It has to be noted that the strongest NOE is not always between these protons. Proton-carbon heteronuclear experiments correlate ^{1}H and ^{13}C chemical shifts in a 2D spectrum. Heteronuclear multiple quantum-coherence spectroscopy (^{1}H-^{13}C-HMQC) makes it possible to determine the correlation between carbon atoms and the directly attached protons. For determination of the position of the glycosidic linkage, heteronuclear multiple-bond correlation spectroscopy (HMBC) can be used. This method yields cross-peaks between ^{1}H and ^{13}C atoms over several chemical bonds. Additionally, three-dimensional (3D)-NMR spectroscopy can be used for

the investigation of the structure of complex carbohydrates. A nonselective 3D ^{1}H-^{13}C-NOE-TOCSY experiment and 3D HMQC-NOE spectroscopy provide detailed information on protons resonating in the bulk signal [64, 65].

The conversion of the NMR data into carbohydrate structures makes use of extensive libraries of reference compounds [66, 67] and computerized databases [68–70], now accessible through the internet [71].

Typical examples of glycan structure analyses as recently carried out in our research group include: the N-glycans of human urokinase [72], monoclonal immunoglobulin (Ig) G1 antibodies [73] and αD-hemocyanin [74], the N- and O-glycans of recombinant human erythropoietin [75], the O-glycans of equine chorionic gonadotropin [76], porcine zona pellucida glycoproteins [77], a bovine seminal plasma protein [78] and jacalin-bound rabbit IgG [79]. In all these cases the analyses, using ^{1}H-NMR spectroscopy as the major technique, were performed on the level of sialylated carbohydrate chains, providing also information about the sialylation patterns. The type of information obtainable from different NMR techniques is shown in Table 8.

Not only for the primary structural analysis but also for the study of the conformation and intramolecular interactions of glycans and glycoproteins/peptides, ^{1}H-NMR spectroscopy has proven to be an invaluable technique, providing data regarding the spatial structure and segmental mobility of carbohydrate chains in solution [80]. The solution conformation of oligosaccharides can be determined through the combined use of internuclear distances obtained from quantitative proton NOE measurements and tortional angles derived from coupling constants, in conjunction with molecular dynamics calculations. These data, together with those obtained from circular dichroism and X-ray diffraction, are crucial for computer-graphics molecular modelling of oligosaccharides and glycopeptides. Recently, some intact glycoproteins have been investigated by

Table 8. NMR parameters and their application for oligosaccharide structure determination

Relevant NMR parameter	Information
Chemical shift, spin couplings ($^{1}H-^{1}H$), spectral integration	monomer composition
Spin couplings ($^{1}H-^{1}H$, $^{13}C-^{1}H$, $^{13}C-^{13}C$)	monomer conformation
Chemical shift, spin couplings ($^{1}H-^{1}H$, $^{13}C-^{1}H$)	anomeric configuration
Chemical shift, NOE, ROE, spin coupling ($^{13}C-^{1}H$)	linkage sites
Nuclear spin relaxation time (T_1, T_2), NOE, spin coupling ($^{13}C-^{1}H$, $^{13}C-^{13}C$)	O-glycoside conformation
Spin couplings ($^{1}H-^{1}H$, $^{13}C-^{1}H$)	hydroxymethyl conformation
Nuclear spin relaxation time (T_1, T_2), NOE	motional/dynamical properties

NMR [81, 82]. For instance, differences in conformational behavior of the unique N-glycan of pineapple stem bromelain glycoprotein with that of the bromelain-derived glycopeptide were demonstrated. For this purpose, models obtained through molecular dynamics simulations of the glycopeptide in water were also applied [83, 84]. The two N-glycans of the free α-subunit of human chorionic gonadotropin (hCG) behave differently with reference to flexibility and interaction with the protein backbone as studied by ^{1}H-NMR spectroscopy [85].

5. Identification and analysis of glycosylation sites

Detailed knowledge of the carbohydrate structures at particular glycosylation sites is not only a prerequisite for understanding their involvement in the biological action of certain glycoproteins, but it also provides information on the extent to which oligosaccharide biosynthesis may be affected by the surrounding protein structure within a given cellular glycosylation system [86]. Furthermore, new types of carbohydrate-protein linkages are still being discovered [87], for instance, the C-linkage of Man-Trp in human RNase2 [88, 89] and the occurrence of phosphoglycosylation [90].

In order to study site-specific glycosylation, a glycoprotein is either chemically and/or enzymatically cleaved into its peptides and glycopeptides, which are then separated by a number of chromatographic steps. Recent examples are the separation of glycosylated peptides derived from recombinant tissue plasminogen activator [91] and recombinant coagulation factor VIIa [92] by use of sequential HPLC and capillary zone electrophoresis.

A commonly applied strategy is the digestion of the glycoprotein with trypsin, chymotrypsin and/or V-8 protease followed by purification of glycopeptides representing individual N-glycosylation sites by RP-HPLC. After identification by amino acid sequence analysis, the glycopeptides are analyzed by ^{1}H-NMR spectroscopy and/or mass spectrometry. This procedure has successfully been applied to the determination of site-specific N-glycosylation of human chorionic gonadotropin [93], recombinant human immunodeficiency virus (HIV)-1 gp120 [32], soybean seed coat peroxidase [94], recombinant bovine lactoferrin [95] and recombinant human interleukin-6 [96].

The analysis of the ratio of N-linked oligosaccharides attached to different glycosylation sites of a glycoprotein requires the purification of site-specific glycopeptides. The proteolytic digestion of reduced and alkylated recombinant erythropoietin (r-EPO) to obtain site-specific glycopeptides was successfully performed with Lys-C and trypsin [97]. The determination and characterization of the N-glycosylation sites of Tamm-Horsfall glycoprotein was performed on Glu-C(V-8 protease)-prepared glycopeptides fractionated by RP-HPLC and Concanavalin A (Con A)-affinity

chromatography. It was shown for the glycoprotein from one male donor that from the eight potential glycosylation sites seven sites were occupied and, remarkably, one of them (Asn251) contains predominantly high-mannose-type structures, $Man_{5-8}GlcNAc_2$ [98].

Recently developed mass spectrometric techniques have enabled the direct analysis of oligosaccharide chains while still attached to peptides, together with the peptide sequence. For instance, the reduced and alkylated glycoprotein is digested with a protease such as trypsin or endoproteinase Glu-C, and an aliquot of the digest is analyzed by HPLC connected directly to an ES mass spectrometer to generate a "peptide mass map". Thus, molecular masses which do not match predicted proteolytically generated peptides are possibly glycopeptides. Tandem ES-MS/CID (collision-induced dissociation)-MS can be used for selected ion monitoring (SIM) of carbohydrate-specific ions [99]. In order to help identify N-glycopeptides, the N-glycans are cleaved from the peptide by PNGase F or Endo H in an aliquot of the digest and rechromatographed to detect changes in the masses or retention times of the components in the digest. By combining the change in mass after different glycosidase treatments, the types of oligosaccharides on the glycopeptide can be determined [100]. Tandem mass spectrometry has also been proven to be useful for the analysis of O-glycopeptides, yielding the peptide sequence and the exact site of glycosylation on the peptide [101].

Recently, a new technique has been reported to gain site-specific compositional data on the oligosaccharides attached to a single amino acid. Sequential solid-phase Edman degradation was used to recover a single glycosylated phenylthiohydantoin (PTH) amino acid from a (preferably desialylated) glycopeptide coupled to an arylamine membrane. The oligosaccharide of the glyco-amino acid is characterized by HPAEC monosaccharide analysis and ES-MS [102, 103]. Although the yield is limited and the oligosaccharide is degraded slowly during repeated cycles of Edman degradation, the approach looks promising for the characterization of heavily O-glycosylated proteins.

6. Concluding remarks

The persistence throughout evolution of glycosylation of proteins, together with the fact that glycosylation of proteins is an expensive operation requiring genetic information for the production of many enzymes, substrates and cofactors, strongly indicates an important function for the variety of glycans. To understand the biological role of glycoprotein glycans, their detailed structure, conformation, and interactions with complementary molecules must be known. Furthermore, the complete characterization of glycoproteins will contribute to elucidating the biosynthetic controls that determine site-specific glycosylation patterns. In recent years the academic

and industrial interests in the carbohydrate part of glycoproteins has grown dramatically, leading to new developments in separation techniques and in methodology to unravel the structure of complex carbohydrates. The progress in MS and NMR spectroscopy for the latter purpose is spectacular.

Acknowledgments
The studies in the authors' laboratory were supported by the Netherlands Foundation for Chemical Research (SON/NWO).

References

1 Montreuil J, Vliegenthart JFG, Schachter H (eds) (1995) *Glycoproteins: new comprehensive biochemistry*, vol. 29a. Elsevier Science, Amsterdam
2 Montreuil J, Vliegenthart JFG, Schachter H (eds) (1996) *Glycoproteins II: new comprehensive biochemistry*, vol. 29b. Elsevier Science, Amsterdam
3 Montreuil J, Vliegenthart JFG, Schachter H (eds) (1996) *Glycoproteins and disease: new comprehensive biochemistry*, vol. 30. Elsevier Science, Amsterdam
4 Allen HJ, Kisailus EC (eds) (1992) *Glycoconjugates: composition, structure and function*. Marcel Dekker, New York
5 Natsuka S, Lowe JB (1994) Enzymes involved in mammalian oligosaccharide biosynthesis. *Curr Opin Struct Biol* 4: 683–691
6 Vance BA, Wu W, Ribaudo RK, Segal DM, Kearse KP (1997) Multiple dimeric forms of human CD69 result from differential addition of N-glycans to typical (Asn-X-Ser/Thr) and atypical (Asn-X-Cys) glycosylation motifs. *J Biol Chem* 272: 23117–23122
7 Gavel Y, Von Heijne G (1990) Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Engineering* 3: 433–442
8 Gooley AA, Classon BJ, Marschalek R, Williams KL (1991) Glycosylation sites identified by detection of glycosylated amino acids released from Edman degradation: the identification of Xaa-Pro-Xaa-Xaa as a motif for Thr-O-glycosylation. *Biochem Biophys Res Commun* 178: 1194–1201
9 Nehrke K, Ten Hagen KG, Hagen FK, Tabak LA (1997) Charge distribution of flanking amino acids inhibits O-glycosylation of several single-site acceptors *in vivo*. *Glycobiology* 7: 1053–1060
10 Lis H, Sharon N (1993) Protein glycosylation: structural and functional aspects. *Eur J Biochem* 218: 1–27
11 Bulet P, Hegy G, Lambert J, Van Dorsselaer A, Hoffmann JA, Hetru C (1995) Insect Immunity. The inducible antibacterial peptide Diptericin carries two O-glycans necessary for biological activity. *Biochemistry* 34: 7394–7400
12 Williams DH (1996) The glycopeptide story – how to kill the deadly "superbugs". *Nat Prod Rep* 13: 469–477
13 Vliegenthart JFG (1994) Studies on glycoprotein-derived carbohydrates. *Biochem Soc Trans* 22: 370–373
14 Vliegenthart JFG (1994) Studies on the carbohydrate chains of glycoproteins. In: Bock K, Clausen H (eds): *Complex carbohydrates in drug research*. Alfred Benzon Symposium 36: 30–41
15 Kamerling JP (1994) Structural studies on glycoprotein glycans. *Pure Appl Chem* 66: 2235–2238
16 Jenkins N, Parekh RB, James DC (1996) Getting the glycosylation right: implications for biotechnology industry. *Nature Biotechnol* 14: 975–981
17 Hounsel EF (ed) (1993) Glycoprotein analysis in biomedicine. *Methods in molecular biology*, vol. 14. Humana Press, Totowa, NJ
18 Sojar HT, Bahl OP (1987) Chemical deglycosylation of glycoproteins. *Methods Enzymol* 138: 341–350

19 Gerken TA, Gupta R, Jentoft N (1992) A novel approach for chemically deglycosylating O-linked glycoproteins. The deglycosylation of submaxillary and respiratory mucins. *Biochemistry* 31: 639–648
20 Kamerling JP, Vliegenthart JFG (1989) Mass Spectrometry. In: AM Lawson (ed): *Clinical biochemistry: principles, methods, applications*, vol. 1. Walter de Gruyter, Berlin, 176–263
21 Townsend RR, Hardy MR (1991) Analysis of glycoprotein oligosaccharides using high-pH anion exchange chromatography. *Glycobiology* 1: 139–147
22 Dwek RA, Edge CJ, Harvey DJ, Dormald MR, Parekh RB (1993) Analysis of glycoprotein-associated oligosaccharides. *Annu Rev Biochem* 62: 65–100
23 Takasaki S, Mizuochi T, Kobata A (1982) Hydrazinolysis of asparagine-linked sugar chains to produce free oligosaccharides. *Methods Enzymol* 83: 263–268
24 Merry AH, Bruce J, Bigge C, Ioannides A (1992) Automated simultaneous release of intact and unreduced N- and O-linked glycans from glycoproteins. *Biochem Soc Trans* 20: 91s
25 Schneider P, Ferguson MAJ (1995) Microscale analysis of glycosylphosphatidylinositol structures. *Methods Enyzmol* 250: 614–630
26 Nilsson B, Zopf D (1982) Gas chromatography and mass spectrometry of hexosamine-containing oligosaccharide alditols as their permethylated *N*-trifluoroacetyl derivatives. *Methods Enzymol* 83: 46–58
27 Zinn AB, Plantner JJ, Carlson DM (1977) Nature of linkages between protein and oligosaccharides. In: MI Horowitz, W Pigman (eds): *The glycoconjugates*. Academic Press, New York, 69–85
28 Tarentino AL, Plummer TH (1994) Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from *Flavobacterium meningosepticum*. *Methods Enzymol* 230: 44–57
29 Ishii-Karakasa I, Iwase H, Hotta K, Tanaka Y, Omura S (1992) Partial purification and characterization of an endo-α-*N*-acetylgalactosaminidase from the culture medium of *Streptomyces* sp. OH-11242. *Biochem J* 288: 475–482
30 Damm JBL, Kamerling JP, Van Dedem GWK, Vliegenthart JFG (1987) A general strategy for the isolation of carbohydrate chains from N,O-glycoproteins and its application to human chorionic gonadotropin. *Glycoconjugate J* 4: 129–144
31 Finne J, Krusius T (1982) Preparation and fractionation of glycopeptides. *Methods Enzymol* 83: 269–277
32 Yeh J, Seals JR, Murphy CI, Van Halbeek H, Cummings RD (1993) Site-specific N-glycosylation and oligosaccharide structures of recombinant HIV-1 gp120 derived from a *Baculovirus* expression system. *Biochemistry* 32: 11087–11099
33 Hunter AP, Games DE (1995) Evaluation of glycosylation site heterogeneity and selective identification of glycopeptides in proteolytic digests of bovine α_1-acid glycoprotein by mass spectrometry. *Rapid Commun Mass Spectrom* 9: 42–56
34 Rohrer JS, Cooper GA, Townsend RR (1993) Identification, quantification and characterization of glycopeptides in reversed-phase HPLC separations of glycoprotein proteolytic digests. *Anal Biochem* 212: 7–16
35 Davies MJ, Smith KD, Carruthers RA, Chai W, Lawson AM, Hounsell EF (1993) Use of a porous graphitised carbon column for the high-performance liquid chromatography of oligosaccharides, alditols and glycopeptides with subsequent mass spectrometry analysis. *J Chromatog* 646: 317–326
36 Hsi KL, Chen L, Hawke DH, Zieske LR, Yuan PM (1991) A general approach for characterizing glycosylation sites of glycoproteins. *Anal Biochem* 198: 238–245
37 Osawa T, Tsuji T (1987) Fractionation and structural assessment of oligosaccharides and glycopeptides by use of immobilized lectins. *Annu Rev Biochem* 56: 21–42
38 Merkle RK, Cummings RD (1987) Lectin affinity chromatography of glycopeptides. *Methods Enzymol* 138: 232–259
39 Shimura K, Kasai K (1997) Affinity capillary electrophoresis: A sensitive tool for the study of molecular interactions and its use in microscale analysis. *Anal Biochem* 251: 1–16
40 Gerwig GJ, Kamerling JP, Vliegenthart JFG (1979) Determination of the absolute configuration of monosaccharides in complex carbohydrates by capillary GLC. *Carbohydr Res* 77: 1–7

41 Jay A (1996) The methylation reaction in carbohydrate analysis. *J Carbohydr Chem* 15: 897–923
42 Hellerqvist CG, Sweetman BJ (1990) Mass spectrometry of carbohydrates. Methods Biochem Anal 34: 91–143
43 Yamashita K, Mizuochi T, Kobata A (1982) Analysis of oligosaccharides by gel filtration. *Methods Enzymol* 83: 105–126
44 Edge CJ, Rademacher TW, Wormald MR, Parekh RB, Butters TD, Wing DR, Dwek RA (1992) Fast sequencing of oligosaccharides: the Reagent Array Analysis Method. *Proc Natl Acad Sci USA* 89: 6338–6342
45 Hermentin P, Doenges R, Witzel R, Hokke CH, Vliegenthart JFG, Kamerling JP, Conradt HS, Mimtz M, Brazel D (1994) A strategy for the mapping of N-glycans by high-performance capillary electrophoresis. *Anal Biochem* 221: 29–41
46 Guttmann A, Starr CM (1995) Capillary and slab gel electrophoresis of oligosaccharides. *Electrophoresis* 16: 993–997
47 Takahashi N, Nakagawa H, Fujikawa K, Kawamura Y, Tomiya N (1995) Three-dimensional elution mapping of pyridylaminated N-linked neutral and sialyl oligosaccharides. *Anal Biochem* 226: 139–146
48 Starr C, Masada RI, Hague C, Skop E, Klock J (1996) Fluorophore-Assisted-Carbohydrate-Electrophoresis, FACE, in the separation, analysis, and sequencing of carbohydrates. *J Chromatography* 720: 295–321
49 Dell A (1989) FAB mass spectrometry of carbohydrates. *Adv Carbohydr Chem Biochem* 45: 19–72
50 Khoo KH, Chatterjee D, Caulfield JP, Morris HR, Dell A (1997) Structural mapping of the glycans from egg glycoproteins of *Schistosoma mansoni* and *Schistosoma japonicum*. identification of novel core structures and terminal sequences. *Glycobiology* 7: 663–677
51 Medzihradszky KF, Maltby DA, Hall SC, Settineri CA, Burlingame AL (1994) Characterization of protein N-glycosylation by reversed-phase microbore liquid chromatography/electrospray mass spectrometry, complementary mobile phases, and sequential exoglycosidase digestion. *J Am Soc Mass Spectrom* 5: 350–358
52 Settineri CA, Burlingame AL (1995) Mass spectrometry of carbohydrates and glycoconjugates. In: EI Rassi Z (ed): *Carbohydrate analysis: high performance liquid chromatography and capillary electrophoresis*. Elsevier Science, Amsterdam, 447–514
53 Morris HR, Dell A, Easton RL, Panico M, Koiestinen H, Koistinen R, Oehninger S, Patankar MS, Seppala M, Clark GF (1996) Gender-specific glycosylation of human glycodelin affects its contraceptive activity. *J Biol Chem* 271: 32159–32170
54 Takahashi N, Lee KB, Nakagawa H, Tsukamoto Y, Masuda K, Lee YC (1998) New N-glycans in horseradish peroxidase. *Anal Biochem* 255: 183–187
55 Treuheit MJ, Costello CE, Halsall HB (1992) Analysis of the five glycosylation sites of human α_1-acid glycoprotein. *Biochem J* 283: 105–112
56 Vorm O, Roepstorff P, Mann M (1994) Improved resolution and very high sensitivity in MALDI-TOF of matrix surfaces made by fast evaporation. *Anal Chem* 66: 3281–3287
57 Vestal ML, Juhasz P, Martin SA (1995) Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 9: 1044–1050
58 Sugiyama E, Hara A, Uemura K, Taketomi T (1997) Application of matrix-assisted laser desorption ionization time-of-flight mass spectrometry with delayed ion extraction to ganglioside analysis. *Glycobiology* 7: 719–724
59 Goletz S, Thiede B, Hanisch F-G, Schultz M, Peter-Katalinic J, Muller S, Seitz O, Karsten U (1997) A sequencing strategy for the localization of O-glycosylation sites of MUC1 tandem repeats by PSD-MALDI mass spectrometry. *Glycobiology* 7: 881–896
60 Rouse JC, Strang A-M, Yu W, Vath JE (1998) Isomeric differentiation of asparagine-linked oligosaccharides by matrix-assisted laser desorption-ionization postsource decay time-of-flight mass spectrometry. *Anal Biochem* 256: 33–46
61 Morris HR, Paxton T, Panico M, McDowell R, Dell A (1997) A novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing. *J Protein Chem* 16: 469–479
62 Vliegenthart JFG, Van Halbeek H, Dorland L (1981) The application of 500-MHz high-resolution ^{1}H-NMR spectroscopy for the structural determination of carbohydrates derived from glycoproteins. *Pure Appl Chem* 53: 45–77

63 Hård K, Vliegenthart JFG (1993) Nuclear magnetic resonance spectroscopy of glycoprotein-derived carbohydrate chains. In: M Fukuda, A Kobata (eds): *Glycobiology, a practical approach*. Oxford University Press, Oxford, 223–242
64 Vuister GW, De Waard P, Boelens R, Vliegenthart JFG, Kaptein R (1989) The use of 3D NMR in structural studies of oligosaccharides. *J Am Chem Soc* 111: 772–774
65 De Waard P, Vuister R, Boelens R, Vliegenthart JFG (1990) Structural studies by $^{1}H/^{13}C$ two-dimensional and three-dimensional HMQC-NOE of natural abundance in complex carbohydrates. *J Am Chem Soc* 112: 3232–3234
66 Vliegenthart JFG, Dorland L, Van Halbeek H (1983) High-resolution, ^{1}H-nuclear magnetic resonance spectroscopy as a tool in the structural analysis of carbohydrates related to glycoproteins. *Adv Carbohydr Chem Biochem* 41: 209–374
67 Kamerling JP, Vliegenthart JFG (1992) High-resolution ^{1}H-nuclear magnetic resonance spectroscopy of oligosaccharide-alditols released from Mucin-type O-glycoproteins. *Biol Magn Reson* 10: 1–194
68 Van Kuik JA, Hård K, Vliegenthart JFG (1992) A ^{1}H-NMR database computer program for the analysis of the primary structure of complex carbohydrates. *Carbohydr Res* 235: 53–68
69 Van Kuik JA, Vliegenthart JFG (1992) Databases of complex carbohydraes. *Trends Biotechnol* 10: 182–185
70 Van Kuik JA, Vliegenthart JFG (1994) A NMR spectroscopic database of complex carbohydrate structures. *Carbohydr Europe* 10: 31–32
71 URL: http://www.boc.chem.uu.nl/sugabase/databases.html
72 Bergwerff AA, Van Oostrum J, Kamerling JP, Vliegenthart JFG (1995) The major N-linked carbohydrate chains from human urokinase; the occurrence of 4-O-sulfate, α(2-6)-sialylated or α(1-3)-fucosylated *N*-acetylgalactosamine-β(1-4)-*N*-acetylglucosamine elements. *Eur J Biochem* 228: 1009–1019
73 Bergwerff AA, Stroop CJM, Murray B, Holtorf AP, Pluschke G, Van Oostrum J, Kamerling JP, Vliegenthart JFG (1995) Variation in N-linked carbohydrate chains in different batches of two chimeric monoclonal IgG1 antibodies produced by different murine SP2/0 transfectoma cell subclones. *Glycoconjugate J* 12: 318–330
74 Lommerse JPM, Thomas-Oates JE, Gielens C, Preaux G, Kamerling JP, Vliegenthart JFG (1997) Primary structure of 21 novel monoantennary and diantennary N-linked carbohydrate chains from α D-hemocyanin of *Helix pomatia*. *Eur J Biochem* 249: 195–222
75 Hokke CH, Bergwerff AA, Van Dedem GWK, Kamerling JP, Vliegenthart JFG (1995) Structural analysis of sialylated N- and O-linked carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells; sialylation patterns and branch location of dimeric *N*-acetyllactosamine units. *Eur J Biochem* 228: 981–1008
76 Hokke CH, Roosenboom MJH, Thomas-Oates JE, Kamerling JP, Vliegenthart JFG (1994) Structure determination of the disialylated poly-(*N*-acetyllactosamine)-containing O-linked carbohydrate chains of equine chorionic gonadotropin. *Glycoconjugate J* 11: 35–41
77 Hokke CH, Damm JBL, Penninkhof B, Aitken RJ, Kamerling JP, Vliegenthart JFG (1994) Structure of the O-linked carbohydrate chains of procine zona pellucida glycoproteins. *Eur J Biochem* 221: 491–512
78 Gerwig GJ, Calvete JJ, Töpfer-Petersen E, Vliegenthart JFG (1996) The structure of O-linked carbohydrate chains of bovine seminal plasma protein PDC-109 revised by ^{1}H-NMR spectroscopy. A correction. *FEBS Lett* 387: 99–100
79 Kabir S, Gerwig GJ (1997) The structural analysis of the O-glycans of the jacalin-bound rabbit immunoglobulin G. *Biochem Molec Biol Int* 42: 769–778
80 Van Halbeek H (1994) NMR developments in structural studies of carbohydrates and their complexes. *Curr Opin Struct Biol* 4: 697–709
81 Fletcher MC, Harrison RA, Lachmann PJ, Neuhaus D (1994) Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. *Structure* 2: 185–199
82 De Beer T, Van Zuylen CWEM, Leeflang BR, Hård K, Boelens R, Kaptein R, Kamerling JP, Vliegenthart JFG (1996) NMR studies of the free α subunit of human chorionic gonadotropin. Structural influences of N-glycosylation and the β subunit on the conformation of the α subunit. *Eur J Biochem* 241: 229–242
83 Lommerse JPM, Kroon-Batenburg LMJ, Kamerling JP, Vliegenthart JFG (1995) Conformational analysis of the xylose-containing N-glycan of pineapple stem bromelain as part of the intact glycoprotein. *Biochemistry* 34: 8196–8206

84 Lommerse JPM, Kroon-Batenburg LMJ, Kroon J, Kamerling JP, Vliegenthart JFG (1995) Conformations and internal mobility of a glycopeptide derived from bromelain using molecular dynamics simulations and NOESY analysis. *J Biomol NMR* 5:79–94
85 Van Zuylen CWEM, Kamerling JP, Vliegenthart JFG (1997) Glycosylation beyond the Asn78-linked GlcNAc residues has a significant enhancing effect on the stability of the α subunit of human chorionic gonadotropin. *Biochem Biophys Res Commun* 232: 117–120
86 Yet MG, Wold F (1990) The distribution of glycan structures in individual N-glycosylation sites in animal and plants glycoproteins. *Arch Biochem Biophys* 278: 356–364
87 Hayes BK, Hart GW (1994) Novel forms of protein glycosylation. *Curr Opin Struct Biol* 4: 692–696
88 Hofsteenge J, Müller DR, De Beer T, Löffler A, Richter WJ, Vliegenthart JFG (1994) New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry* 33: 13524–13530
89 Hofsteenge J, Löffler A, Müller DR, Richter WJ, De Beer T, Vliegenthart JFG (1996) Protein C-glycosylation. *Techniques Protein Chem* VII: 163–171
90 Haynes PA (1998) Phosphoglycosylation: a new structural class of glycosylation? *Glycobiology* 8: 1–5
91 Wu SL (1997) The use of sequential high-performance liquid chromatography and capillary zone electrophoresis to separate the glycosylated peptides from recombinant tissue plasminogen activator to a detailed level of microheterogeneity. *Anal Biochem* 253: 85–97
92 Weber PL, Kornfelt T, Klausen NK, Lunte SM (1995) Characterization of glycopeptides from recombinant coagulation factor VIIa by high-performance liquid chromatography and capillary zone electrophoresis using ultraviolet and pulsed electrochemical detection. *Anal Biochem* 225: 135–142
93 Weisshaar G, Hiyama J, Renwick AGC (1991) Site-specific N-glycosylation of human chorionic gonadotropin-structural analysis of glycopeptides by one- and two-dimensional ^{1}H-NMR spectroscopy. *Glycobiology* 1: 393–404
94 Gray JSS, Montgomery R (1997) The N-glycosylation sites of soybean seed coat peroxidase. *Glycobiology* 7: 679–685
95 Lopez M, Coddeville B, Langridge J, Plancke Y, Sautiere P, Chaabihi H, Chirat F, Harduin-Lepers A, Cerutti M, Verbert A, Delannoy P (1997) Microheterogeneity of the oligosaccharides carried by recombinant bovine lactoferrin expressed in *Mamestra brassicae* cells. *Glycobiology* 7: 635–651
96 Orita T, Oh-eda M, Hasegawa M, Kuboniwa H, Esaki K, Ochi N (1994) Polypeptide and carbohydrate structure of recombinant human interleukin-6 produced in Chinese hamster ovary cells. *J Biochem* 115: 345–350
97 Sasaki H, Ochi N, Dell A, Fukuda M (1988) Site-specific glycosylation of human recombinant erythropoietin: analysis of glycopeptides or peptides at each glycosylation site by fast atom bombardment mass spectrometry. *Biochemistry* 27: 8618
98 Van Rooijen JJM, Voskamp AF, Kamerling JP, Vliegenthart JFG (1999) Glycosylation sites and site-specific glycosylation in human Tamm-Horsfall glycoprotein. *Glycobiology* 9: 21–30
99 Reinhold VN, Reinhold BB, Costello CE (1995) Carbohydrate molecular weight profiling, sequence, linkage, and branching data: ES-MS and CID. *Anal Chem* 67: 1772–1784
100 Carr SA, Hemling ME, Bean MF, Roberts GD (1991) Integration of mass spectrometry in analytical biotechnology. *Anal Chem* 63: 2802–2824
101 Medzihradszky KF, Gillece-Castro BL, Settineri CA, Townsend RR, Masiarz FR, Burlingame AL (1990) Structure determination of O-linked glycopeptides by tandem mass spectrometry. *Biomed Environ Mass Spectrom* 19: 777–781
102 Gooley AA, Pisano A, Packer NH, Ball M, Jones A, Alewood PF, Redmond JW, Williams KL (1994) Characterization of a single glycosylated asparagine site on a glycopeptide using solid-phase Edman degradation. *Glycoconjugate J* 11: 180–186
103 Gooley AA, Williams KL (1997) How to find, identify and quantitate the sugars on proteins. *Nature* 385: 557–559