8.1 Glycoproteins: Occurrence and Significance

Valentin Wittmann
Fachbereich Chemie, Universität Konstanz, 78457 Konstanz, Germany
mail@valentin-wittmann.de

1	**Introduction**	1736
2	**Glycoproteins**	1737
2.1	Monosaccharide Constituents	1738
2.2	Carbohydrate–Peptide Linkages	1738
2.2.1	*N*-Glycosides	1740
2.2.2	*O*-Glycosides	1742
2.2.3	GPI Anchors	1744
2.2.4	Uncommon Linkages	1746
2.3	Methods of Glycoprotein-Glycans Analysis	1746
2.3.1	Release of Glycans from Glycoprotein	1747
2.3.2	Labeling and Fractionation of Glycans	1747
2.3.3	Sequencing of Glycans	1748
2.4	Structure of *N*-Glycans	1749
2.4.1	Diversity of *N*-Glycans	1750
2.4.2	Cell-Type Specific Expression of *N*-Glycans	1752
2.5	Structure of *O*-Glycans	1753
2.5.1	Mucin-Type *O*-Glycans	1754
2.5.2	Non-Mucin *O*-Glycans	1755
2.6	Proteoglycans	1758
2.6.1	Carbohydrate–Peptide Linkage Region	1759
2.6.2	Structure of Glycosaminoglycans	1759
3	**Peptidoglycan**	1760
4	**Glycopeptide Antibiotics**	1762
4.1	Discovery and Medical Use	1763
4.2	Structure and Classification	1763

Abstract

Protein glycosylation is regarded as the most complex form of post-translational modification leading to a heterogeneous expression of glycoproteins as mixtures of glycoforms. This chapter describes the structure and occurrence of glycoproteins with respect to their glycan chains. Discussed are different carbohydrate–peptide linkages including GPI anchors, common structures of *N*- and *O*-glycans, and the structure of glycosaminoglycans contained in proteoglycans. Also covered are the bacterial cell wall polymer peptidoglycan and the glycopeptide

antibiotics of the vancomycin group. Properties and functions of the glycans contained in glycoproteins are dealt with in the next chapter of this book.

Keywords

Glycoproteins; Glycopeptides; Glycoforms; Microheterogeneity; *N*-Glycans; *O*-Glycans; Antigenic determinants; GPI anchors; Mucins; Proteoglycans

Abbreviations

AGEs	advanced glycation end products
CE	capillary electrophoresis
DATDH	2,4-diacetamido-2,4,6-trideoxyhexose
ER	endoplasmic reticulum
ERGIC	ER-Golgi intermediate compartment
EGF	epidermal growth factor
ECM	extracellular matrix
GPI	glycosyl-phosphatidylinositol
GPC	gel permeation chromatography
HPAEC	high pH anion exchange chromatography
HPLC	high performance liquid chromatography
IGF-I	insulin-like growth factor-I
MS	mass spectrometry
***m*-DAP**	*meso*-diaminopimelic acid
MRSA	methicillin-resistant *Staphylococcus aureus*
OT	oligosaccharyl transferase
PAGE	polyacrylamide gel electrophoresis
PNGase	peptide *N*-glycanase
RAAM	reagent array analysis method
SBA	soybean agglutinin
tPA	tissue plasminogen activator
VSG	variant surface glycoprotein

1 Introduction

This chapter covers natural glycoconjugates in which carbohydrates are covalently linked to amino acids which themselves are part of a peptide or protein. Three classes of biologically important compounds fulfill this criterion:

- glycoproteins including their substructures, the glycopeptides,
- the bacterial cell wall polymer peptidoglycan, and
- glycopeptide antibiotics.

Glycoproteins constitute the most diverse of these classes of glycoconjugates. Their substructures, the glycopeptides, occur as natural and artificial degradation products of glycoproteins during metabolism and structural analysis, respectively, or they are obtained synthetically in

order to probe their structure and/or function. Peptidoglycan is commonly not regarded as a glycopeptide. It forms a bag-like macromolecule (called the sacculus) wrapping bacteria and thereby contributing essentially to their mechanical stability. However, it is connected with glycopeptides not only by virtue of its composition but also because its biosynthesis is inhibited by (beside β-lactam antibiotics) glycopeptide antibiotics. The latter are secondary metabolites produced in actinomycetes and are dealt with at the end of this chapter.

2 Glycoproteins

Glycoproteins [1,2,3,4,5,6,7,8,9,10,11,12,13] consist of carbohydrate covalently linked with protein and are ubiquitous in all forms of life. Once thought to be restricted to eukaryotes, protein glycosylation is now being increasingly reported in prokaryotes [14,15,16,17]. The carbohydrate content of glycoproteins is variable from less than 1%, as in some collagens, to over 99%, as in glycogen. The carbohydrate may be in the form of a monosaccharide or disaccharide, but more frequently in the form of oligosaccharides and polysaccharides (up to hundreds of monosaccharides in size), or their derivatives (e. g. sulfo- or phospho-substituted), linear or branched, generally referred to as glycans. The number of glycans present in a glycoprotein differs greatly from a single to more than a hundred. The carbohydrate chains are attached to the polypeptide backbone by characteristic carbohydrate–peptide linkages.

Being the most complex co- and post-translational modification, glycosylation is found in most proteins including enzymes, antibodies, receptors, hormones, cytokines, and structural proteins. Glycoproteins occur inside cells, both in the cytoplasm and in subcellular organelles, in extracellular fluids as well as embedded in cell membranes. In the latter case the glycans are located extracellularly. Especially blood serum is a rich source of glycoproteins. Of the almost 100 proteins which have been identified in this fluid, almost all are glycosylated. Serum albumin is one of the rare exceptions in this respect although a genetic variant which is glycosylated has been discovered [18].

A striking feature of almost all glycoproteins is the polymorphism associated with their glycan moieties, a phenomenon known as microheterogeneity. This type of diversity derives from the fact that glycans are secondary gene products and is manifested in that individual molecules of a given glycoprotein carry different oligosaccharides at the same glycosylation site of the protein backbone. The resulting variants, referred to as glycoforms, were first observed in 1962 by Schmid et al. in α_1-acid glycoprotein from human serum by electrophoresis [19]. This glycoprotein contains five glycosylation sites modified with di-, tri-, and tetra-antennary glycans of the *N*-acetyllactosamine type [20]. In addition, the glycans can be fucosylated [21] and sialylated at different levels. Microheterogeneity is also observed in proteins with a single glycosylation site like chicken ovalbumin. Close to 20 different oligosaccharides have been identified at that site (Asn-293) [22]. One of the very rare cases in which microheterogeneity is absent is soybean agglutinin (SBA), a plant glycoprotein with a single uniform oligosaccharide per subunit, namely $Man_9(GlcNAc)_2$ [23]. This is mainly why SBA is the best source for a preparative isolation of this oligosaccharide. Up to now, many thousands of primary glycan structures have been characterized. There are several glycan databases available via the internet, for example, Glycosciences.de provided by the German Cancer Research Center (http://www.glycosciences.de) [24], KEGG Glycan, part of the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/glycan) [25], and the Glycan Database of the Consor-

tium for Functional Glycomics (http://www.functionalglycomics.org) [26]. These resources are increasingly linking to each other, and additional levels of integration are in development, including the EuroCarbDB initiative (http://www.eurocarbdb.org).

The observation of microheterogeneity gives rise to many interesting questions regarding the origin of this phenomenon and its biological relevance. Whereas in the early view the heterogeneity of the glycans was thought to be random and mainly a result of the lack of fidelity in their synthesis, today it appears that under constant physiological conditions the populations in a set of glycoforms are reproducible and highly regulated. Moreover, the populations change under certain conditions such as cell growth, cell differentiation, and disease, including malignant transformation, suggesting that the presence of different but defined glycoforms is indeed required for the normal functioning of an organism.

Today we know there is no single unifying function for the carbohydrates present in glycoproteins [27]. Perhaps their major function is to participate in numerous physiological and pathological molecular recognition events. In addition, they may modify the physical, chemical, and biological properties of the proteins they are attached to. It has been shown that the glycans alter charge and solubility of proteins and influence the conformation and dynamic properties of the polypeptide chain. Because of their large size, the oligosaccharides may cover functionally important areas of the proteins and thus regulate their interactions with other biomolecules or protect the protein from proteolytic degradation. Properties and functions of glycoproteins are covered in ➋ Chap. 8.2 of this book.

2.1 Monosaccharide Constituents

For a long time less than a dozen monosaccharides were considered to be the main constituents of glycoproteins. Beside these "classical" saccharides, however, refinement of the analytical methods brought up many new monosaccharides, originally considered as rare but which now appear to be more common than previously thought. ➋ *Table 1* [6,28,29,30] gives an overview of the classical and examples of rare monosaccharide constituents of glycoproteins.

2.2 Carbohydrate–Peptide Linkages

There are three major types of linkages between carbohydrates and proteins [1,2,3,4,5,6,7,8,9, 30,31]:

- *N*-glycosidic, between the reducing end monosaccharide and the side chain amide group of asparagine (*N*-glycans),
- *O*-glycosidic, between the reducing end monosaccharide and the side chain hydroxy group of most commonly serine and threonine but also hydroxyproline, hydroxylysine, and tyrosine (*O*-glycans), and
- via ethanolamine phosphate, between the *C*-terminal amino acid of the protein and an oligosaccharide attached to phosphatidylinositol, generally known as the glycosyl-phosphatidylinositol (GPI) anchor.

Beside these, a number of uncommon linkages are found in nature including *C*-glycosides and carbohydrates linked via a phosphodiester bridge. The non-enzymatic condensation of

Table 1
Monosaccharides found in glycoproteins [6,28,29,30]

Monosaccharide[a]	Abbreviation	Comment
"Classical"		
Galactose	Gal	
Glucose	Glc	Mainly in collagens
Mannose	Man	
N-Acetylgalactosamine	GalNAc	
N-Acetylglucosamine	GlcNAc	
L-Fucose	Fuc	
L-Arabinofuranose	Ara*f*	In plant glycoproteins
Xylose	Xyl	In proteoglycans and plant glycoproteins
Glucuronic acid	GlcA	In proteoglycans
L-Iduronic acid	IdoA	In proteoglycans
N-Acetylneuraminic acid	Neu5Ac	Mainly in higher vertebrates and invertebrates
Rare		
2-Acetamido-4-amino-2,4,6-trideoxyglucose		*Clostridium symbiosum*
6-Deoxyaltrose		Salmonid fish eggs
3-Deoxy-D-*glycero-galacto*-nonulosonic acid	Kdn	Salmonid fish eggs
2,3-Diacetamido-2,3-dideoxymannuronic acid		*Bacillus stearothermophilus*
2-*O*-Methyl-L-fucose	Fuc2Me	Nematodes
Galactofuranose	Gal*f*	Bacteria, trypanosoma, fungi
3-*O*-Methylgalactose	Gal3Me	Snail
4-*O*-Methylgalactose	Gal4Me	Nematodes
6-*O*-Methylgalactose	Gal6Me	Algae
Galactose-3-sulfate	Gal3*S*	Thyroglobulin, mucins in cystic fibrosis
N-Acetylgalactosamine-4-sulfate	GalNAc4*S*	Pituitary hormones, Tamm–Horsfall glycoprotein, urokinase
3-*O*-Methylglucose	Glc3Me	*Methanothermus fervidus*
3-*O*-Methyl-*N*-acetylglucosamine	GlcNAc3Me	*Clostridium thermocellum*
N-Acetylglucosamine-6-sulfate	GlcNAc6*S*	Thyroglobulin
Gulose	Gul	Algae
3-*O*-Methylmannose	Man3Me	Snail
Mannose-4-sulfate	Man4*S*	Ovalbumin
Mannose-6-sulfate	Man6*S*	Ovalbumin, slime mold
Mannose-6-methylphosphate	Man6*P*Me	Slime mold
N-Acetylmannosamine	ManNAc	*Clostridium symbiosum*

Table 1 (continued)
Monosaccharides found in glycoproteins [6,28,29,30]

Monosaccharide[a]	Abbreviation	Comment
4,8-Anhydro-*N*-acetylneuraminic acid		Edible bird's nest
8-*O*-Methyl-9-*O*-acetyl-*N*-glycolylneuraminic acid	Neu5Gc8Me9Ac	Starfish
8-*O*-Methyl-7,9-di-*O*-acetyl-*N*-glycolylneuraminic acid	Neu5Gc7,9$Ac_2$8Me	Starfish
L-Rhamnose	Rha	Eubacteria
Pseudaminic acid (5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-L-*manno*-nonulosonic acid)	Pse	*Campylobacter jejuni* flagellin, pili of *Pseudomonas aeruginosa*
2,4-Diacetamido-2,4,6-trideoxyhexose	DATDH	Pili of *Neisseria meningitidis* and *Neisseria gonorrhoeae*
N-Acetylfucosamine	FucNAc	Pili of *Pseudomonas aeruginosa*

[a] Unless otherwise stated, all monosaccharides are of D-configuration and the ring form is pyranose

reducing monosaccharides, such as glucose, with the side chain amino group of lysine leads via initial Schiff bases and subsequent rearrangements to the formation of so-called advanced glycation end products (AGEs). AGEs have been implicated in alterations of proteins during aging and long-term diabetes [32,33] and are not dealt with in this chapter.

2.2.1 *N*-Glycosides

Until recently, the only type of *N*-glycosidic linkage that had been characterized in glycoproteins regardless of their origin was the *N*-acetylglucosaminyl(β1–N)asparagine bond discovered by Neuberger et al. [34] and almost simultaneously by others [35,36]. During biosynthesis, the enzyme oligosaccharyl transferase (OT) transfers a triantennary tetradecasaccharide ($Glc_3Man_9GlcNAc_2$) from dolichol pyrophosphate to the amide nitrogen of an Asn side chain in the nascent polypeptide [37]. In the subsequent trimming process this oligosaccharide is modified by action of several glycosylhydrolases and glycosyltransferases resulting in a structural diversity of glycans sharing a common core pentasaccharide (➲ *Fig. 1*). The primary peptide sequence requirements for OT-catalyzed glycosylation include a minimum Asn-Xaa-Ser/Thr (sometimes also Cys) tripeptide recognition motif where Xaa can be any of the 20 natural amino acids except proline. However, it is found that not all such consensus tripeptides (sequons) are glycosylated. A plausible explanation arises from mechanistic studies on the OT-catalyzed glycosylation of linear and constrained model peptides [38]. It could be demonstrated that only peptides which are able to adopt an Asx-turn conformation [39] are efficient substrates of OT leading to the proposed mechanism of oligosaccharyl transfer shown in ➲ *Fig. 2a* [40,41]. In this model, the apparent failure of the glycosylation machinery most likely results from conformational influence by the neighboring polypeptide sequence that may override the potential for the tripeptide acceptor sequence to adopt an Asx-turn and therefore limit its compatibility with the OT active site. However, although necessary, formation of an Asx-turn might not be sufficient [42]. An alternate mechanistic model which can also explain both specificity and enhanced amide nitrogen nucleophilicity was proposed by Bause

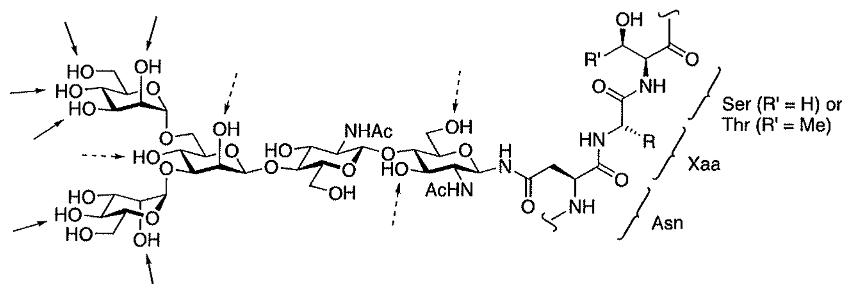


Figure 1
Core pentasaccharide Man(α1–6)[Man(α1–3)]Man(β1–4)GlcNAc(β1–4)GlcNAc of *N*-glycoproteins attached to the consensus sequence Asn-Xaa-Ser/Thr. *Solid arrows* indicate the points of attachment of the outer arm saccharides forming carbohydrate chains, called antennae. In addition, the inner-core may be substituted by several monosaccharides (*dashed arrows*)

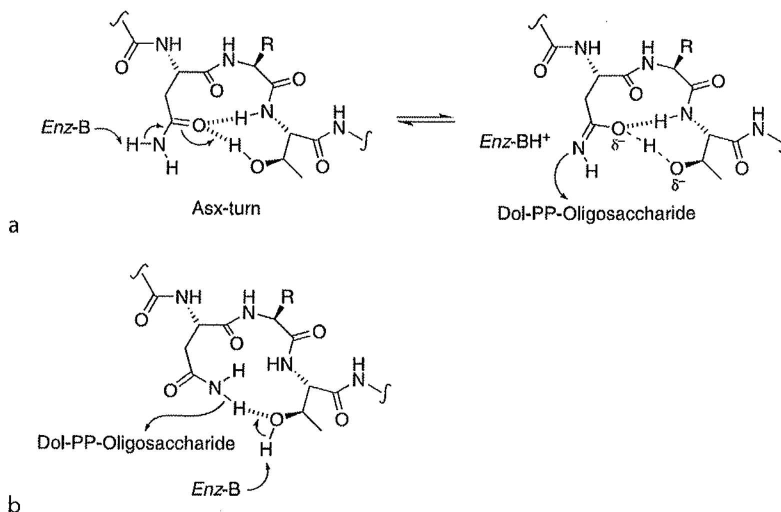


Figure 2
Proposed mechanisms of asparagine activation for oligosaccharyl transferase by (a) Imperiali et al. [40,41] and (b) Bause et al. [43,44,45]

et al. [43,44,45]. In this model the hydroxyamino acid acts as a hydrogen-bond acceptor rather than a donor (➲ *Fig. 2b*).

During recent years, linkages between asparagine side chains and other carbohydrates such as α- [46] and β-glucose [47], β-*N*-acetylgalactosamine [48] and L-rhamnose[49] have been discovered mainly in bacterial glycoproteins [50]. However, the β-glucosyl–asparagine linkage has been found also in the mammalian protein laminin [51]. Whenever carefully studied, the saccharides are attached to the sequon Asn-Xaa-Ser/Thr except for the glycopeptide nephritogenoside in which glucose is α-linked to the amide side chain of the *N*-terminal tripeptide Asn-Pro-Leu [46]. Another example of an *N*-glycosidic bond is the linkage of β-glucose to

the guanidino group of arginine which has been reported to occur in amylogenin, a self-glucosylating protein from sweet corn [52].

2.2.2 *O*-Glycosides

In contrast to *N*-glycoproteins, biosynthesis of the glycan chains of *O*-glycoproteins is an entirely post-translational process and starts with addition of a single monosaccharide to a hydroxy amino acid of the protein backbone [53]. Therefore, a variety of carbohydrate–protein linkages are found in *O*-glycoproteins (➲ *Fig. 3*). The *N*-acetylgalactosaminyl (α1–O) serine/threonine linkage (also called Tn antigen) was first demonstrated in mucins and is widely distributed in nature in the so-called mucin-type glycoproteins. An increasingly important *O*-glycosidic bond is that between β-*N*-acetylglucosamine and the serine/threonine hydroxyl [54,55,56,57]. This type of attachment is present predominantly in intracellular glycoproteins (nuclear pore, chromatin proteins, transcription factors, and cytoplasmic inclusions) and indeed represents the first reported example of glycosylated proteins found outside of the secretory channels. Interestingly, no other sugars are connected to the *N*-acetylglucosamine residue. The *N*-acetylglucosaminyl (β1–O) serine/threonine modification shares many common traits with protein phosphorylation. Both are dynamic modifications processed by specific enzymes that modify serine/threonine residues and rapidly respond to extracellular stimuli.

In the epidermal growth factor (EGF) modules of several blood coagulation and fibrinolytic proteins, L-fucose was characterized to be α-linked to serine or threonine in the consensus sequence Cys-Xaa-Xaa-Gly-Gly-Ser/Thr-Cys [58]. Later on, the consensus sequence was broadened to Cys-$(Xaa)_{3-5}$-Ser/Thr-Cys to account for observations made with the proteins Notch, Serrate/Jagged, and Delta [59]. EGF modules also contain glucose β-*O*-linked to serine within the consensus sequence Cys-Xaa-Ser-Xaa-Pro-Cys, an example being the bovine blood clotting factor IX [60]. α-Galactose bound to serine is found in plant glycoproteins like extensin or potato lectin and in the glycoproteins from the cellulosome, an extracellular complex of cellulases, produced by cellulolytic bacteria [61]. When human insulin-like growth factor-I (IGF-I) was expressed in yeast, a new form of IGF-I was characterized in addition to IGF-I in which mannose α-glycosidically was linked to threonine [62]. The mannosyl Ser/Thr linkage [63] has also been reported in Ser/Thr-rich domains of the bovine peripheral nerve α-dystroglycan [64] and in rat brain proteoglycans [65,66]. Xylosyl (β1–O) serine is found in animal proteoglycans and, at present, considered to be confined to these glycoproteins. Pseudaminic acid (5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-L-*manno*-nonulosinic acid) α-glycosidically linked to serine/threonine is an unusual bond that has been recently identified in *Campylobacter jejuni* flagellin [67,68]. It represents the first example of an acidic monosaccharide directly linked to protein. In the pili of *Neisseria meningitidis*, a 2,4-diacetamido-2,4,6-trideoxyhexose (DATDH) as part of the trisaccharide Gal(β1–4)Gal(α1–3)DATDH is linked to serine or threonine with unknown anomeric configuration [69]. In the pili of *Neisseria gonorrhoeae*, a disaccharide Gal(α1–3)DATDH has been reported to be *O*-glycosidically linked to serine [70] in contrast to an earlier publication in which Gal(α1–3)GlcNAc was assigned to be this disaccharide [71]. The *N*-acetylfucosaminyl (β1–O) serine/threonine linkage represents another recently described bacterial linkage that occurs in the pili of *Pseudomonas aeruginosa* [72].

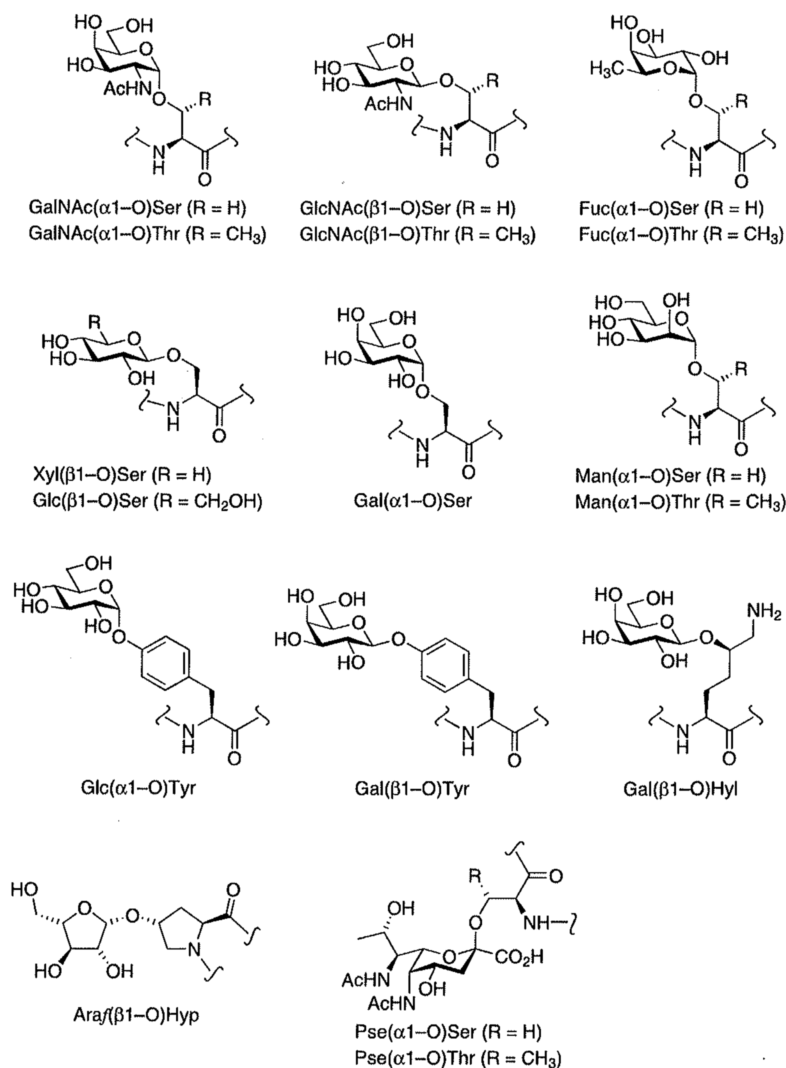


Figure 3
Carbohydrate–peptide linkages found in *O*-glycoproteins

O-Glycosylation is, however, not limited to serine and threonine. In glycogenin, the priming enzyme for glycogen synthesis, glucose is α-linked to the phenolic hydroxy group of a tyrosine residue [73,74] and galactosyl (β1–O) tyrosine has been identified in glycoproteins of the crystalline surface layers (*S*-layers) of eubacteria [75,76]. In the collagens, β-galactose is linked to hydroxylysine [77,78]. Hydroxyproline, finally, is glycosylated with α- [79] and β-L-arabinofuranose [80,81,82] in certain plant glycoproteins, with β-galactose in plants and eubacteria [83,84], and with *N*-acetylglucosamine in cytoplasmic glycoprotein of *Dictyostelium* [85].

2.2.3 GPI Anchors

Glycosyl-phosphatidylinositol anchors provide a means for anchoring proteins with a wide variety of structures and functions to the external surface of the plasma membrane of eukaryotic cells. The existence of lipid anchors for membrane proteins was initially demonstrated by the ability of phosphatidylinositol-specific phospholipase C (PI-PLC) to release alkaline phosphatase from various mammalian tissues into the medium [86,87]. Later, it was observed that fatty acids and ethanolamine were attached to the carboxy terminus of the rat Thy-1 antigen which provided direct evidence for a glycolipid anchor being covalently attached to protein [88]. In 1985 these and other results were combined to establish a common mode of membrane attachment via a GPI anchor covalently bound to the *C*-terminus through ethanolamine [89]. Today, we know well over 100 proteins which are GPI-anchored (reviewed in [90,91,92,93,94,95,96,97,98]) and the structures of several GPI anchors have been elucidated, the first being that on the *Trypanosoma brucei* variant surface glycoprotein (VSG) [99,100]. All characterized GPI anchors share a common core of ethanolamine-PO_4-6Man(α1–2) Man(α1–6)Man(α1–4)GlcN(α1–6)*myo*-Ino-1-PO_4 [99,100].

All characterized GPI anchors share a common core of ethanolamine-PO_4-6Man(α1–2) Man(α1–6)Man(α1–4)GlcN(α1–6)*myo*-Ino-1-PO_4-lipid. This may be further processed in a cell-type and protein-specific manner (➋ *Fig. 4*). The rat brain Thy-1 anchor, for example, contains additional Man(α1–2) and GalNAc(β1–4) residues, whereas the tetrasaccharide backbone of the trypanosome VSG anchor is modified with a branched chain of α-linked galactoses. In human erythrocyte acetylcholinesterase [101] and decay accelerating factor [102] the first and second mannoses carry additional phosphoethanolamines. Attached to the phosphoinositol are lipids of varying chain length and saturation including diacylglyc-

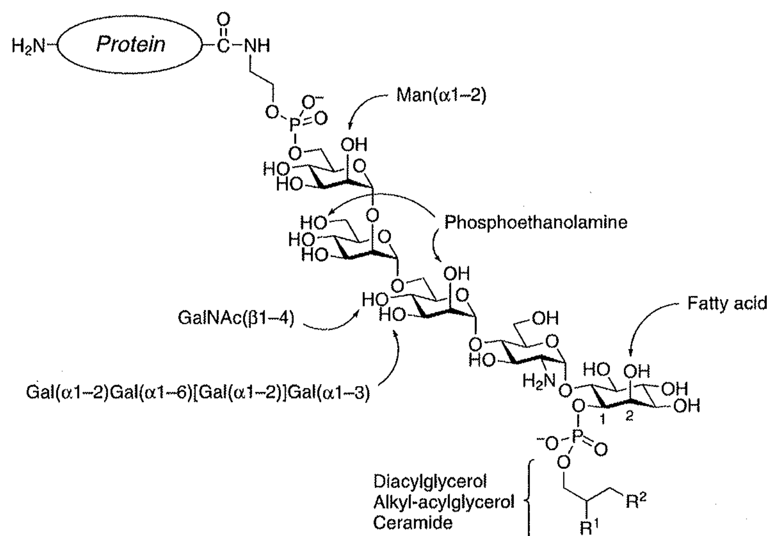


Figure 4
Minimal structure and some modifications of GPI anchors

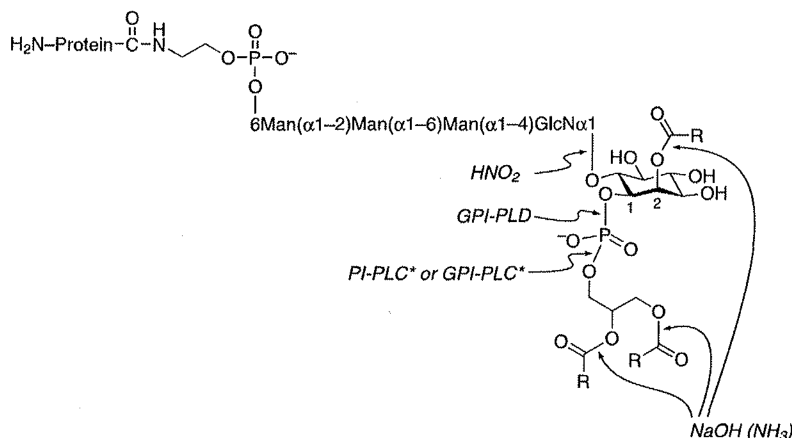


Figure 5

Enzymatic and chemical cleavage sites of GPI anchors most commonly used for the identification and structural analysis of GPI anchors. In addition, all phosphoric esters are cleavable by aqueous hydrofluoric acid. GPI-PLC*, GPI-specific phospholipase C; GPI-PLD, GPI-specific phospholipase D; PI-PLC*, phosphatidyinositol-specific phospholipase C. * Cleaves only if the inositol is unmodified

erol, alkyl-acylglycerol, stearoyl-*lyso*glycerol, or ceramide. In the VSG anchor, for example, dimyristyl glycerol has been identified. Furthermore, the inositol may be acylated with an additional fatty acid, most commonly palmitoyl, at position 2 of the inositol ring.

Biosynthesis of GPI anchors starts with the core structure assembly by sequential addition of UDP-GlcNAc (followed by *N*-deacetylation), dolichol-phosphate-mannose, and phosphoethanolamine to phosphatidylinositol and culminates in the en bloc transfer to protein shortly after the protein is synthesized. However, the biosynthetic pathways can differ strikingly between different organisms with respect to specific modifications and fatty acid remodeling occurring after completion of the core glycan. This also applies for the point when certain modifications are introduced, e. g. before or after the transfer of the GPI-moiety to the protein. GPI anchors can be cleaved at defined positions by an array of enzymatic and chemical methods, respectively (➋ *Fig. 5*). Thus, it becomes possible to identify GPI-anchored proteins and, moreover, analyze the structure and biosynthesis of GPI anchors [103].

The most obvious function of GPI anchors is to provide a stable, oriented attachment of proteins onto membranes, usually the extracellular surface of plasma membranes. However, there is much discussion over whether this highly conserved, multiple-step, complex mechanism for anchoring proteins has additional functions and it has been proposed [96,98,104,105] that GPI anchors:

- increase the lateral mobility of proteins,
- mediate the release or secretion of proteins,
- target proteins to apical surfaces,
- mediate endocytosis or protein turnover,
- play a role in protein sorting in the secretory and endocytic pathways, and
- participate in transmembrane signaling mechanisms.

In the meantime, several GPIs and GPI-anchored peptides have been synthesized as reviewed in [106].

2.2.4 Uncommon Linkages

In 1994 a previously unknown type of carbohydrate–protein linkage has been identified in human RNase 2 (identical to RNase U_s) with a mannose residue *α-C*-glycosidically attached to C2 of the indole ring of Trp-7 as a result of a post-translational modification (➔ *Fig. 6*) [107,108,109]. In the meantime, *C*-mannosylation of tryptophan has been found in several proteins, including interleukin-12 [110] and properdin [111]. Recently, a method predicting the location of *C*-mannosylation sites in proteins was developed using a neural network approach [112]. Another uncommon linkage is the attachment of sugar to protein via a phosphodiester bridge [113] which had been first characterized in the lysosomal proteinase I of *Dictyostelium discoideum* in which *α-N*-acetylglucosamine-1-phosphate is bound to the side chain hydroxyl of serine [114,115]. Attachment of *α*-mannosyl phosphate to serine has been observed in several major proteins of *Leishmania* species [116].

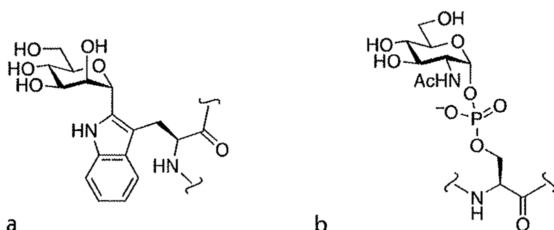


Figure 6
Uncommon carbohydrate–peptide linkages identified in (a) human RNase 2 and (b) proteinase I of *D. discoideum*

2.3 Methods of Glycoprotein-Glycans Analysis

Structural analysis of glycans contained in glycoproteins [117,118,119,120,121,122,123,124,125] is a challenging task due to the lack of a general oligosaccharide sequencing method like those available for peptides or oligonucleotides. The situation is further complicated by the phenomenon of microheterogeneity which makes only minute amounts of homogenous material available. Fractionation of individual glycoforms of a given glycoprotein can be achieved using capillary electrophoresis but is not generally applicable because the differences in physicochemical properties between neutral glycoforms are often relatively small. Therefore, protein glycosylation analysis is usually carried out after release of the glycans either from the whole glycoprotein or from glycopeptides obtained by proteolytic digestion. A general strategy consists of four steps:

1. Release of glycans from their conjugate polypeptide.
2. Labeling of released glycans.
3. Fractionation of glycan mixtures.
4. Sequencing of individual glycans.

2.3.1 Release of Glycans from Glycoprotein

In order to release glycans from a glycoprotein a general method is required that is independent of the protein to which the saccharides are attached. In this respect, chemical release is often preferred over the use of endoglycosidases or glycoamidases. Hydrazinolysis [126] has initially been described for release of *N*-glycans producing intact glycans with a free reducing terminus (➲ *Fig. 7*) but later on it was shown that under controlled conditions the method is also suitable for *O*-glycans [127]. In the case of *O*-glycans, however, to some extent degradation of reducing-end monosaccharides (so-called peeling) remains a problem. As long as a reducing saccharide is not required at the cleavage point, *O*-glycans can be more mildly released in the form of oligosaccharide alditols by β-elimination and subsequent reduction upon treatment with alkaline borohydride [128]. This procedure, however, prevents subsequent labeling of the glycans by reductive amination (see below). To avoid this drawback, modified β-elimination procedures that deliver glycans with intact reducing ends have been developed [129,130].

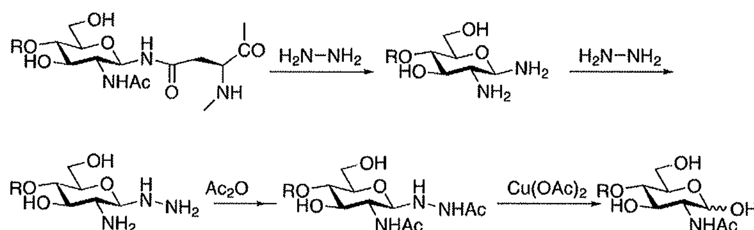


Figure 7
Hydrazinolytic release of *N*-linked glycans

Enzymatic release [131] of *N*-glycans is most frequently effected by the peptide *N*-glycanases (PNGases) F or A which, under denaturing conditions, generally cleave the $C\gamma$–$N\delta$ bond of glycosylated asparagines. Released are the intact *N*-linked glycans as glycosylamines, which are readily converted to regular glycans. Asparagine, in turn, is converted into aspartic acid resulting in a mass difference of 1 Dalton that may be used to assess the glycosylation site by mass spectrometry. In contrast to PNGase A, PNGase F is not able to cleave *N*-glycans containing an (α1–3)-linked fucose attached to the reducing-terminal GlcNAc residue. Alternatively, endoglycosidases can be applied which cleave between the first and second *N*-acetylglucosamine residue attached to asparagine. The various enzymes differ in their substrate specificities. Endo H, for example, cleave oligomannose-type and most hybrid-type glycans, whereas, Endo F2 and Endo F3 cleave certain complex-type chains [131]. Enzymatic release of *O*-glycans is much more difficult since most *O*-glycanases currently available are highly specific. Therefore, after PNGase treatment of *N*,*O*-glycoproteins, remaining *O*-linked glycans are often released by alkaline borohydride degradation.

2.3.2 Labeling and Fractionation of Glycans

Unless mass spectrometric analysis of the glycan pool is carried out, the released sugars need to be tagged to enable them to be detected during separation. Classically, the reducing end of the oligosaccharides had been reduced with alkaline sodium borotritide [132].

Today, radioactive labeling has been largely replaced by reductive amination with fluorescent compounds like 2-aminobenzamide [133], anthranilic acid [133], 8-aminonaphthalene-1,3,6-trisulfonic acid [134], 2-aminopyridine [135], 2-aminoacridone [136], or 1-aminopyrene-3,6,8-trisulfonate [137]. These fluorophores are compatible with a range of separation techniques including HPLC (normal phase, reversed phase, and weak anion exchange), high pH anion exchange chromatography (HPAEC), lectin-affinity chromatography, polyacrylamide gel electrophoresis (PAGE), capillary electrophoresis (CE), and gel permeation chromatography (GPC). Derivatization with ionizable functional groups has gained importance for high-resolution analysis of oligosaccharides by CE [138,139] and to enhance their otherwise low ionization efficiencies enabling sensitive detection of glycans by mass spectrometry even in the presence of peptides [140].

2.3.3 Sequencing of Glycans

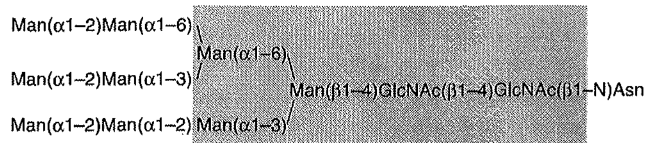
For complete structural analysis of oligosaccharides, a combination of several physical, chemical, and biochemical techniques including NMR spectroscopy, mass spectrometry (MS), and chemical and enzymatic degradation is required. Mass spectrometry has developed a key technology for the analysis of glycans in glycomics [119,120,121,122,125]. The advantages of MS techniques are high sensitivity (detection limits of oligosaccharides in the low pico- to femtomolar range), high accuracy, and the possibility of being directly coupled to many separation methods. ESI-MS/MS, ESI-ion trap-MS^n, and MALDI-TOF/TOF-MS are currently the most frequently used methods. Structural information may be obtained from two types of fragmentation: cleavages of glycosidic bonds between monosaccharide units provide information on glycan sequence and branching, whereas cleavages within sugar rings (so-called cross-ring cleavages) provide extended information of branching and monosaccharide linkage positions. It is important, however, to keep in mind that mass spectrometry is not able to distinguish between different stereoisomeric sugars such as GlcNAc/GalNAc or Gal/Man/Glc all of which have the same mass. Therefore, knowledge of biosynthetic pathways of the organism which the glycans were derived from is used to reduce the number of possible glycans that are in agreement with measured data. Several algorithms for automated interpretation of MS data with integrated links to glycan databases have been developed and progress in this area has been reviewed [141].

Enzymatic analysis using highly specific exoglycosidases is a powerful means of determining the sequence of glycans especially if only very small amounts of material (picomoles or less) are available. This method is based on the determination of the susceptibility of a glycan to a series of exoglycosidases of defined specificity. In the case of the reagent array analysis method (RAAM) [142,143], the process has been automated. The glycan is divided into several aliquots and each aliquot is incubated with a defined mixture of exoglycosidases called a reagent array. Degradation occurs in each vial until a linkage is reached which is resistant to all exoglycosidases present in that mix. The remaining "stop point" fragments containing the labeled reducing end saccharides are characterized by size (GPC or MS). From these data, a computer program constructs the sequence of the glycan. Later, an improved approach was developed involving the simultaneous digestion of aliquots of a total pool of fluorescently labeled oligosaccharides with a series of multiple enzyme arrays [144]. For further details on the analysis of glycoproteins, the reader is referred to ➲ Chap. 8.5 of this book.

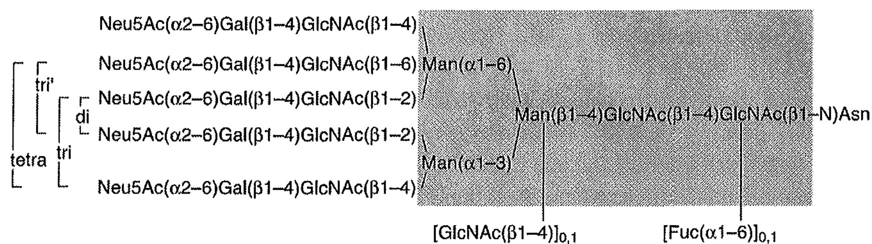
2.4 Structure of *N*-Glycans

Because of their biosynthesis, all *N*-glycoproteins share the common core pentasaccharide Man(α1–6)[Man(α1–3)]Man(β1–4)GlcNAc(β1–4)GlcNAc. The tremendous diversity of asparagine-linked oligosaccharides derives from attachment of several monosaccharides as well as up to five different carbohydrate chains, the so-called antennae, to the core. On the basis of the structure and the location of the glycans attached to the trimannosyl core, *N*-glycoproteins can be classified into three main groups (➋ *Fig. 8*). These are:

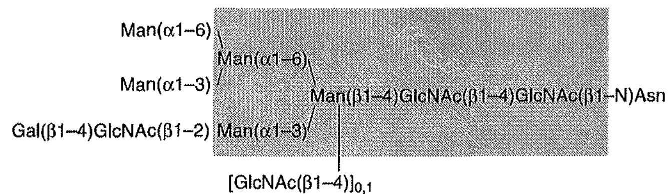
Oligomannose



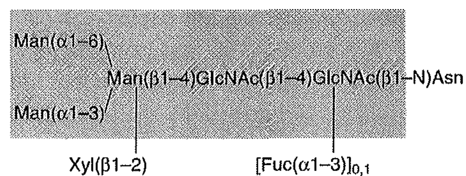
Complex



Hybrid



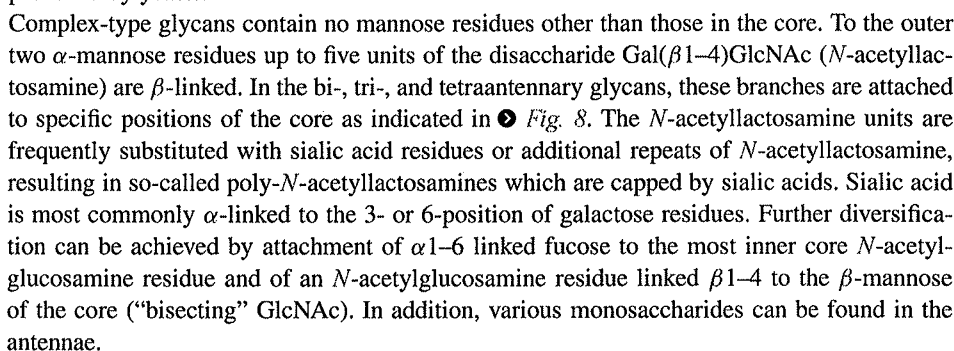
Xylose-containing



■ Figure 8
Examples of the main types of *N*-linked oligosaccharides of glycoproteins. The structure within the *gray box* represents the pentasaccharide core common to all *N*-glycans

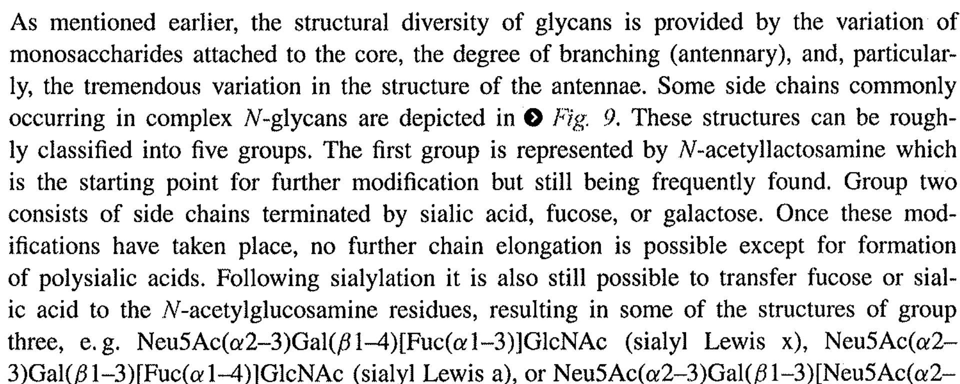
- oligomannose (high mannose),
- complex, and
- hybrid type.

It has been suggested to add the xylose-containing type as a fourth group [28].
Oligomannose-type glycans usually contain two to six α-mannose residues bound to the core. The largest high mannose oligosaccharide thus contains nine mannose and two *N*-acetylglucosamine residues which were originally discovered in bovine thyroglobulin [145]. High molecular weight oligomannose chains with up to 100–200 mannose residues are, however, produced by yeasts.



Complex-type glycans contain no mannose residues other than those in the core. To the outer two α-mannose residues up to five units of the disaccharide Gal(β1–4)GlcNAc (*N*-acetyllactosamine) are β-linked. In the bi-, tri-, and tetraantennary glycans, these branches are attached to specific positions of the core as indicated in ➲ *Fig. 8*. The *N*-acetyllactosamine units are frequently substituted with sialic acid residues or additional repeats of *N*-acetyllactosamine, resulting in so-called poly-*N*-acetyllactosamines which are capped by sialic acids. Sialic acid is most commonly α-linked to the 3- or 6-position of galactose residues. Further diversification can be achieved by attachment of α1–6 linked fucose to the most inner core *N*-acetylglucosamine residue and of an *N*-acetylglucosamine residue linked β1–4 to the β-mannose of the core ("bisecting" GlcNAc). In addition, various monosaccharides can be found in the antennae.
Oligosaccharides of the hybrid typehave characteristic features of both complex-type and oligomannose-type glycans. One or two mannose residues are α-linked to the Man(α1–6) branch of the core pentasaccharide and one or two antennae like those found in complex-type oligosaccharides are β-linked to the Man(α1–3) branch of the core. This type of glycan results from partial processing of oligomannose-type glycans produced during biosynthesis and subsequent addition of sugars.
The fourth group of *N*-glycans is named xylose-containing and is characterized by a xylose residue linked β1–2 to the β-mannose of the core. Often, there is also fucose α1–3 linked to the innermost *N*-acetylglucosamine of the core.

2.4.1 Diversity of *N*-Glycans



As mentioned earlier, the structural diversity of glycans is provided by the variation of monosaccharides attached to the core, the degree of branching (antennary), and, particularly, the tremendous variation in the structure of the antennae. Some side chains commonly occurring in complex *N*-glycans are depicted in ➲ *Fig. 9*. These structures can be roughly classified into five groups. The first group is represented by *N*-acetyllactosamine which is the starting point for further modification but still being frequently found. Group two consists of side chains terminated by sialic acid, fucose, or galactose. Once these modifications have taken place, no further chain elongation is possible except for formation of polysialic acids. Following sialylation it is also still possible to transfer fucose or sialic acid to the *N*-acetylglucosamine residues, resulting in some of the structures of group three, e.g. Neu5Ac(α2–3)Gal(β1–4)[Fuc(α1–3)]GlcNAc (sialyl Lewis x), Neu5Ac(α2–3)Gal(β1–3)[Fuc(α1–4)]GlcNAc (sialyl Lewis a), or Neu5Ac(α2–3)Gal(β1–3)[Neu5Ac(α2–

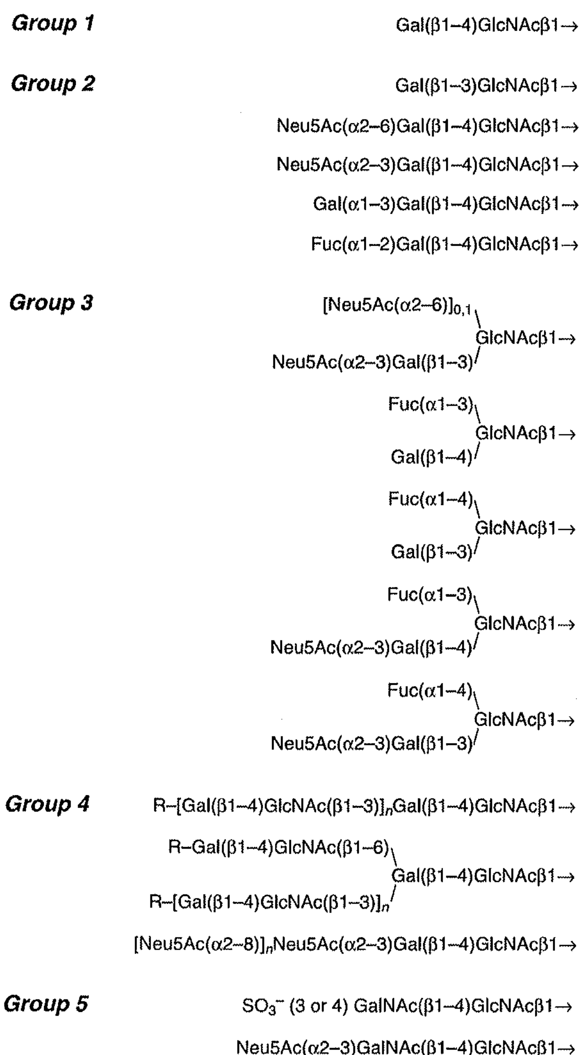


Figure 9
Commonly occurring structures of complex-type *N*-glycan side chains. The *arrows* indicate the point of attachment to the pentasaccharide core

6)]GlcNAc. On the other hand, fucosylation of a subterminal *N*-acetylglucosamine residue halts the chain elongation in nonsialylated, fucosylated form. Group four consists of long-chain structures containing either linear or branched poly-*N*-acetyllactosamine or α2–8 linked sialic acid repeatedly added to α2–3 linked sialic acid, forming polysialyl side chains. Group five side chains contain *N*-acetylgalactosamine instead of galactose. These structures may be further modified by sialylation or sulfation. There are many more variations in *N*-glycan chains identified to date and many more are expected to be found in future studies.

2.4.2 Cell-Type Specific Expression of *N*-Glycans

As described above, the enormous diversity of *N*-glycans is largely provided by variation of structures in the antennae. However, expression of specific carbohydrate epitopes, so-called antigenic determinants, is highly controlled and often cell-type specific. Critical findings revealed that poly-*N*-acetyllactosamines that are contained in many complex-type *N*-glycoproteins (but also in *O*-glycoproteins and glycolipids) provide a preferable backbone for many of the cell-type specific glycosylations, such as the ABH(O), I/i, and Lewis (Le^a, Le^x, and sialylated forms thereof) blood group and differentiation antigenic determinants [146]. Poly-*N*-acetyllactosamine repeats are not uniformly distributed among different antennae attached to the trimannosyl core. Because of the branch specificity of β1–3-*N*-acetylglucosaminyltransferase—the key enzyme for formation and elongation of poly-*N*-acetyllactosamine chains—poly-*N*-acetyllactosamine extensions are more common at the α1–6 linked mannose of the core. Tetraantennary glycans are particularly good acceptor substrates for poly-*N*-acetyllactosamine chain formation. Termination of chain elongation by sialylation leads to structures summarized in ➋ *Fig. 10*.

Glycosylation patterns of glycoproteins (and glycolipids) in human erythrocytes and granulocytes are well-established examples for cell-type specific expression of glycans. Some important carbohydrate antigenic determinants are listed in ➋ *Table 2*. Erythrocytes and granulocytes directly differentiate from the same precursor stem cells. In fetal erythrocytes, poly-*N*-acetyllactosamine is linear and expresses blood type i activity. In adult erythrocytes, poly-*N*-acetyllactosamines are branched at galactose, forming Gal(β1–4)GlcNAc(β1–3)[Gal(β1–4)GlcNAc(β1–6)]Gal(β1–4)GlcNAc, the blood group I antigen. The change from i to I specificity occurs gradually during the first year of life. Terminal galactose residues of the branched poly-*N*-acetyllactosamine chains of erythrocytes are then modified by α1–2 linked fucose, forming the H antigen. This reaction is catalyzed by an α1–2-fucosyltransferase. The H structure, Fuc(α1–2)Gal(β1–4/3)GlcNAc, is present in *O*-type blood group individuals. This structure now serves as the acceptor for α1–3-*N*-acetylgalactosaminyltransferase or α1–3-galactosyltransferase, forming A or B blood group antigens, respectively.

In granulocytes, no branching of poly-*N*-acetyllactosamines is observed. α1–3-Fucosyltransferase-catalyzed modification of the linear chains leads to the Lewis x antigen. If α2–3-sialyltransferase acts before fucosylation, sialyl Lewis x is formed. These carbohydrate structures are involved in a number of cell-cell interactions during the inflammatory response and

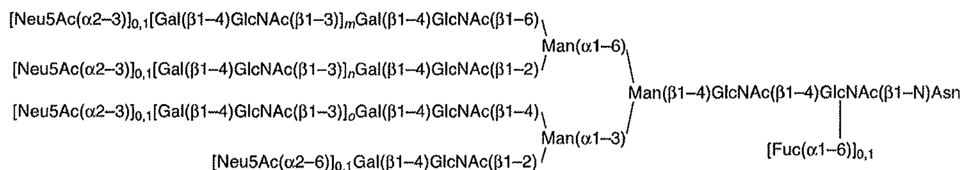


Figure 10
Representative structures of tetraantennary poly-*N*-acetyllactosamine containing *N*-glycans based on structures reported in various cells. The length of the *N*-acetyllactosamine repeats generally decreases in the order $m > n > o$. Sialic acids attached to poly-*N*-acetyllactosamine chains are usually α2–3 linked whereas α2–6 sialylation preferentially occurs at the side chain linked β1–2 to the Man(α1–3) branch of the core

Table 2
Antigenic determinants. LacNAc stands for Gal(β1–4)GlcNAc

Determinant structure	Name
Fuc(α1–2)Gal(β1–4/3)GlcNAc	Blood group H (O), type 1 or 2[a]
GalNAc(α1–3)[Fuc(α1–2)]Gal(β1–4/3)GlcNAc	Blood group A, type 1 or 2[a]
Gal(α1–3)[Fuc(α1–2)]Gal(β1–4/3)GlcNAc	Blood group B, type 1 or 2[a]
Gal(β1–3)[Fuc(α1–4)]GlcNAc	Blood group Lewis a (Le^a)
Neu5Ac(α2–3)Gal(β1–3)[Fuc(α1–4)]GlcNAc	Sialyl Lewis a (sLe^a)
Fuc(α1–2)Gal(β1–3)[Fuc(α1–4)]GlcNAc	Blood group Lewis b (Le^b)
Gal(β1–4)[Fuc(α1–3)]GlcNAc	Lewis x (Le^x)
Neu5Ac(α2–3)Gal(β1–4)[Fuc(α1–3)]GlcNAc	Sialyl Lewis x (sLe^x)
Fuc(α1–2)Gal(β1–4)[Fuc(α1–3)]GlcNAc	Lewis y (Le^y)
[LacNAc(β1–3)]$_n$	Blood group i
LacNAc(β1–3)[LacNAc(β1–6)]LacNAc	Blood group I

[a] Type 1 contains the sequence Gal(β1–3)GlcNAc, type 2 contains the sequence Gal(β1–4)GlcNAc

malignant transformations [147,148]. Production of Lewis x and sialyl Lewis x is restricted to myeloid cell lineage and only granulocytes and monocytes are enriched with them in blood cells. In contrast, expression of the ABH blood group antigens is restricted to erythroid cells, and no other blood cells contain ABH blood group antigens. This cell-type specific expression of glycans is caused by the presence of certain key glycosyltransferases that are essential for formation of the corresponding carbohydrate structures and of which expression is restricted to that particular cell type.

2.5 Structure of *O*-Glycans

The biosynthesis of *O*-glycans differs fundamentally from that of the *N*-glycans. *O*-Glycosylation does not begin with the transfer of an oligosaccharide from a dolichol precursor to the nascent peptide chain which is then further processed. Instead, *O*-glycosylation is mainly a post-translational and post-folding event and is initiated by glycosyltransferase-catalyzed addition of a single monosaccharide to a side chain hydroxyl of the protein. Depending on the carbohydrate–peptide linkage, various types of *O*-glycans can be distinguished. By definition, mucin-type *O*-glycoproteins are characterized by the GalNAc(α1–O)Ser/Thr bond [149,150]. This structural element, often referred to as the Tn antigen, was first identified in the mucins but is also found in other glycoproteins. Subsequent stepwise enzymatic glycosylation of the monoglycosylated polypeptide and further modification by sulfation and acetylation leads to a high degree of heterogeneity. Mucin-type *O*-glycan chains typically consist of one to 20 monosaccharide residues and they carry many of the blood group antigens and recognition signals required for intercellular and molecular interactions as well as cancer-associated and differentiation antigens.

2.5.1 Mucin-Type *O*-Glycans

Biosynthesis of mucin-type *O*-glycans starts with the transfer of *N*-acetylgalactosamine (GalNAc) from UDP-GalNAc to the side chain of serine or threonine catalyzed by an *N*-acetylgalactosaminyltransferase. More than 20 homologous *N*-acetylgalactosaminyltransferases are known to date [151]. These transferases are expressed in a tissue-specific manner and have overlapping but different substrate specificities. As a consequence, the site-specificity of mucin *O*-glycosylation is tissue specific and no simple consensus sequence can be found. However, based on statistical analysis of occupied *O*-glycosylation sites, some general rules can be deduced [9]. *O*-Glycosylation occurs preferentially at exposed regions of the protein surface like β-turns or regions with extended conformations. Such areas are usually of low hydrophobicity and, consequently, seldom contain amino acids like Trp, Leu, Ile, and Phe. Instead, Ser, Thr, and Pro are commonly found before and after occupied *O*-glycosylation sites. Also, no bulky amino acids are found near *O*-glycosylation sites, probably due to steric hindrance. However, in contrast to *N*-linked glycans which are usually well separated from each other, *O*-glycans frequently occur in clusters on short peptide sections consisting of repeating units of Ser, Thr, and Pro. It seems to be a general feature that Thr is glycosylated more efficiently than Ser. On the basis of statistical analyses, algorithms allowing the prediction of *O*-glycosylation sites have been developed [152]. The location of the subcellular compartment where *O*-glycosylation is initiated is still controversial and may depend on the type of *N*-acetylgalactosaminyltransferase. Several suggestions that have been made include subregions of the endoplasmic reticulum (ER), a proximal Golgi compartment, an intermediate ER-Golgi compartment (ERGIC), and the Golgi apparatus [153,154,155]. Subsequent stepwise enzymatic elongation of the monoglycosylated polypeptide then leads to several core structures, eight of which have been identified to date (➋ *Fig. 11*). However, under certain circumstances (e. g. in particular transformed cells), elongation does not occur and the Tn

Core 1 Gal(β1–3)GalNAc(α1–O)Ser/Thr

Core 2
GlcNAc(β1–6)\
GalNAc(α1–O)Ser/Thr
Gal(β1–3)/

Core 3 GlcNAc(β1–3)GalNAc(α1–O)Ser/Thr

Core 4
GlcNAc(β1–6)\
GalNAc(α1–O)Ser/Thr
GlcNAc(β1–3)/

Core 5 GalNAc(α1–3)GalNAc(α1–O)Ser/Thr

Core 6 GlcNAc(β1–6)GalNAc(α1–O)Ser/Thr

Core 7 GalNAc(α1–6)GalNAc(α1–O)Ser/Thr

Core 8 Gal(α1–3)GalNAc(α1–O)Ser/Thr

Figure 11
Core structures of *O*-glycans. Core 1 is also called T antigen

antigen is exposed which might be modified to sialyl Tn (sTn) (Neu5Ac(α2–6)GalNAc(α1–O)Ser/Thr) [156,157,158].

Depending on the cell type and its activation and differentiation status, the core structures can be further elongated or terminated by various glycosyltransferases resulting in a large number of *O*-glycans, of which some are depicted in ➋ *Fig. 12*. Many of these glycosyltransferases have been isolated and their genes have been cloned [159,160]. There are glycosyltransferases which specifically act on *O*-glycans (mainly those responsible for assembly of the core structures) and there are others with an activity restricted to *N*-glycans. However, a third group is able to assemble both *N*- and *O*-glycans. The antigenic determinants mentioned in ➋ *Sect. 2.4.2* are generated by such transferases, i. e. the structures listed in ➋ *Table 2* are found on both *N*- and *O*-glycans.

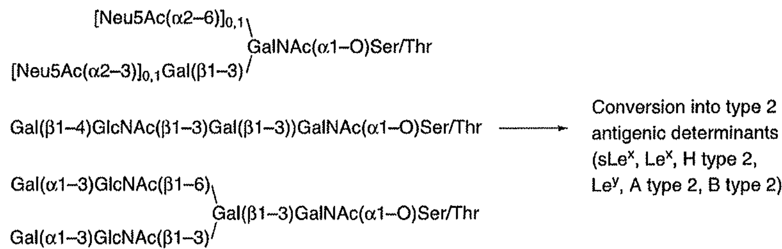
Core 1 and core 2 are the most common core structures in mucins and other secreted and cell surface glycoproteins. Core 1 (the T antigen) is usually not exposed but is monosialylated (Neu5Ac(α2–3)Gal(β1–3)GalNAc(α1–O)Ser/Thr and Neu5Ac(α2–6)[Gal(β1–3)]GalNAc(α1–O)Ser/Thr) or disialylated (Neu5Ac(α2–6)[Neu5Ac(α2–3)Gal(β1–3)]GalNAc(α1–O)Ser/Thr). Core 1 as well as other *O*-glycans is often elongated by *N*-acetyllactosamine, which is a substrate for conversion into the type 2 antigenic determinants sLe^x, Le^x, H type 2, Le^y, A type 2, and B type 2. Core 1 is, however, a poor acceptor for poly-*N*-acetyllactosamine formation. To efficiently achieve this type of extension, core 1 first has to be transformed into core 2. Then, poly-*N*-acetyllactosamine formation can take place at the new GlcNAc(β1–6) branch. Thus, the expression of core 2 *N*-acetylglucosaminyltransferase (core 2 GlcNAcT), the enzyme responsible for the conversion of core 1 into core 2, is rate limiting for the biosynthesis of poly-*N*-acetyllactosamine extensions. The fact that poly-*N*-acetyllactosamine extensions provide an excellent backbone for the expression of antigenic determinants involved in numerous recognition events highlights the biological importance of core 2 GlcNAcT.

So far, core 3 and 4 have been found only in mucins. Core 5 is contained in glycoproteins from several species and has been reported in human adenocarcinoma [161] and meconium [162,163]. It has been speculated that core 6 might be generated by a β-galactosidase-catalyzed degradation of core 2 [159]. Core 7 occurs in bovine submaxillary mucin [164]. All core structures may be unsubstituted or elongated and may carry terminal antigens. Elongation can result in type 1 chains (containing the sequence Gal(β1–3)GlcNAc) or type 2 chains (containing the sequence Gal(β1–4)GlcNAc, e. g. i antigen, poly-*N*-acetyllactosamines) and branched structures (I antigen). Neu5Ac(α2–3 and –6), Fuc(α1–2, –3, and –4), GalNAc(α1–3 and –6), GalNAc(β1–4), GlcNAc(α1–4), Gal(α1–3), and sulfated residues are typically found as terminal structures on *O*-glycans although sialic acids, sulfate, and Fuc(α1–3 or –4) may be also attached to internal carbohydrate residues. Many of the terminal structures are of the Lewis and ABH blood group system (➋ *Table 2*) and are also found on *N*-glycoproteins and glycolipids.

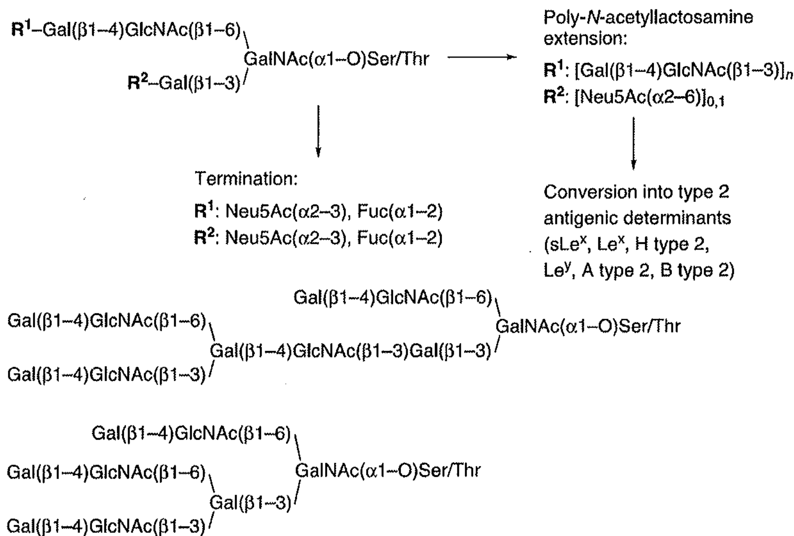
2.5.2 Non-Mucin *O*-Glycans

Beside the GalNAc(α1–O)Ser/Thr linkage of the mucin-type *O*-glycans a number of additional *O*-linkages have been identified. These include mainly fucose α-*O*-linked to Ser/Thr and glucose β-*O*-linked to Ser found in the epidermal growth factor domains of different proteins [58] and *N*-acetylglucosamine β-*O*-linked to Ser/Thr on cytoplasmic and nuclear proteins.

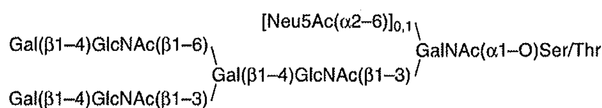
Structures generated from Core 1



Structures generated from Core 2



Structures generated from Core 3



Structures generated from Core 4

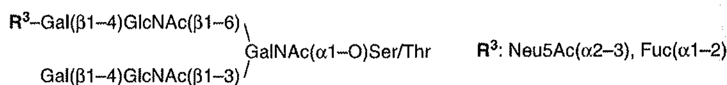


Figure 12
Structures of *O*-glycans commonly found in mucin-type *O*-glycoproteins

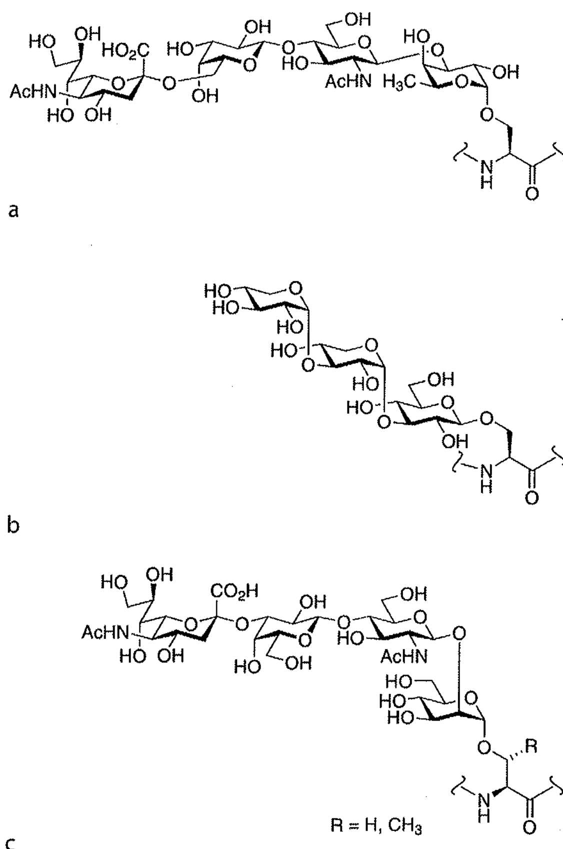


Figure 13
Some structures of non-mucin *O*-glycans (cf. text)

O-Linked fucose has been found on the EGF domains of urokinase, tissue plasminogen activator (tPA), factor VII, and factor XII, attached to the consensus sequence Cys-Xaa-Xaa-Gly-Gly-Ser/Thr-Cys. The fucosyltransferase responsible for the initiation of this type of glycosylation has been cloned [165]. On human (but not bovine) factor IX the tetrasaccharide Neu5Ac(α2–6)Gal(β1–4)GlcNAc(β1–3)Fuc(α1–O)Ser (➲ *Fig. 13a*) was found [166,167]. In other EGF domains only Fuc(α1–O)Ser/Thr was detected.

On the same EGF modules glucose *β*-*O*-linked to the Ser of the conserved consensus sequence Cys-Xaa-Ser-Xaa-Pro-Cys occurs. In several proteins like the human and bovine blood clotting factors VII and IX, human and bovine protein Z, tPA, thrombospondin, and murine fetal antigen 1 (mFA1) Xyl(α1–3)Xyl(α1–3)Glc(β1–O)Ser (➲ *Fig. 13b*) and substructures thereof (Xyl(α1–3)Glc(β1–O)Ser and Glc(β1–O)Ser) have been identified. The transferase activity responsible for addition of the second xylose has been detected [168].

The *N*-acetylglucosaminyl (β1–O) serine/threonine linkage is common on cytoplasmic and nuclear proteins [54,55,56,57]. No other sugars are connected to the GlcNAc residue. Formation of this type of modification is independent of the activity of the glycosylation machinery

in the endoplasmic reticulum and the Golgi apparatus. Addition of *O*-GlcNAc is catalyzed by a highly conserved *O*-GlcNAc transferase cloned independently by two groups [169,170]. The specificity of this enzyme resembles that of proline-directed kinases. *O*-GlcNAc is found in serine/threonine-rich sequences near Pro and Val residues. Often a negative charge is located nearby. Since the turnover of the *O*-GlcNAc glycan is faster than that of the peptide to which the sugar is attached, *O*-GlcNAc glycosylation is assumed to be reversible. Further evidence for this is the existence of a soluble *N*-acetyl-β-D-glucosaminidase [171]. It is likely that *O*-GlcNAc glycosylation plays a regulatory role that is analogous to protein phosphorylation. The mannosyl (α1–O) Ser/Thr linkage is uncommon in mammalian cells. In bovine peripheral nerve α-dystroglycan clusters of Neu5Ac(α2–3)Gal(β1–4)GlcNAc(β1–2)Man(α1–O)Ser/Thr (➋ *Fig. 13c*) have been observed in Ser/Thr-rich domains [64]. A similar glycan (Neu5Ac(α2–3)Gal(β1–4)GlcNAc(β1–3)Man(α1–O)Ser/Thr) has been found in rat brain proteoglycans [65,66].

2.6 Proteoglycans

Proteoglycans consist of glycosaminoglycan chains covalently attached to a protein backbone [172,173,174,175,176,177]. Therefore, they can be classified as glycoproteins. However, for historical reasons and because they differ substantially from other glycoproteins, they are usually treated as a separate class of compounds. Glycosaminoglycans are linear polysaccharides consisting of hexosamine residues like *N*-sulfonylglucosamine (GlcN*S*), *N*-acetylglucosamine (GlcNAc), or *N*-acetylgalactosamine (GalNAc) alternating with glucuronic acid (GlcA) or L-iduronic acid (IdoA) or galactose (Gal) residues. Different types of glycosaminoglycans are characterized by specific disaccharide repeat patterns. These oligosaccharides are substituted to varying degrees with sulfate linked to free amino and/or hydroxyl groups and range in size from ca. 15 disaccharide units to several hundred. The glycosaminoglycans contained in proteoglycans are chondroitin sulfate, dermatan sulfate, heparan sulfate, heparin, and keratan sulfate. Usually one type of glycosaminoglycan predominates but hybrid proteoglycans do exist. Another glycosaminoglycan (hyaluronan or hyaluronic acid) is not covalently attached to protein. A striking feature of all glycosaminoglycans is their negative charge caused by their sulfate and carboxyl groups which distinguishes these molecules from other glycoproteins. In addition to the glycosaminoglycan chains some proteoglycans contain *O*- and *N*-linked oligosaccharides similar or identical to those of various classes of glycoproteins. Proteoglycans are widely distributed in the animal and plant kingdom and they possess both structural and regulatory roles [177,178,179,180]. They are found in the extracellular matrix (ECM) and associated with specialized structures of the ECM like basement membranes. They also occur as part of plasma membranes where they function as receptors and co-receptors, and they are found intracellularly and intravesicularly in many hematopoietic cells. Proteoglycans have important physiological and homeostatic roles, e. g. during development, inflammation, and the immune response. Glycosaminoglycans such as heparan sulfate may encode information that is (cell) specific, spatially and temporally regulated, and instructive. Many diseases like chondrodystrophies [181], atherosclerosis [182], different types of cancer [183], or Alzheimer's disease [184] are related to abnormalities in the biosynthesis and processing of proteoglycans [172].

2.6.1 Carbohydrate–Peptide Linkage Region

Chondroitin sulfate, dermatan sulfate, heparan sulfate, and heparin are linked with their reducing end to the common core tetrasaccharide GlcA(β1–3)Gal(β1–3)Gal(β1–4)Xyl with the Xyl β-*O*-linked to a Ser residue (● *Fig. 14*), usually adjacent to a Gly in the protein backbone [172,173,174,177]. The GlcA residue of this tetrasaccharide is considered as part of the linkage region because the transferase involved in the addition of this residue to the Gal-Gal disaccharide is different from the enzyme that catalyzes the incorporation of GlcA into the rest of the glycosaminoglycan. The xylose may be phosphorylated and one or both galactose residues may be sulfated.

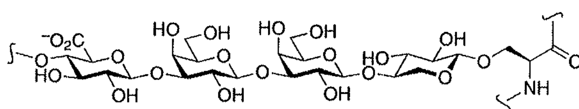


Figure 14
Structure of the linkage region GlcA(β1–3)Gal(β1–3)Gal(β1–4)Xyl(β1–O)Ser of proteoglycans

Skeletal keratan sulfate and corneal keratan sulfate are linked to core protein through *O*-linked and *N*-linked oligosaccharides, respectively, of the general types described in ● *Sect. 2.4* and ● *Sect. 2.5*.

2.6.2 Structure of Glycosaminoglycans

Hyaluronan (hyaluronic acid) is the simplest glycosaminoglycan. It consists of alternating GlcA and GlcNAc residues which are (β1–3) and (β1–4) linked, respectively (● *Fig. 15*) [172, 173,174,177]. Hyaluronan is not sulfated and is not bound to protein, thus it is not a component of proteoglycans.

Chondroitin sulfate glycosaminoglycans are composed of the disaccharide repeat unit GlcA(β1–3)GalNAc(β1–4) with variable amounts of sulfation at the positions indicated in ● *Fig. 15*. Dermatan sulfate is formed from the precursor chondroitin sulfate by action of an appropriate uronosyl epimerase which converts GlcA into IdoA, giving rise to the disaccharide repeat unit IdoA(α1–3)GalNAc(β1–4) [185]. Thus, both glycosaminoglycans are found on the same protein core. However, there is still no unambiguous convention as to what the proportion of IdoA must be to name a glycosaminoglycan dermatan sulfate. Ordinarily, IdoA is only formed adjacent to 4-sulfated GalNAc and not adjacent to 6-sulfated or non-sulfated GalNAc.

Heparan sulfate glycosaminoglycans and heparin are more complex. They contain two main types of disaccharide residues, although the proportions differ greatly between heparan sulfate and heparin. Heparan sulfate consists of areas of non-sulfated GlcA(β1–4)GlcNAc(α1–4) and areas of sulfated IdoA(α1–4)GlcN*S*(α1–4) disaccharide repeat units with the latter being obtained from the former by action of modifying enzymes during biosynthesis. Heparin differs from heparan sulfate by its higher content of IdoA, *N*-sulfate, and *O*-sulfate with most of the disaccharide units containing 2-sulfated IdoA alternating with GlcN*S* which is partially 6-sulfated and occasionally 3-sulfated as well. On some heparan sulfates occasional 3-sulfation of

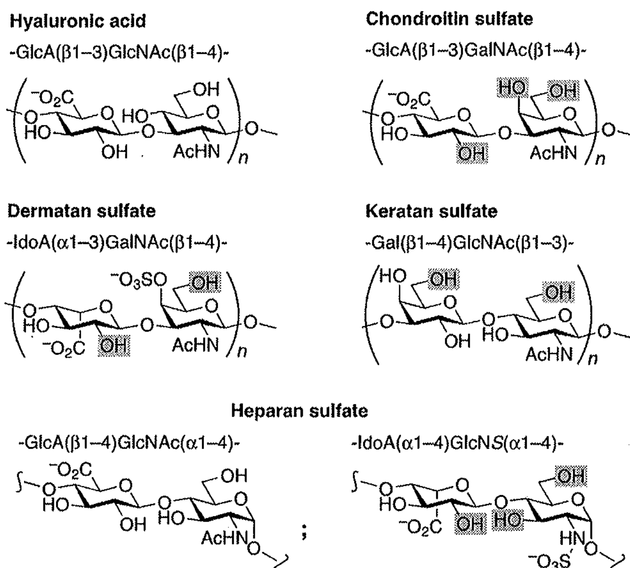


Figure 15
Disaccharide repeats of different glycosaminoglycans. Hydroxyl groups which may be modified by sulfation are marked gray

GlcN*S* is also found. Although heparan sulfate and heparin are structurally similar one should keep in mind that both are found on different core proteins.

Keratan sulfate glycosaminoglycans are sulfated poly-*N*-acetyllactosamines of the type found in *N*- and *O*-glycoproteins. The degree of sulfation which takes place at the primary hydroxyl groups is variable along the oligosaccharide chains with a higher degree at the non-reducing end.

3 Peptidoglycan

Peptidoglycan (synonym: murein) [186,187,188] is the essential cell wall polymer of most eubacteria, both Gram-positive and Gram-negative. Although it is not regarded as a glycopeptide it will be briefly treated in this chapter since it consists of carbohydrates covalently attached to peptides. Furthermore, knowledge of peptidoglycan structure is helpful for an understanding of the mode of action of the glycopeptide antibiotics of the vancomycin group. Peptidoglycan is composed of three structural features: a matrix of polysaccharide chains, the pentapeptide side chains, and the interstrand peptide cross-links, called interpeptide bridges (➋ *Fig. 16a*). The glycan component of all peptidoglycans consists of alternating β1–4 linked GlcNAc and *N*-acetylmuramic acid (MurNAc) (a GlcNAc with a lactyl ether at O-3) residues. Average glycan-chain lengths of ca. 8 to 40 disaccharide units have been reported depending on the method used for determining them [188,189,190]. This structure is common to all bacterial peptidoglycans examined, with only minor variations. In *Staphylococcus aureus* for

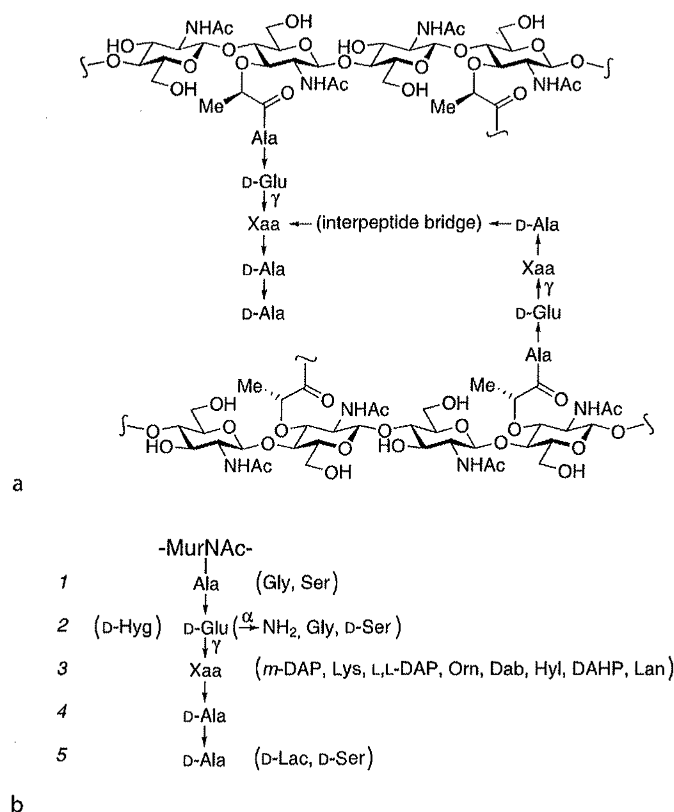


Figure 16
(a) Part of the common primary structure of bacterial peptidoglycan. *Arrows* indicate the polarity of peptide chains pointing from the *N*-terminus towards the *C*-terminus. (b) Variations in the pentapeptide chain. Residues in *parentheses* may replace corresponding amino acids or, in the case of α-carboxylate modification of γ-D-Glu, may be missing. *Abbreviations*: Dab, 2,4-diaminobutyric acid; DAHP, 2,6-diamino-3-hydroxypimelic acid; DAP, diaminopimelic acid; Hyg, *threo*-3-hydroxyglutamic acid; Hyl, hydroxylysine; Lac, lactate; Lan, lanthionine

example, up to 50% of the MurNAc residues are 6-*O*-acetylated [191] and in *Micrococcus lysodeikticus*, up to 40% of the MurNAc residues are not attached to peptide chains [192]. Attached to the carboxylate of MurNAc is a pentapeptide with the common sequence Ala-γ-D-Glu-Xaa-D-Ala-D-Ala (non-specification of the amino acid configuration implies the natural L-form), where Xaa is *meso*-diaminopimelic acid (*m*-DAP) for Gram-negative bacteria and Lys for most Gram-positive bacteria, with some variations [193]. During peptidoglycan biosynthesis, the pentapeptide is cross-linked through amino acid Xaa to another peptide strand. This transpeptidation involves attack of the ε-amino group of the *m*-DAP or Lys residue (or amino terminus of the interpeptide bridge) onto the penultimate D-Ala of another chain, resulting in loss of the terminal D-Ala of the second peptide chain and peptide bond formation. As mentioned, variations in the pentapeptide side chains are possible and they are summarized in *Fig. 16b*. Variations in position 1 are rare and γ-D-Glu is found universally at position 2.

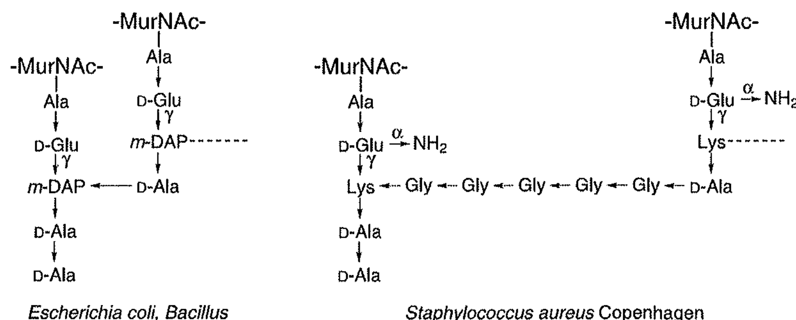


Figure 17
Peptide cross-links of *Escherichia coli*, *Bacillus*, and *Staphylococcus aureus* Copenhagen peptidoglycan

However, its α-carboxylate is often amidated or linked to an additional amino acid and in *Microbacterium lacticum* γ-D-Glu is replaced by *threo*-3-hydroxy-D-glutamic acid (3-Hyg). Most variation is found in position 3. If *m*-DAP (D,L-configuration) is incorporated, the L center is found in the pentapeptide chain and the D center in the side chain. Positions 4 and 5 (D-Ala-D-Ala) were for a long time thought to be invariant. However, the incidence of bacterial resistance to vancomycin-type glycopeptide antibiotics which recognize specifically the *N*-acyl-D-Ala-D-Ala terminus led to the discovery that these resistant strains contain altered substituents at position 5 as indicated in *Fig. 16b* [194,195,196].

Peptidoglycan strands are connected to varying degrees by a network of cross-links formed between the pentapeptide side chains. The structures of these cross-links which are quite heterogeneous have been reviewed [186,189,193] and this feature has been used for taxonomic classification of bacteria [193]. Here only two examples are presented (*Fig. 17*). In *Escherichia coli* and strains of *Bacillus* the simplest type of linkage is realized, i. e. the direct attachment of the ε-amino group of *m*-DAP to D-Ala at position 4 of the second peptide. If L-lysine is found in position 3 of the peptide chain, which is the case for most Gram-positive bacteria, usually 1–5 amino acid residues (L-configured or glycine) are incorporated between the ε-amino function of Lys and the D-Ala of the second peptide . As such, the interpeptide bridge of the well-studied *Staphylococcus aureus* Copenhagen strain consists of pentaglycine. However, other strains of *Staphylococcus* contain Ala, Ser, or Thr.

Currently, there is no generally accepted model for the three-dimensional structure of peptidoglycan. Two alternative models in which the glycan strands are arranged either parallel (classical model) or perpendicular ('scaffold' model) to the cytoplasmic membrane are discussed [188,190]. However, the models do not need to be mutually exclusive. It may also be that the cell wall exists as a mosaic of microdomains with different structures, or perhaps new cell wall is inserted in one form before being converted to another [197].

4 Glycopeptide Antibiotics

The glycopeptide antibiotics are glycosylated secondary metabolites of bacteria and fungi that are synthesized by non-ribosomal peptide synthetases. The term glycopeptide antibiot-

ic often solely refers to vancomycin and its analogues excluding other compounds, such as the bleomycins, ramoplanin, and the mannopeptimycins, that also belong to the group of glycopeptide antibiotics [198]. This section, however, will focus on the vancomycin group because it represents the biggest class of glycopeptide antibiotics.

4.1 Discovery and Medical Use

Vancomycin was discovered in the mid-1950s in a soil sample from the jungles of Borneo by coworkers of the American pharmaceutical company Eli Lilly [199]. It is produced as a secondary metabolite by the microorganism *Streptomyces orientalis* (reclassified as *Amycolatopsis orientalis*) [200] and displays antibiotic activity against the majority of Gram-positive bacteria. The discovery of vancomycin was followed a year later by the isolation of ristocetin [201]. These two antibiotics were recognized as belonging to a chemical class of antibiotics called glycopeptides [194,198,202,203,204]. Vancomycin was first used clinically in 1959. However, toxicity problems due to inadequacies during its purification precluded its widespread use in the early years and vancomycin was for the moment overshadowed by semisynthetic penicillins and later cephalosporins. However, with improved purification techniques and as the result of the emergence of multiple resistant and methicillin resistant *Staphylococcus aureus* (MRSA), vancomycin has become the drug of choice in the fight against these 'superbugs' [203]. Ristocetin appeared to be a promising antibacterial agent in the late 1950s but it was withdrawn following the death of several patients [205]. Avoparcin has shown growth-promoting activity in farm animals and has been marketed in Europe since 1976 as a feed additive. It is also used in the dairy industry for improvement of milk production of lactating dairy cows. Today two members of the group of the glycopeptide antibiotics are in clinical use—vancomycin and teicoplanin. Together with the aminoglycoside gentamycin, they are the antibiotics of last resort in our hospitals. However, in the meantime vancomycin resistant *S. aureus* strains are also being observed [206,207].

4.2 Structure and Classification

The first structural work on vancomycin were degradation studies carried out by Marshall [208] followed by extensive NMR examinations in the laboratory of Williams [209]. On the basis of these pioneering studies and on the X-ray analysis of the degradation product CDP-1 [210], Harris and Harris published the complete structure of vancomycin in 1982 [211]. The determination of the vancomycin structure then served as the basis for the structural characterization of hundreds of natural and semisynthetic glycopeptide antibiotics.

The glycopeptide antibiotics of the vancomycin group consist of a peptide backbone of 7 amino acids (see ➲ *Fig. 18* for selected examples). The side chains of amino acids 2 and 4 as well as 4 and 6 and in most cases also 5 and 7 are linked to each other, either via biaryl ether or direct biaryl bonds. The resulting structural elements give rise to the phenomenon of atropisomerism. Bound to this peptide core are 1–4 saccharides (mono- to tetrasaccharides). The vancomycin-type glycopeptide antibiotics can be subdivided in four or five groups based on chemical structure [198,204,212]. Group I, or the vancomycin type, has aliphatic amino acids at positions 1 and 3 while groups II–IV, illustrated by avoparcin, ristocetin, and

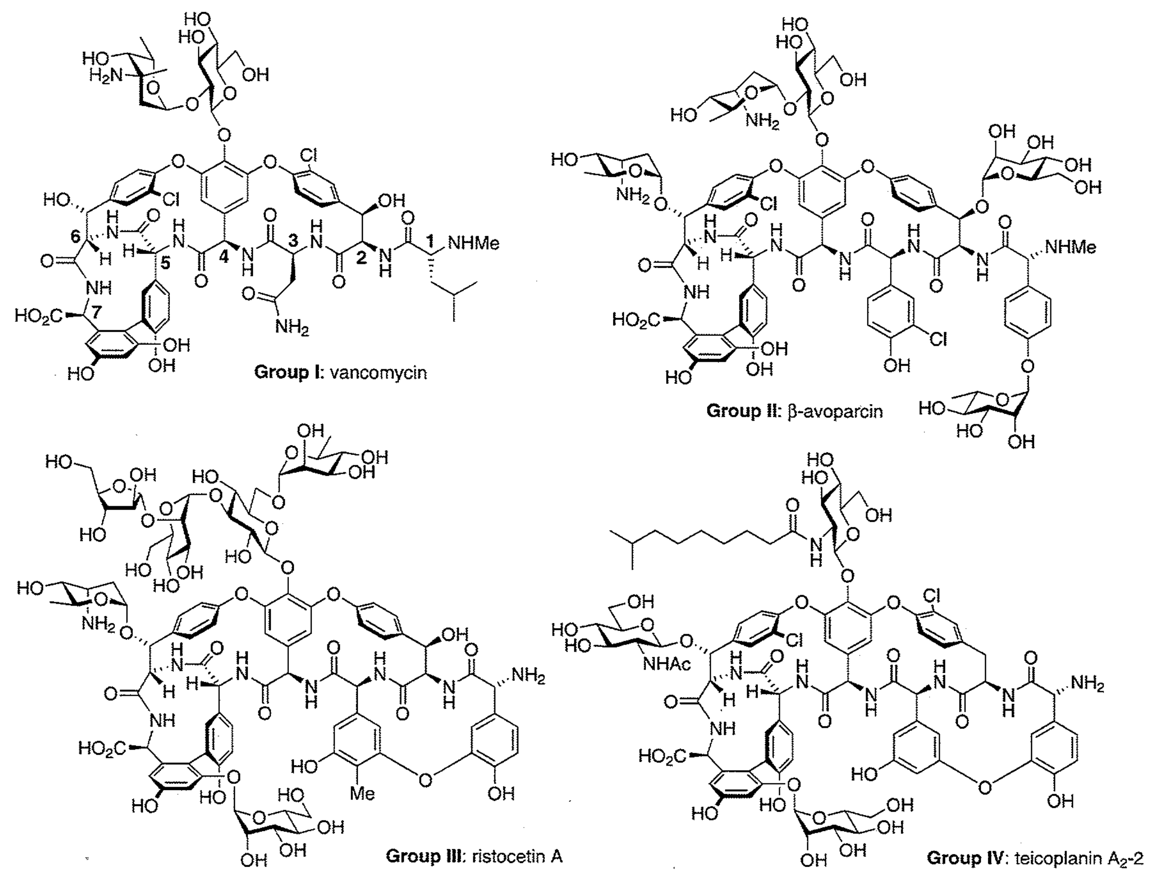


Figure 18
Selected structures of vancomycin-type glycopeptide antibiotics belonging to groups I–IV

teicoplanin, respectively, are characterized by aromatic residues at these positions. Members of group III and IV have an additional biaryl ether bridge between the side chains of amino acids 1 and 3. Type IV glycopeptide antibiotics could be considered a subgroup of III since the arrangement of the amino acids in the peptide core is the same. The antibiotics of this group have a fatty acid residue attached to an amino sugar. Finally, a fifth type of antibiotics can be defined containing a characteristic tryptophan residue at position 2. However, the members of this group identified so far (chloropeptin I, complestatin, and kistamycin A and B) are not glycosylated.

More than four centuries after its discovery and some 20 years after its structure elucidation, total syntheses of vancomycin [213,214,215,216,217] and its aglycon [218,219,220] have been accomplished by several research groups.

References

1. Fukuda M, Hindsgaul O (eds) (1994) Molecular Glycobiology. IRL Press, Oxford
2. Montreuil J, Vliegenthart JFG, Schachter H (eds) (1995) Glycoproteins. Elsevier, Amsterdam
3. Montreuil J, Vliegenthart JFG, Schachter H (eds) (1997) Glycoproteins II. Elsevier, Amsterdam
4. Gabius HJ, Gabius S (eds) (1997) Glycosciences: Status and Perspectives. Chapman & Hall, Weinheim
5. Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J (eds) (1999) Essentials of Glycobiology. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
6. Lis H, Sharon N (1993) Eur J Biochem 218:1
7. Dwek RA (1996) Chem Rev 96:683
8. Rudd PM, Dwek RA (1997) Crit Rev Biochem Mol Biol 32:1
9. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G (1998) Crit Rev Biochem Mol Biol 33:151
10. Taylor ME, Drickamer K (2006) Introduction to Glycobiology, 2 edn. Oxford University Press, Oxford
11. Lehle L, Strahl S, Tanner W (2006) Angew Chem 118:6956
12. Wittmann V (ed) (2007) Glycopeptides and Glycoproteins: Synthesis, Structure, and Application (Topics in Current Chemistry, vol. 267). Springer-Verlag, Berlin Heidelberg New York
13. Arnold JN, Wormald MR, Sim RB, Rudd PM, Dwek RA (2007) Annu Rev Immunol 25:21
14. Hitchen PG, Dell A (2006) Microbiology 152:1575
15. Szymanski CM, Wren BW (2005) Nat Rev Microbiol 3:225
16. Upreti RK, Kumar M, Shankar V (2003) Proteomics 3:363
17. Benz I, Schmidt MA (2002) Mol Microbiol 45:267
18. Brennan SO, Myles T, Peach RJ, Donaldson D, George PM (1990) Proc Natl Acad Sci USA 87:26
19. Schmid K, Binette JP, Kamiyama S, Pfister V, Takahashi S (1962) Biochemistry 1:959
20. Fournet B, Strecker G, Montreuil J, Vliegenthart JFG, Schmid K, Binette JP (1978) Biochemistry 17:5206
21. Van Halbeek H, Dorland L, Vliegenthart JFG, Montreuil J, Fournet B, Schmid K (1981) J Biol Chem 256:5588
22. Kato Y, Iwase H, Hotta K (1984) Anal Biochem 138:437
23. Ashford DA, Dwek RA, Rademacher TW, Lis H, Sharon N (1991) Carbohydr Res 213:215
24. Lütteke T, Bohne-Lang A, Loss A, Goetz T, Frank M, von der Lieth C-W (2006) Glycobiology 16:71R
25. Hashimoto K, Goto S, Kawano S, Aoki-Kinoshita KF, Ueda N, Hamajima M, Kawasaki T, Kanehisa M (2006) Glycobiology 16:63R
26. Raman R, Venkataraman M, Ramakrishnan S, Lang W, Raguram S, Sasisekharan R (2006) Glycobiology 16:82R
27. Varki A (1993) Glycobiology 3:97
28. Vliegenthart JFG, Montreuil J (1995) Primary structure of glycoprotein glycans. In: Montreuil J, Vliegenthart JFG, Schachter H (eds) Glycoproteins. Elsevier, Amsterdam, p 13

29. Sharon N, Lis H (1997) Glycoproteins: Structure and Function. In: Gabius H-J, Gabius S (eds) Glycosciences: Status and Perspectives. Chapman & Hall, Weinheim, p 133
30. Spiro RG (2002) Glycobiology 12:43R
31. Taylor CM (1998) Tetrahedron 54:11317
32. Cho SJ, Roman G, Yeboah F, Konishi Y (2007) Curr Med Chem 14:1653
33. Bierhaus A, Hofmann MA, Ziegler R, Nawroth PP (1998) Cardiovasc Res 37:586
34. Johansen PG, Marshall RD, Neuberger A (1961) Biochem J 78:518
35. Nuenke RH, Cunningham LW (1961) J Biol Chem 236:2451
36. Yamashina I, Makino M (1962) J Biochem 51:359
37. Helenius A, Aebi M (2004) Ann Rev Biochem 73:1019
38. Imperiali B, Shannon KL, Rickert KW (1992) J Am Chem Soc 114:7942
39. Abbadi A, McHarfi M, Aubry A, Premilat S, Boussard G, Marraud M (1991) J Am Chem Soc 113:2729
40. Imperiali B, Shannon KL, Unno M, Rickert KW (1992) J Am Chem Soc 114:7944
41. Imperiali B (1997) Acc Chem Res 30:452
42. Gibbs BS, Coward JK (1999) Bioorg Med Chem 7:441; see also corrigendum on page 2121
43. Bause E, Legler G (1981) Biochem J 195: 639
44. Bause E, Breuer W, Peters S (1995) Biochem J 312:979
45. Bause E, Wesemann M, Bartoschek A, Breuer W (1997) Biochem J 322:95
46. Shibata S, Takeda T, Natori Y (1988) J Biol Chem 263:12483
47. Wieland F, Heitzer R, Schaefer W (1983) Proc Natl Acad Sci USA 80:5470
48. Paul G, Lottspeich F, Wieland F (1986) J Biol Chem 261:1020
49. Messner P, Sleytr UB (1988) FEBS Lett 228:317
50. Review: Messner P (1997) Glycoconjugate J 14:3
51. Schreiner R, Schnabel E, Wieland F (1994) J Cell Biol 124:1071
52. Singh DG, Lomako J, Lomako WM, Whelan WJ, Meyer HE, Serwe M, Metzger JW (1995) FEBS Lett 376:61
53. Wopereis S, Lefeber DJ, Morava E, Wevers RA (2006) Clin Chem 52:574
54. Hart GW (1997) Annu Rev Biochem 66:315
55. Snow DM, Hart GW (1998) Int Rev Cytol 181:43
56. Zachara NE, Hart GW (2002) Chem Rev 102:431
57. Zachara NE, Hart GW (2004) Biochim Biophys Acta 1673:13
58. Harris RJ, Spellman MW (1993) Glycobiology 3:219
59. Shao L, Haltiwanger RS (2003) Cell Mol Life Sci 60:241
60. Hase S, Nishimura H, Kawabata S-I, Iwanaga S, Ikenaka T (1990) J Biol Chem 265:1858
61. Gerwig GJ, Kamerling JP, Vliegenthart JFG, Morag E, Lamed R, Bayer AE (1992) Eur J Biochem 205:799
62. Gellerfors P, Axelsson K, Helander A, Johansson S, Kenne L, Lindqvist S, Pavlu B, Skottner A, Fryklund L (1989) J Biol Chem 264:11444
63. Endo T (1999) Biochim Biophys Acta 1473: 237
64. Chiba A, Matsumura K, Yamada H, Inazu T, Shimizu T, Kusunoki S, Kanazawa I, Kobata A, Endo T (1997) J Biol Chem 272:2156
65. Krusius T, Finne J, Margolis RK, Margolis RU (1986) J Biol Chem 261:8237
66. Margolis RK, Margolis RU (1993) Experientia 49:429
67. Thibault P, Logan SM, Kelly JF, Brisson J-R, Ewing CP, Trust TJ, Guerry P (2001) J Biol Chem 276:34862
68. Young KT, Davis LM, DiRita VJ (2007) Nat Rev Microbiol 5:665
69. Stimson E, Virji M, Makepeace K, Dell A, Morris HR, Payne G, Saunders JR, Jennings MP, Barker S, Panico M, Blench I, Moxon ER (1995) Mol Microbiol 17:1201
70. Hegge FT, Hitchen PG, Aas FE, Kristiansen H, Lovold C, Egge-Jacobsen W, Panico M, Leong WY, Bull V, Virji M, Morris HR, Dell A, Koomey M (2004) Proc Natl Acad Sci USA 101:10798
71. Parge HE, Forest KT, Hickey MJ, Christensen DA, Getzoff ED, Tainer JA (1995) Nature 378:32
72. Castric P, Cassels FJ, Carlson RW (2001) J Biol Chem 276:26479
73. Alonso MD, Lomako J, Lomako WM, Whelan WJ (1995) FASEB J 9:1126
74. Lomako J, Lomako WM, Whelan WJ (2004) Biochim Biophys Acta 1673:45
75. Bock K, Schuster-Kolbe J, Altman E, Allmaier G, Stahl B, Christian R, Sleytr UB, Messner P (1994) J Biol Chem 269:7137
76. Messner P, Christian R, Neuninger C, Schulz G (1995) J Bacteriol 177:2188

77. Butler WT, Cunningham LW (1966) J Biol Chem 241:3882
78. Spiro RG (1967) J Biol Chem 242:4813
79. Yamagishi T, Matsuda K, Watanabe T (1976) Carbohydr Res 50:63
80. Lamport DTA (1967) Nature 216:1322
81. Lamport DTA, Miller DH (1971) Plant Physiol 48:454
82. Allen KA, Desai NN, Neuberger A, Creeth JM (1978) Biochem J 171:665
83. Strahm A, Amado R, Neukom H (1981) Phytochem 20:1061
84. Qi W, Fong C, Lamport DTA (1991) Plant Physiol 96:848
85. Teng-umnuay P, Morris HR, Dell A, Panico M, Paxton T, West CM (1998) J Biol Chem 273:18242
86. Ikezawa H, Yamanegi M, Taguchi R, Miyashita T, Ohyabu T (1976) Biochim Biophys Acta 450:154
87. Low MG, Finean JB (1977) Biochem J 167:281
88. Campbell DG, Gagnon J, Reid KBM, Williams AF (1981) Biochem J 195:15
89. Ferguson MAJ, Low MG, Cross GA (1985) J Biol Chem 260:14547
90. Ferguson MAJ, Williams AF (1988) Annu Rev Biochem 57:285
91. Cross GAM (1990) Annu Rev Immunol 8:83
92. Englund PT (1993) Annu Rev Biochem 62:121
93. McConville MJ, Ferguson MAJ (1993) Biochem J 294:305
94. Field MC, Menon AK (1993) Glycolipid anchoring of cell surface proteins. In: Schlesinger MJ (ed) Lipid Modifications of Proteins. CRC Press, Boca Raton, FL, p 83
95. Eckert V, Gerold P, Schwarz RT (1997) GPI-Anchors: Structure and Functions. In: Gabius H-J, Gabius S (eds) Glycosciences: Status and Perspectives. Chapman & Hall, Weinheim, p 223
96. Cole RN, Hart GW (1997) Glycosyl-phosphatidylinositol anchors: structure, biosynthesis and function. In: Montreuil J, Vliegenthart JFG, Schachter H (eds) Glycoproteins II. Elsevier, Amsterdam, p 69
97. Ikezawa H (2002) Biol Pharm Bull 25:409
98. Hoessli DC, Tauzin S, Nasir ud D, Borisch B (2007) Curr Org Chem 11:619
99. Ferguson MAJ, Homans SW, Dwek RA, Rademacher TW (1988) Science 239:753
100. Homans SW, Ferguson MAJ, Dwek RA, Rademacher TW, Anand R, Williams AF (1988) Nature 333:269
101. Deeg MA, Humphrey DR, Yang SH, Ferguson TR, Reinhold VN, Rosenberry TL (1992) J Biol Chem 267:18573
102. Walter BI, Roberts WL, Rosenberry TI, Ratnoff WD, Medof ME (1990) J Immunol 144:1030
103. Ferguson MAJ (1992) Chemical and enzymic analysis of glycosyl-phosphatidylinositol anchors. In: Hooper NM, Turner AJ (eds) Lipid Modification of Proteins: A Practical Approach. IRL Press, Oxford, p 191
104. Mayor S, Riezman H (2004) Nat Rev Mol Cell Biol 5:110
105. Baumann NA, Menon AK, Rancour DM (2000) Functions of Glycosyl Phosphatidylinositols. In: Ernst B, Hart GW, Sinaÿ P (eds) Carbohydrates in Chemistry and Biology, vol 4. Wiley-VCH, Weinheim, p 757
106. Guo Z, Bishop L (2004) Eur J Org Chem 3585-3596
107. Hofsteenge J, Mueller DR, de Beer T, Löffler A, Richter WJ, Vliegenthart JFG (1994) Biochemistry 33:13524
108. de Beer T, Vliegenthart JFG, Löffler A, Hofsteenge J (1995) Biochemistry 34:11785
109. Loeffler A, Doucey M-A, Jansson AM, Mueller DR, de Beer T, Hess D, Meldal M, Richter WJ, Vliegenthart JFG, Hofsteenge J (1996) Biochemistry 35:12005
110. Doucey M-A, Hess D, Blommers MJJ, Hofsteenge J (1999) Glycobiology 9:435
111. Hartmann S, Hofsteenge J (2000) J Biol Chem 275:28569
112. Julenius K (2007) Glycobiology 17:868
113. Haynes PA (1998) Glycobiology 8:1
114. Gustafson GL, Gander JE (1984) Methods Enzymol 107:172
115. Mehta DP, Ichikawa M, Salimath PV, Etchison JR, Haak R, Manzi A, Freeze HH (1996) J Biol Chem 271:10897
116. Guha-Niyogi A, Sullivan DR, Turco SJ (2001) Glycobiology 11:45R
117. Hounsell EF (1997) Methods of Glycoconjugate Analysis. In: Gabius H-J, Gabius S (eds) Glycosciences: Status and Perspectives. Chapman & Hall, Weinheim, p 15
118. Rudd PM, Dwek RA (1997) Curr Opin Biotechnol 8:488
119. Kameyama A (2006) Trends Glycosci Glycotechnol 18:323
120. Haslam SM, North SJ, Dell A (2006) Curr Opin Struct Biol 16:584
121. Geyer H, Geyer R (2006) Biochim Biophys Acta 1764:1853

122. Budnik BA, Lee RS, Steen JAJ (2006) Biochim Biophys Acta 1764:1870
123. Bardor M, Cabrera G, Rudd PM, Dwek RA, Cremata JA, Lerouge P (2006) Curr Opin Struct Biol 16:576
124. Mechref Y, Novotny MV (2002) Chem Rev 102:321
125. Zamfir AD, Bindila L, Lion N, Allen M, Girault HH, Peter-Katalinic J (2005) Electrophoresis 26:3650
126. Kobata A, Endo T (1993) In: Fukuda M, Kobata A (eds) Glycobiology: A Practical Approach. Oxford University Press, Oxford, p 79
127. Patel T, Bruce J, Merry A, Bigge C, Wormald M, Jaques A, Parekh R (1993) Biochemistry 32: 679
128. Piller F, Piller V (1993) In: Fukuda M, Kobata A (eds) Glycobiology: A Practical Approach. Oxford University Press, Oxford, p 291
129. Huang Y, Mechref Y, Novotny MV (2001) Anal Chem 73:6063
130. Huang Y, Konse T, Mechref Y, Novotny MV (2002) Rapid Commun Mass Spectrom 16:1199
131. Review: O'Neill RA (1996) J Chromatogr A 720:201
132. Yoshima H, Mizuochi T, Ishii M, Kobata A (1980) Cancer Res 40:4276
133. Bigge JC, Patel TP, Bruce JA, Goulding PN, Charles SM, Parekh RB (1995) Anal Biochem 230:229
134. Jackson P (1996) Mol Biotechnol 5:101
135. Honda S, Makino A, Suzuki S, Kakehi K (1990) Anal Biochem 191:228
136. Jackson P (1991) Anal Biochem 196:238
137. Guttman A, Chen F-TA, Evangelista A, Cooke N (1996) Anal Biochem 233:234
138. Chiesa C, O'Neill RA (1994) Electrophoresis 15:1132
139. Guttman A, Chen FT, Evangelista RA (1996) Electrophoresis 17:412
140. Shinohara Y, Furukawa J-I, Niikura K, Miura N, Nishimura S-I (2004) Anal Chem 76:6989
141. von der Lieth C-W, Lütteke T, Frank M (2006) Biochim Biophys Acta 1760:568
142. Edge CJ, Rademacher TW, Wormald MR, Parekh RB, Butters TD, Wing DR, Dwek RA (1992) Proc Natl Acad Sci USA 89:6338
143. Rudd PM, Guile GR, Kuster B, Harvey DJ, Opdenakker G, Dwek RA (1997) Nature 388:205
144. Rudd PM, Morgan BP, Wormald MR, Harvey DJ, van den Berg CW, Davis SJ, Ferguson MAJ, Dwek RA (1997) J Biol Chem 272:7229
145. Ito S, Yamashita K, Spiro RG, Kobata A (1977) J Biochem 81:1621
146. Watkins WM (1995) Molecular basis of antigenic specificity in the ABO, H and Lewis blood-group systems. In: Montreuil J, Vliegenthart JFG, Schachter H (eds) Glycoproteins. Elsevier, Amsterdam, p 313
147. Simanek EE, McGarvey GJ, Jablonowski JA, Wong C-H (1998) Chem Rev 98:833
148. Kannagi R (2002) Curr Opin Struct Biol 12:599
149. Hang HC, Bertozzi CR (2005) Bioorg Med Chem 13:5021
150. Hanisch F-G (2001) Biol Chem 382:143
151. Ten Hagen KG, Fritz TA, Tabak LA (2003) Glycobiology 13:1R
152. Julenius K, Molgaard A, Gupta R, Brunak S (2005) Glycobiology 15:153
153. Perez-Vilar J, Hidalgo J, Velasco A (1991) J Biol Chem 266:23967
154. Roth J, Wang Y, Eckhardt AE, Hill RL (1994) Proc Natl Acad Sci USA 91:8935
155. Roth J (1995) Compartmentation of glycoprotein biosynthesis. In: Montreuil J, Vliegenthart JFG, Schachter H (eds) Glycoproteins. Elsevier, Amsterdam, p 287
156. Yamashita Y, Chung YS, Horie R, Kannagi R, Sowa M (1995) J Natl Cancer Inst 87:441
157. Brockhausen I, Schutzbach J, Kuhns W (1998) Acta Anat 161:36
158. Brockhausen I, Yang J, Lehotay M, Ogata S, Itzkowitz S (2001) Biol Chem 382:219
159. Brockhausen I (1995) Biosynthesis of O-glycans of the N-acetylgalactosamine-α-Ser/Thr linkage type. In: Montreuil J, Vliegenthart JFG, Schachter H (eds) Glycoproteins. Elsevier, Amsterdam, p 201
160. Brockhausen I, Schachter H (1997) Glycosyltransferases Involved in *N*- and *O*-Glycan Biosynthesis. In: Gabius H-J, Gabius S (eds) Glycosciences: Status and Perspectives. Chapman & Hall, Weinheim, p 79
161. Kurosaka A, Nakajima H, Funakoshi J, Matsuyama M, Nagayo T, Yamashina I (1983) J Biol Chem 258:11594
162. Capon C, Leroy Y, Wieruszeski J-M, Ricart G, Strecker G, Montreuil J, Fournet B (1989) Eur J Biochem 182:139
163. Hounsell EF, Lawson AM, Feeney J, Gooi HC, Pickering NJ, Stoll MS, Lui SC, Feizi T (1985) Eur J Biochem 148:367
164. Chai W, Hounsell EF, Cashmore GC, Rosankiewicz JR, Bauer CJ, Feeney J, Feizi T, Lawson AM (1992) Eur J Biochem 203:257

165. Wang Y, Lee GF, Kelley RF, Spellman MW (1996) Glycobiology 6:837
166. Nishimura H, Takao T, Hase S, Shimonishi Y, Iwanaga S (1992) J Biol Chem 267:17520
167. Harris RJ, van Halbeek H, Glushka J, Basa LJ, Ling VT, Smith KJ, Spellman MW (1993) Biochemistry 32:6539
168. Minamida S, Aoki K, Natsuka S, Omichi K, Fukase K, Kusumoto S, Hase S (1996) J Biochem 120:1002
169. Kreppel LK, Blomberg MA, Hart GW (1997) J Biol Chem 272:9308
170. Lubas WA, Frank DW, Krause M, Hanover J (1997) J Biol Chem 272:9316
171. Dong DL, Hart GW (1994) J Biol Chem 269:19321
172. Wight TN (1999) Biosynthesis of Proteoglycans. In: Pinto BM (ed) Carbohydrates and their Derivatives Including Tannins, Cellulose, and Related Lignins (Comprehensive Natural Products Chemistry), vol 3. Elsevier, Amsterdam, p 161
173. Silbert JE, Bernfield M, Kokenyesi R (1997) Proteoglycans: a special class of glycoproteins. In: Montreuil J, Vliegenthart JFG, Schachter H (eds) Glycoproteins II. Elsevier, Amsterdam, p 1
174. Kresse H (1997) Proteoglycans - Structure and Functions. In: Gabius H-J, Gabius S (eds) Glycosciences: Status and Perspectives. Chapman & Hall, Weinheim, p 201
175. Mulloy B, Rider CC (2006) Biochem Soc Trans 34:409
176. Taylor KR, Gallo RL (2006) FASEB J 20:9
177. Bülow HE, Hobert O (2006) Annu Rev Cell Dev Biol 22:375
178. Wight TN, Kinsella MG, Qwarnstrom EE (1992) Curr Opin Cell Biol 4:793
179. Williams KJ, Fukui IV (1997) Curr Opin Lipidol 8:253
180. Bishop JR, Schuksz M, Esko JD (2007) Nature 446:1030
181. Superti-Furga A, Haestbacka J, Rossi A, Van der Harten JJ, Wilcox WR, Cohn DH, Rimoin DL, Steinmann B, Lander ES, Gitzelmann R (1996) Ann N Y Acad Sci 785:195
182. Wight TN (1996) In: Fuster V, Ross R, Topol EJ (eds) Atherosclerosis and Coronary Artery Disease. Raven Publishers, Philadelphia, PA, p 421
183. Iozzo RV (1988) Cancer Metastasis Rev 7:39
184. Snow AD, Wight TN (1989) Neurobiol Aging 10:481
185. It should be kept in mind that the anomeric configuration of the uronic acid remains the same. It is only the terminology that has changed due to the conversion of a D-sugar (GlcA) into an L-sugar (IdoA)
186. Bugg TDH (1999) Bacterial Peptidoglycan. Biosynthesis and its Inhibition. In: Pinto BM (ed) Carbohydrates and Their Derivatives Including Tannins, Cellulose, and Related Lignins (Comprehensive Natural Products Chemistry, vol 3). Elsevier, Amsterdam, p 241
187. Labischinski H, Maidhof N (1994) Bacterial peptidoglycan: overview and evolving concepts. In: Ghuysen J-M, Hakenbeck R (eds) Bacterial Cell Wall (New Compr Biochem, vol 27). Elsevier, Amsterdam, p 23
188. Vollmer W, Hoeltje J-V (2004) J Bacteriol 186:5978
189. Ghuysen JM (1968) Bacteriol Rev 32:425
190. Dmitriev B, Toukach F, Ehlers S (2005) Trends Microbiol 13:569
191. Tipper DJ, Strominger JL, Ensign JC (1967) Biochemistry 6:906
192. Leyh-Bouille M, Ghuysen J-M, Tipper DJ, Strominger JL (1966) Biochemistry 5:3079
193. Schleifer KH, Kandler O (1972) Bacteriol Rev 36:407
194. Williams DH, Bardsley B (1999) Angew Chem, Int Ed 38:1173
195. Walsh C (2000) Nature 406:775
196. Wright GD (2003) Curr Opin Chem Biol 7: 563
197. Young KD (2006) Trends Microbiol 14:155
198. Wolter F, Schoof S, Süssmuth RD (2007) Top Curr Chem 267:143
199. McCormick MH, Stark WM, Pittenger GE, Pittenger RC, McGuire JM (1956) Antibiot Annu 1955–1956:606
200. Lechevalier MP, Prauser H, Labeda DP, Ruan J-S (1986) Int J Syst Bacteriol 36:29
201. Philip JE, Schenck JP, Hargie MP (1957) Antibiot Annu 1955–1956:699
202. Nagarajan R (1994) Glycopeptide Antibiotics. Marcel Dekker, New York
203. Williams DH (1996) Nat Prod Rep 13:469
204. Nicolaou KC, Boddy CNC, Bräse S, Winssinger N (1999) Angew Chem Int Ed 38:2096
205. Howard MA, Firkin BG (1971) Thromb Diath Haemorrh 26:362
206. Hiramatsu K (1998) Drug Resist Updates 1:135
207. Levy SB (1998) Sci Am 3:46
208. Marshall FJ (1965) J Med Chem 8:18
209. Williams DH, Kalman JR (1977) J Am Chem Soc 99:2768

210. Sheldrick GM, Jones PG, Kennard O, Williams DH, Smith GA (1978) Nature 271:223
211. Harris CM, Harris TM (1982) J Am Chem Soc 104:4293
212. Lancini GC (1989) Fermentation and biosynthesis of glycopeptide antibiotics. In: Bushell ME, Graefe U (eds) Bioactive Metabolites from Microorganisms (Prog Ind Microbiol, vol 27). Elsevier, Amsterdam, p 283
213. Nicolaou KC, Mitchell HJ, Jain NF, Winssinger N, Hughes R, Bando T (1999) Angew Chem Int Ed 38:240
214. Nicolaou KC, Li H, Boddy CNC, Ramanjulu JM, Yue T-Y, Natarajan S, Chu X-J, Bräse S, Rübsam F (1999) Chem Eur J 5:2584
215. Nicolaou KC, Boddy CNC, Li H, Koumbis AE, Hughes R, Natarajan S, Jain NF, Ramanjulu JM, Bräse S, Solomon ME (1999) Chem Eur J 5:2602
216. Nicolaou KC, Koumbis AE, Takayanagi M, Natarajan S, Jain NF, Bando T, Li H, Hughes R (1999) Chem Eur J 5:2622
217. Nicolaou KC, Mitchell HJ, Jain NF, Bando T, Hughes R, Winssinger N, Natarajan S, Koumbis AE (1999) Chem Eur J 5:2648
218. Evans DA, Wood MR, Trotter BW, Richardson TI, Barrow JC, Katz JL (1998) Angew Chem Int Ed 37:2700
219. Boger DL, Miyazaki S, Kim SH, Wu JH, Loiseleur O, Castle SL (1999) J Am Chem Soc 121:3226
220. Boger DL, Miyazaki S, Kim SH, Wu JH, Castle SL, Loiseleur O, Jin Q (1999) J Am Chem Soc 121:10004