Glycobiology vol. 12 no. 4 pp. 43R–56R, 2002

MINI REVIEW

Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds

Robert G. Spiro

Departments of Biological Chemistry and Medicine, Harvard Medical School, and the Joslin Diabetes Center, One Joslin Place, Boston, MA 02215, USA

Accepted on January 30, 2002

Formation of the sugar–amino acid linkage is a crucial event in the biosynthesis of the carbohydrate units of glycoproteins. It sets into motion a complex series of posttranslational enzymatic steps that lead to the formation of a host of protein-bound oligosaccharides with diverse biological functions. These reactions occur throughout the entire phylogenetic spectrum, ranging from archaea and eubacteria to eukaryotes. It is the aim of this review to describe the glycopeptide linkages that have been found to date and specify their presence on well-characterized glycoproteins. A survey is also made of the enzymes involved in the formation of the various glycopeptide bonds as well as the site of their intracellular action and their affinity for particular peptide domains is evaluated. This examination indicates that 13 different monosaccharides and 8 amino acids are involved in glycoprotein linkages leading to a total of at least 41 bonds, if the anomeric configurations, the phosphoglycosyl linkages, as well as the GPI (glycophosphatidylinositol) phosphoethanolamine bridge are also considered. These bonds represent the products of *N*- and *O*-glycosylation, *C*-mannosylation, phosphoglycation, and glypiation. Currently at least 16 enzymes involved in their formation have been identified and in many cases cloned. Their intracellular site of action varies and includes the endoplasmic reticulum, Golgi apparatus, cytosol, and nucleus. With the exception of the Asn-linked carbohydrate and the GPI anchor, which are transferred to the polypeptide *en bloc,* the sugar–amino acid linkages are formed by the enzymatic transfer of an activated monosaccharide directly to the protein. This review also deals briefly with glycosidases, which are involved in physiologically important cleavages of glycopeptide bonds in higher organisms, and with a number of human disease states in which defects in enzymatic transfer of saccharides to protein have been implicated.

Key words: consensus sequence of glycoproteins/diseases of protein glycosylation/glycosyltransferases/*N*-glycosylation/*O*-glycosylation

Introduction

It has been appreciated for some time that the attachment of sugar residues is the most complicated co- or posttranslational modification that a protein can undergo (Spiro, 1973). Indeed, the modification of proteins through enzymatic glycosylation is an event that reaches beyond the genome and is controlled by factors that differ greatly among cell types and species. Many elaborate glycosylation routes have been identified in a host of organisms that lead to the mature carbohydrate units on glycoproteins that are secreted by cells or become components of its membranes, cytoplasm, or nucleus. The defining event in the biogenesis of peptide-linked oligosaccharides is clearly the formation of the sugar–amino acid bond; this in most instances determines the nature of the carbohydrate units that will subsequently be formed by the cellular enzymatic machinery, which in turn influences the protein's biological activity.

Since the description of the GlcNAc-β-Asn linkage in ovalbumin by Neuberger and colleagues (Johansen *et al.*, 1961), glycopeptide linkages have been described involving almost every functional group occurring on peptide chains and most of the commonly occurring monosaccharide residues, so that a multitude of diverse sugar–amino acid combinations have been described. With the recognition that eubacteria and archaea (Lechner and Wieland, 1989; Messner, 1997) produce glycoproteins in addition to eukaryotes, the glycopeptide bond has attained the broadest possible phylogenetic distribution.

It is the aim of this review to describe the great variety of glycopeptide linkages that have been reported to date as well as to indicate their distribution among well-defined glycoproteins. The large number of enzymes involved in the formation of sugar–protein bonds that have presently been characterized from various sources will also be surveyed and their affinity for certain domains of peptide chains evaluated. This information can be of value for the production of recombinant glycoproteins. Glycosidases that have been implicated in physiologically relevant scission of the sugar–amino acid linkage will be briefly described, as will human diseases in which alterations in the attachment of carbohydrate to protein have been observed.

Nature and distribution of glycopeptide linkages of glycoproteins

At the present stage of our knowledge an impressive variety of carbohydrate–peptide linkages have been described that are distributed among glycoproteins found in essentially all living organisms, ranging from eubacteria to eukaryotes. In the latter group they are distributed over a broad phylogenetic spectrum reaching from unicellular organisms, such as yeast and trypanosomes, to the highly differentiated tissues of the animal and plant kingdoms (Table I). Thirteen different monosaccharides and 8 amino acid types participate in these bonds so that at

Table I. Nature and distribution of sugar–amino acid linkages of glycoproteins[a]

	Linkage			Phylogenetic distribution			
Type of bond	Amino acid	Sugar	Configuration[b]	Eukaryotes	Archaea[c]	Bacteria[c]	Examples[d]
N-glycosyl	Asn	GlcNAc	β	+	+	+	Ovalbumin, fetuin, insulin receptor
	Asn	Glc	β	+	+	–	Laminin, *H. halobium* S-layer
	Asn	GalNAc	*	–	+	–	*H. halobium* S-layer
	Asn	Rha	*	–	–	+	*S. sanguis* cell wall
	Arg	Glc	β	+	–	–	Sweet corn amylogenin
O-glycosyl	Ser/Thr[e]	GalNAc	α	+	–	–	Mucins, fetuin, glycophorin
	Ser/Thr	GalNAc	β	–	+	–	*A. thermoaerophilus* S-layer
	Ser/Thr	GlcNAc	β	+	–	–	Nuclear and cytoplasmic proteins
	Ser/Thr	Gal	α	+	–	+	Earthworm collagen, *B. cellulosoleum*
	Ser/Thr	Man	α	+	–	–	Yeast mannoproteins
	Ser/Thr	Man	*	+	–	+	α-dystroglycan, *F. meningosepticum*
	Ser/Thr	Fuc	α	+	–	–	Coagulation and fibrinolytic factors
	Ser/Thr	Pse[f]	α	–	–	+	*C. jejuni* flagellins
	Ser/Thr	DiActrideoxyhexose[g]	*	–	–	+	*N. meningitidis* pili
	Ser	Glc	β	+	–	–	Coagulation factors
	Ser	FucNAc	β	–	–	+	*P. aeruginosa* pili
	Ser	Xyl	β	+	–	–	Proteoglycans
	Ser	Gal	α	+	–	–	Cell walls of plants
	Thr	Man	α	–	–	+	*M. tuberculosis* secreted glycoproteins
	Thr	Man	*	+	–	–	Clamworm collagen
	Thr	GlcNAc	α	+	–	–	Dictyostelium[h], *T. cruzi*
	Thr	GlcNAc	*	+	–	–	Rho proteins (GTPases)
	Thr	Glc	*	+	–	–	Rho proteins (GTPases)
	Thr	Gal	*	+	+	–	*H. halobium* S-layer, vent worm collagen
	Hyl[i]	Gal	β	+	–	–	Collagen, C1q complement, core specific lectin
	Hyp[i]	Ara[j]	α	+	–	–	Plant cell walls
	Hyp	Ara	β	+	–	–	Potato lectin
	Hyp	Gal	β	+	–	–	Wheat endosperm
	Hyp	GlcNAc	*	+	–	–	*Dictyostelium* cytoplasmic proteins
	Tyr	Glc	α	+	–	–	Muscle and liver glycogenin
	Tyr	Glc	β	–	–	+	*C. thermohydrosulfuricum* S-layer
	Tyr	Gal	β	–	–	+	*T. thermohydrosulfuricus* S-layer
C-mannosylation	Trp[k]	Man	α	+	–	–	RNase 2, interleukin 12, properdin
Phosphoglycosyl	Ser	GlcNAc	α-1-P	+	–	–	*Dictyostelium* proteinases
	Ser	Man	α-1-P	+	–	–	*L. mexicana* acid phosphatase
	Ser	Fuc	β-1-P	+	–	–	*Dictyostelium* proteins
	Ser	Xyl	*-1-P	+	–	–	*T. cruzi* cell surface
Glypiation	Pr-C-(O)-EthN-6-P-Man[l]			+	+	–	*T. brucei* VSG, Thy-1, *Sulfolobus* acidocaldarius proteins

[a]References are given in the text.
[b]Refers to anomeric configuration of glycopeptide or glycosylphosphate bonds; *asterisks* indicate that configuration has not yet been established.
[c]Also known as archaebacteria and eubacteria, respectively.
[d]Further examples and details are presented in the text.
[e]Ser/Thr indicates that linkages to both amino acids have been found in a given protein while Ser or Thr by itself indicates that at the present time a bond to only one of these two amino acids has been observed in the proteins examined.
[f] Pse refers to pseudaminic acid (5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-L-manno-nonulosinic acid).
[g]DiActrideoxyhexose refers to 2,4-diacetamido-2,4,6-trideoxyhexose.
[h]In this table *Dictyostelium* always refers to the *discoideum* species.
[i]The abbreviations Hyl and Hyp refer to hydroxylysine and hydroxyproline, respectively.
[j]The Ara linkages have been reported to be in the furanosidic form.
[k]The mannose is linked to C-2 of the indole ring of tryptophan.
[l]The mannose is attached to the C-terminal end of the protein by a phosphoethanolamine bridge.

least 31 sugar–amino acid combinations exist. If the known anomeric configurations of the glycosidic bonds are taken into account this number rises to a minimum of 37. With the additional consideration of the phosphoglycosyl linkages and the glycophosphatidylinositol (GPI) phosphoethanolamine bridge, a total of at least 41 linkages are found to occur (Table I).

The glycopeptide bonds can be arranged in five quite distinct groups, as shown in Figure 1. In many cases, more than one type of sugar–amino acid bond can occur in the same protein, depending on the available enzymatic machinery as well as the amino acid sequence and conformation

N-*glycosidic bonds*

The β-glycosylamine linkage of GlcNAc to Asn represents the most widely distributed carbohydrate–peptide bond and is the site of attachment for a large variety of complex and poly-mannose oligosaccharides (Spiro, 1973; Montreuil, 1980) in proteins with demonstrated biological importance (Varki, 1993). The *GlcNAc*-β-*Asn* bond was initially described in ovalbumin (Johansen *et al.*, 1961) and because this protein contains only a modest amount of carbohydrate (~3%), the characterization of this linkage was considered to be a major accomplishment. Soon thereafter the *GlcNAc*-β-*Asn* bond was observed in a vast array of proteins in eukaryotes, including plasma proteins, thyroglobulins, hormones, enzymes, cell surface receptors, immunoglobulins, and lectins (Spiro, 1973; Montreuil, 1980). Indeed, the *N*-linked carbohydrate units are frequently found together with *O*-linked oligosaccharides on proteins such as fetuin (Spiro and Bhoyroo, 1974), glycophorin (Marchesi *et al.*, 1976), IgG immunoglobulins (Fanger and Smyth, 1972), yeast mannoproteins (Herscovics and Orlean, 1993), insulin receptor (Collier and Gorden, 1991), thyroid cell surface glycoproteins (Edge and Spiro, 1997), and even molecules that are generally considered to be primarily carriers of carbohydrate units attached by *O*-glycosidic bonds, such as mucins (Perez-Vilar *et al.*, 1996), proteoglycans (Lohmander *et al.*, 1980; Parthasarathy and Spiro, 1984), and collagens (Nayak and Spiro, 1991). Though the *GlcNAc-Asn* linkage prevails in eukaryotic cells, it has also been observed in archaea and eubacteria, as in *Thermoplasma acidophilum* (Yang and Haug, 1979) and *Streptococcus sanguis* (Erickson and Herzberg, 1993), respectively. Moreover, it has been known for some time that the surface layer (S-layer) of the achaebacterium *Halobacter halobium* (also known as *H. salinarium)* contains carbohydrate units linked to Asn through Glc and GalNAc residues (Lechner and Wieland, 1989; Messner, 1997). A mammalian *Glc*-β-*Asn* linkage has so far been noted only in mammalian laminin (Schreiner *et al.*, 1994), and a *Rha-Asn* bond has uniquely been reported to occur in the cell wall of *S. sanguis* (Erickson and Herzberg, 1993). The recognition of the *GlcNAc*-β- *Asn* bond has been greatly facilitated by the ability of bacterial peptide *N*-glycosidase to release carbohydrate units attached to protein by this linkage (Maley *et al.*, 1989).

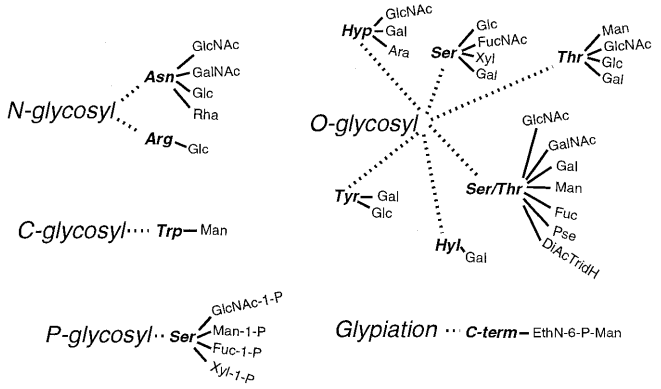

Fig. 1. Diagrammatic representation of the five distinct types of sugar–peptide bonds that have currently been identified. The anomeric configuration, physiologic distribution, and glycoprotein examples of the various linkages are presented in Table I. The abbreviations used are the same as in Table I. Furthermore, *DiAcTridH* refers to 2,4,-diacetamido, 2,4,6-trideoxyhexose, and *C-term* refers to the carboxy-terminal amino acid residue.

The report that amylogenin, which is believed to be a self-glycosylating protein from sweet corn, contains a *Glc*-β-*Arg* linkage to the guanidino group of the Arg provides another example of an *N*-glycosyl bond (Singh *et al.*, 1995).

O-*glycosidic bonds*

Linkages in which the sugar is attached to an amino acid containing a hydroxyl group occur in great variety of proteins, not only in regard to the partners in this linkage but also in different anomeric configurations (Table I). Every amino acid with a hydroxyl functional group (i.e., Ser, Thr, Tyr, Hyp [hydroxyproline], and Hyl [hydroxylysine]) has been implicated.

The *GalNAc*-α-*Ser/Thr* linkage has been considered a hallmark of mucins where it occurs in clusters. However, a wide variety of glycoproteins contain this linkage (Spiro, 1973; Sadler, 1984), such as fetuin, human gonadotropins, glycophorin, and antifreeze glycoproteins, which indicates that such *O*-linked oligosaccharides frequently also occur in other highly diverse molecules. Though at present it would appear that this linkage is limited to eukaryotes, its β-anomer (*GalNAc*-β-*Ser/Thr*) has been reported to occur in the S-layer of the archaebacterium *Aneurinibacillus thermoaerophilus* (Schäffer *et al.*, 1999).

GlcNAc-β-*Ser/ Thr* represents an increasingly important linkage that is widely dispersed among eukaryotes, from protozoa to higher mammals. It is distinctive in that it is found in nuclear and cytoskeletal proteins and indeed represents the first reported example of glycosylated proteins found outside of the secretory channels (Hart, 1997). In contrast to most other peptide-linked monosaccharides, the β-linked *GlcNAc-Ser/Thr* does not become further substituted by other sugars, remaining a simple monosaccharide modification of the protein to which it is attached. Indeed this property permitted the use of radio-labeled UDP-Gal with purified GlcNAc-galactosyltransferase to demonstrate its presence in minute amounts (Holt *et al.*, 1987) and this continues to be used as an effective probe. Though the *GlcNAc*-β-*Ser/Thr* bond appears to be confined to intracellular glycoproteins, the α-linkage of GlcNAc to Thr has been found in cell surface and secreted glycoproteins from *Trypanosoma cruzi* (Previato *et al.*, 1998) and *Dictyostelium discoideum* (Jung *et al.*, 1998).

Gal-α-*Ser/Thr* has been reported to be present in the cuticle collagens of the earthworm, *Lumbricus terrestris* (Muir and Lee, 1970) and clamworm, *Nereis virens* (Spiro and Bhoyroo, 1980), where it appears both unsubstituted and as di- and tri-α-linked galactose oligosaccharides. Because these collagens do not contain Hyl, the possibility of linkage to this amino acid, as occurs in vertebrate collagens, is excluded. Gal linked apparently only to Thr has also been found in vent worm cuticle collagen (Mann *et al.*, 1996). The *Gal*-α-*Ser/Thr* bond has also been

observed in eubacteria, where it is present on the cellulosomes of *Bacteroides cellulosolvens* and *Clostridium thermocellum* (Gerwig *et al.*, 1993). In archaea the S-layer glycoproteins of *H. halobium* contain clusters of glucosylgalactose disaccharides linked solely to Thr by a bond with an as yet undetermined anomeric configuration (Mescher and Strominger, 1976). Furthermore, it has been reported that single Gal residues α-linked to Ser are present in the cell wall of *Phaseolus coccineus* (O'Neill and Selvendran, 1980) and other higher plants (Lamport *et al.*, 1973).

The well-studied mannoproteins of the yeast cell wall are known to contain the *Man*-α-*Ser/Thr* glycopeptide linkage (Herscovics and Orlean, 1993). A *Man-Ser/Thr* carbohydrate–peptide bond with an as yet unknown anomeric configuration has also been identified in the α-dystroglycans of peripheral nerve (Endo, 1999) and in brain proteoglycans and glycoproteins (Finne *et al.*, 1979; Yuen *et al.*, 1997), as well as several proteins secreted by *Flavobacterium meningosepticum* (Plummer *et al.*, 1995). Furthermore, oligosaccharides α-linked to Thr by a Man residue have been found in the secreted 45-kDa glycoproteins of *Mycobacterium tuberculosis* (Dobos *et al.*, 1996). A unique GlcUAα1-6Man disaccharide linked to Thr with an as yet unspecified anomeric configuration has been found in the cuticle collagen of *Nereis* (Spiro and Bhoyroo, 1980).

The *Fuc*-α-*Ser/Thr* and *Glc*-β-*Ser* linkages can be considered together (Table I) because they appear to be primarily found in epidermal growth factor (EGF) domains (Harris and Spellman, 1993) of mutimodular proteins such as coagulation and fibrinolytic factors. Proteins containing *Fuc*-α-*Ser/Thr* oligosaccharides include urokinase (Buko *et al.*, 1991), human coagulation factors VII (Bjoern *et al.*, 1991), IX (Nishimura *et al.*, 1992a), and XII (Harris *et al.*, 1992) and Notch1 (Moloney *et al.*, 2000). Fucose appears in the mature glycoproteins either alone or as the inner component of short oligosaccharides. The site of glucose attachment has so far been found to be limited to Ser, and in a number of instances this sugar residue is the attachment point of one or two Xyl residues (Harris and Spellman, 1993). Coagulation factors VII and IX as well as human plasma protein Z (Nishimura *et al.*, 1989) have been shown to contain the *Glc*-β-*Ser* bond, as has bovine thrombospondin (Nishimura *et al.*, 1992b).

The *Pse*-α-*Ser/Thr* and *DiActrideoxyhexose-Ser/Thr* linkages (see Table I for abbreviations) are unusual bonds that have recently been found in eubacteria. Pseudaminic acid, which occurs as multiple substituents in *Campylobacter jejuni* flagellin (Thibault *et al.*, 2001), is of particular interest because it represents the first report of an acidic monosaccharide directly linked to protein. The DiActrideoxyhexose-containing glycopeptide bond has been identified in the pili of *Neisseria meningitides*, where it is the linkage sugar of a digalactose-containing trisaccharide unit (Stimson *et al.*, 1995). *FucNAc*-β-*Ser/Thr* represents another recently described eubacterial linkage which occurs in the pili of *Pseudomonas aeruginosa* 1244 (Castric *et al.*, 2001); the FucNAc attaches a trisaccharide containing xylose and a derivative of pseudaminic acid to the protein.

It has been known for some time that the attachment of the chondroitin sulfate and heparan sulfate glycosaminoglycan chains of mammalian proteoglycans is mediated by a *Xyl*-β-*Ser* bond (Kjellén and Lindahl, 1991; Esko and Zhang, 1996). This stands in contrast to corneal and skeletal keratan sulfate chains, which are linked to the protein by *GlcNAc*-β-*Asn* and *GalNAc*-α-*Ser/Thr* linkages, respectively (Kjellén and Lindahl, 1991).

Glc-Thr and *GlcNAc-Thr* are unusual glycopeptide bonds in that they are pathologically generated by the action of *Clostridium* toxins (*vide infra*) on the Rho family of low-molecular-mass GTPases (Busch and Aktories, 2000).

All vertebrate and invertebrate collagens including those of basement membranes, with the exception of the previously mentioned worm cuticle collagens, manifest the *Gal*-β-*Hyl* glycopeptide linkages (Spiro, 1969, 1972a) along their peptide chains. The Hyl-linked saccharide can occur as unsubstituted Gal residues or in the form of Glcα1-2Gal disaccharides (Spiro, 1967). Moreover, the collagenous regions of C1q complement (Shinkai and Yonemasu, 1979) and the hepatic core specific lectin (Colley and Baenziger, 1987) have been shown to contain Gal-β-Hyl bonds. In contrast to the *O*-glycosidic linkages to Ser and Thr, which can be split by β-elimination, the Gal–Hyl bond is stable to even strong alkali treatment (Spiro, 1972b).

Gal and Ara saccharides linked to Hyp are features of plant glycoproteins. The *Gal*-β-*Hyp* glycopeptide bond has been found in wheat endosperm (Strahm *et al.*, 1981), gum arabic from *Acacia senegal* (Qi *et al.*, 1991), and the cell wall of *Chlamydomonas* green algae (Miller *et al.*, 1972). Plant cell walls ranging the phylogenetic spectrum from land plants to green algae have been shown to contain the *Ara*-α-*Hyp* linkage (Yamagishi *et al.*, 1976), which is a the primary glycopeptide bond of arabinogalactans (Kieliszewski *et al.*, 1995); on the other hand the *Ara*-β-*Hyp* combination has been reported to occur in potato lectin (Allen *et al.*, 1978).

Recently a *GlcNAc-Hyp* bond has been characterized in cytoplasmic glycoprotein of *Dictyostelium* (Teng-umnuay *et al.*, 1998). More specifically, this linkage was found to attach a pentasaccharide on the Skp1 component of the Skp1-cullin-F-box-protein complex (SCF), which is involved in the ubiquitination of various cell and other regulatory proteins.

Glycogenin, the protein primer for glycogen synthesis, has been shown to have the most internal sugar linked to protein by a *Glc*-α-*Tyr* bond in both muscle and liver (Smythe and Cohen, 1991). On the other hand, Glc linked to Tyr by a β-glycosidic linkage (*Glc*-β-*Tyr*) has been found in the S-layer of eubacteria including *Clostridium thermohydrosulfuricum* (Messner *et al.*, 1992) and *Thermoanaerobacterium thermosaccharolyticum D120-70* (Schäffer *et al.*, 2000). In another variant of the latter species (L111-69) a *Gal*-β-*Tyr* linkage has been identified (Bock *et al.*, 1994).

C-*mannosyl bonds*

An entirely novel carbohydrate–protein linkage involving the attachment of an α-mannosyl residue to C-2 of the Trp through a C-C bond, was described recently (de Beer *et al.*, 1995). Unlike the *N*- and *O*-glycosyl linkages, this glycopeptide bond does not involve an amino acid functional group. This linkage has been so far found in mammalian proteins including RNase2 (same as RNase Us) (de Beer *et al.*, 1995), interleukin-12 (Doucey *et al.*, 1999), and properdin (Hartmann and Hofsteenge, 2000).

Phosphoglycosyl bonds

Attachment of sugar to protein via a phosphodiester bond represents another quite distinct type of glycopeptide linkage

(Haynes, 1998) in which GlcNAc, Man, Xyl, and Fuc have been found to be involved (Table I). The *GlcNAc-α-1-P-Ser* linkage has been found in various proteins from *Dictyostelium* including proteinase-1 (Mehta *et al.*, 1996). *Man-α-1-P-Ser* has been observed in several major proteins of *Leishmania* species (Guha-Niyogi *et al.*, 2001), and *Xyl-1-P-Ser* has been found in *T. cruzi* (Haynes, 1998). Furthermore, evidence for the presence of *Fuc-β-1-P-Ser* in *Dictyostelium* has also been obtained (Srikrishna *et al.*, 1998).

Glypiated linkage

A major carbohydrate–protein connection is the GPI anchor. In this bond Man is linked to phosphoethanolamine, which in turn is attached to the terminal carboxyl group of the protein. This linkage is widely distributed among biologically important cell surface glycoproteins of eukaryotes, including the variant surface glycoproteins (VSGs) of trypanosomes and the Thy-1 antigen (Ferguson, 1999). Recently GPI-linked proteins have also been detected in the archaebacterium, *Sulfolobus acidocaldarius* (Kobayashi *et al.*, 1997).

Biosynthesis of glycopeptide linkages

At present the enzymes involved in the formation of at least 16 glycopeptide bonds have been identified and purified to various extents; moreover, a substantial number of these transferases have been cloned and in most instances the subcellular site of their action has been determined (Table II). With the exception of the *GlcNAc-β-Asn* bond and the GPI anchor, the sugar–amino acid linkage is formed by the direct enzymatic transfer of an activated monosaccharide or monosaccharide-1-phosphate to a specific amino acid residue in the polypeptide chain (Table II). Oligosaccharides are then generated by the sequential enzymatic attachment of sugars to the peptide-linked component. In the case of the Asn bond and the GPI anchor, a preassembled carbohydrate unit is added to the protein in the endoplasmic reticulum (ER), although by quite different mechanisms (*vide infra*). A strict amino acid consensus sequence has so far been established only for the *GlcNAc-β-Asn* linkage, but distinct glycosylation motifs have been observed for a number of other glycopeptide bonds (Table III).

N-*glycosylation*

It has been known for some time that the *GlcNAc-β-Asn* bond is established in eukaryotes through the cotranslational transfer of a preassembled dolichol-linked triglucosylated polymannose oligosaccharide which subsequently undergoes varying degrees of processing to the large array of *N*-linked carbohydrate units (Cummings, 1992; Moremen *et al.*, 1994; Burda and

Table II. Enzymes involved in the synthesis of glycopeptide bonds[a]

Linkage[b]	Enzyme[c]	Glycosyl donor	Location	Source[d]	Cloned[e]
GlcNAc-β-Asn	Oligo[f]-tr	Dol-PP-Oligo	ER	Liver, pancreas oviduct, yeast	Yes (multiple subunits)
GalNAc-α-Ser/Thr	GalNAc-tr	UDP-GalNAc	Golgi	Colostrum, submaxillary gland	Yes (multiple enzymes)
GlcNAc-α-Thr	GlcNAc-tr	UDP-GlcNAc	Golgi	Trypanosomes, *Dictyostelium*	No
GlcNAc-β-Ser/Thr	GlcNAc-tr	UDP-GlcNAc	Cytosol, nucleus	Liver, blood	Yes
Man-α-Ser/Thr	Man-tr	Dol-P-Man	ER	Yeast	Yes (multiple enzymes)
Fuc-α-Ser/Thr	Fuc-tr	GDP-Fuc	Golgi	CHO cells, liver	Yes
Xyl-β-Ser	Xyl-tr	UDP-Xyl	ER, Golgi	Cartilage, choriocarcinoma	Yes
Glc-Thr	Clostridial cytotoxin	UDP-Glc	Cytosol	*C. difficile* and *sordelli*	Yes
GlcNAc-Thr	Clostridial cytotoxin	UDP-GlcNAc	Cytosol	*C. novyi*	Yes
Gal-β-Hyl	Gal-tr	UDP-Gal	Golgi	Kidney, cartilage	No
GlcNAc-Hyp	GlcNAc-tr	UDP-GlcNAc	Cytosol	*Dictyostelium*	Yes
Glc-α-Tyr	Glycogenin[g]	UDP-Glc	Cytosol	Liver, muscle	Yes
GlcNAc-α-1-P-Ser	GlcNAc-1-P- tr	UDP-GlcNAc	Golgi	*Dictyostelium*	No
Man-α-1-P-Ser	Man-1-P-tr	GDP-Man	Golgi	*Leishmania mexicana*	No
Man-α-Trp	Man-tr	Dol-P-Man	ER	Rat liver	No
Pr-C(O)EthN-6-P-Man	Transamidase	GPI[f]	ER	Yeast	Yes (multiple subunits)

[a]References are given in the text.
[b]See Table I for nature of the bonds.
[c]*tr* refers to saccharide:polypeptide transferase.
[d]Tissues or cells from which enzyme has been examined after varying degrees of purification.
[e]See text for details.
[f]Abbreviations are Oligo, $Glc_3Man_9GlcNAc_2$; GPI, glycosylphosphatidylinositol.
[g]Autoglucosylation.

Table III. Amino acid consensus sequences or glycosylation motifs for the formation of glycopeptide bonds[a]

Glycopeptide bond[b]	Consensus sequence or peptide domain[c]
GlcNAc-β-Asn	*Asn*-X-Ser/Thr (X = any amino acid except Pro)
Glc-β-Asn	*Asn*-X-Ser/Thr
GalNAc-α-Ser/Thr	Repeat domains rich in Ser, Thr, Pro, Gly, Ala in no special sequence
GlcNAc-α-Thr	Thr rich domain near Pro residues
GlcNAc-β-Ser/Thr	Ser/Thr rich domains near Pro, Val, Ala, Gly
Man-α-Ser/Thr	Ser/Thr rich domains
Fuc-α-Ser/Thr	EGF modules (Cys-X-X-Gly-Gly-*Thr*/*Ser*-Cys)
Glc-β-Ser	EGF modules (Cys-X-*Ser*-X-Pro-Cys)
Xyl-β-Ser	*Ser*-Gly (Ala) (in the vicinity of one or more acidic residues)
Glc/GlcNAc-Thr	Rho: *Thr*-37[d]; Ras, Rac and Cdc42: *Thr*-35[d]
Gal-Thr	Gly-X-*Thr* (X = Ala, Arg, Pro, Hyp, Ser) (vent worm)[e]
Gal-β-Hyl	Collagen repeats (X-*Hyl*-Gly)
Ara-α-Hyp	Repetitive Hyp rich domains (e.g., Lys-Pro-*Hyp*-Hyp-Val)
GlcNAc-Hyp	Skp1: *Hyp*-143[d]
Glc-α-Tyr	Glycogenin: Tyr-194[d]
GlcNAc-α-1-P-Ser	Ser rich domains (e.g., Ala-*Ser*-Ser-Ala)
Man-α-1-P-Ser	Ser rich repeat domains
Man-α-Trp[f]	*Trp*-X-X-Trp
Man-6-P-EthN-C(O)-Pr	GPI[g] attached after cleavage of C-terminal peptide

[a]References are given in the text.
[b]Only the glycopeptide linkages for which some information is available are listed; where known the anomeric configuration is indicated.
[c]The information given is based on structural, site-directed mutagenesis and/or *in vitro* biosynthetic studies. Except for the *Asn-X-Ser/Thr* consensus sequence, the amino acid motifs presented are generally open to reservations, which are discussed in the text. An amino acid in italics indicates that it is the site of the sugar attachment; X indicates that the amino acid can be of a variable nature unless otherwise stated.
[d]Where glycopeptide bonds appear to be limited to specific proteins with established amino acid sequence, the specific residue involved in the glycopeptide bond is indicated by its number.
[e]Indicates the biological source of the sequence; the sources of other sequences are given in Table I and the text.
[f]Linkage is defined in text and footnote to Table I.
[g]GPI refers to glycophosphatidylinositol. The nature of the C-terminal peptide released by the transamidase before attachment of the GPI takes place is described in the text.

Aebi, 1999; Spiro, 2000). A rather strict consensus sequence, *Asn-X-Ser/Thr* (Table III) was postulated (Marshall, 1974), and this has been supported by numerous subsequent studies employing structural, mutagenic, and *in vitro* approaches. Although the *Asn-X-Ser/Thr* sequence occurs frequently in proteins, it does not necessarily indicate the actual presence of an *N*-glycosidic linkage, most probably due to conformational factors (Apweiler *et al.*, 1999). *In vitro* studies have shown that replacement of Thr by Ser residues resulted in a pronounced decrease in glycosyl transfer (Bause and Legler, 1981). Moreover, it has been proposed that the Ser or Thr is required for a hydrogen-bond donor function in enzyme binding and in oligosaccharide transfer, although cysteine in its reduced form could take the place of the hydroxyamino acid. The negative effect of Pro as the X amino acid has been attributed to its interference with the ability of the peptide chain to adopt and stabilize a loop conformation (Bause, 1983). The oligosaccharyltransferase resisted purification until it was shown that it can be stabilized by the inclusion of phosphatidylcholine in the preparation and assay buffers (Chalifour and Spiro, 1988). Presently the oligosaccharyltransferase has been isolated from a number of eukaryotic cells and shown to be a heterooligomeric ER membrane complex (Silberstein and Gilmore, 1996; Yan and Lennarz, 1999). In yeast, nine different transmembrane subunits have been identified, and it has been shown that subunit OST1p recognizes the consensus sequence (Yan and Lennarz, 1999); cloning of several subunits has already been achieved (Knauer and Lehle, 1999).

In the formation of the *N*-glycosidic bonds of archaea, C55-60 dolichol monophosphate oligosaccharides have been implicated (Lechner and Wieland, 1989) and it has been suggested that in these primitive organisms *N*-glycosylation takes place on the outer surface of the cell membrane and that the *Asn-X-Ser/Thr* consensus sequence also is operative. Recently it has been reported that homologues of the highly conserved STT3 oligosaccharyltransferase subunit have been observed in archaea and also in the eubacterium *Campylobacter jejuni* 81-176 (Wacker *et al.*, 2001). However, gene replacement studies conducted on the Asn-bonds that occur in *H. halobium* (Table I) indicated that the *Glc*-β-*Asn* bond in contrast to the *GalNAc*-β-*Asn* does not have a strict requirement for the *Asn-X-Ser/Thr* consensus sequence; this led to the suggestion that distinct enzymes may be responsible for the formation of these two *N*-glycosidic bonds (Zeitler *et al.*, 1998).

O-*glycosylation*

The biosynthesis of the *GalNAc*-α-*Ser/Thr* bond has been extensively studied in eukaryotic cells, and it has become

evident through the studies of a number of investigators that a family of at least nine GalNAc-transferases exists (Clausen and Bennett, 1996; Ten Hagen *et al.*, 2001). Indeed, it has been suggested that these enzymes work in concert in a hierarchical manner to form the clustered Ser/Thr-linked oligosaccharides that frequently occur in the "mucin"-type of glycoprotein (Ten Hagen *et al.*, 2001). Several of these enzymes have been cloned and though it has become evident that they are distinct gene products and may be distributed on different chromosomes, they are generally homologous to each other (Clausen and Bennett, 1996). Although these enzymes act on characteristic peptide regions (Table III), no specific consensus sequence has been identified despite numerous intensive investigations; this may very well be due to the multiplicity of the GalNAc-transferases. Because they are frequently assayed without prior separation, overlapping but distinct substrate specificities may therefore be masked. In general however, this linkage is found in clusters of Ser/Thr residues with a β-turn near Pro and at a distance from charged amino acids. *In vitro* studies suggest that Thr is favored over Ser for α-GalNAc modification (Elhammer *et al.*, 1993). Immunoelectron microscopic studies (Roth *et al.*, 1994), in agreement with subcellular fractionation investigations (Hirschberg *et al.*, 1998), have indicated that α-GalNAc-transfer occurs in the cis-Golgi; however the multiple enzymes in this family make it possible that some act in a pre-Golgi or ER compartment, as had previously been suggested. Indeed it is not yet known if the entire GalNAc-transferase family occurs in every cell or species or if there is a selective distribution of the various enzyme isoforms.

Membrane GlcNAc-transferases that form the *GlcNAc*-α-*Ser/Thr* linkages have been characterized in *Dictyostelium* (Jung *et al.*, 1998) and *T. cruzi* (Previato *et al.*, 1998). These appear to be distinct from the enzyme that generates the *GlcNAc*-β-*Ser/Thr* bond. Indeed it was reported that the optimal peptide for the cytosolic GlcNAc-transferase responsible for the formation of the *GlcNAc*-β-*Ser/Thr* linkage is not a substrate for the latter enzyme (Previato *et al.*, 1998). It has been suggested that the *GlcNAc*-α-*Ser/Thr* linkage might have substituted for the α-GalNAc bond in more primitive eukaryotes before the epimerase that converts GlcNAc to GalNAc had evolved (Jung *et al.*, 1998). Studies on acceptor sites have indicated that the GlcNAc-transferase acts on clustered Thr residues near Pro and studies on *Dictyostelium* have indicated that these peptide sequences are similar to those reported for the addition of α-GalNAc residues in mammalian tissues (Jung *et al.*, 1998).

The GlcNAc-transferase responsible for the genesis of the *GlcNAc*-β-*Ser/Thr* linkage was the first glycopeptide-forming enzyme to be localized outside of the channels of the secretory apparatus (Table II); it is widely distributed among eukaryotes and has a highly conserved primary sequence (Hart, 1997). This enzyme has been purified from rat liver cytosol (Haltiwanger *et al.*, 1992) and rabbit blood (Lubas *et al.*, 1997) and has been cloned from rat liver (Kreppel *et al.*, 1997) as well as *C. elegans* and human liver (Lubas *et al.*, 1997). This transferase has taken on importance not only because of the biologically relevant proteins on which it acts but also from the finding that the Ser/Thr residues it glycosylates appear to be identical to those that can undergo *O*-phosphorylation. This has suggested the possibility that there is a reciprocal relationship between these two peptide modifications in a potential regulatory cycle in which cytosolic β-*N*-acetylglucosaminidase plays a key role (Comer and Hart, 2000). Although no specific amino acid consensus sequence has as yet been found, some information relating to the polypeptide domains that it favors has been obtained (Table III) (Haltiwanger *et al.*, 1997).

The biosynthesis of the *Man*-α-*Ser/Thr* linkage has been studied most extensively in yeast (Herscovics and Orlean, 1993; Strahl-Bolsinger *et al.*, 1999). It has been demonstrated that the formation of this glycopeptide bond takes place in the ER (Haselbeck and Tanner, 1983) and moreover that the mannosyltransferase uses a dolichol-linked monosaccharide rather than a sugar nucleotide as the glycosyl donor (Table II). It has become apparent in recent years that seven genes for the protein *O*-mannosyltransferase (PMT1-7) with extensive shared homology are present in *S. cerevisiae* (Strahl-Bolsinger *et al.*, 1999) of which two have been cloned (Lussier *et al.*, 1995). The mannosylation of proteins from higher eukaryotes has not yet been defined, but a human homolog of the PMT1 transferase gene has recently been reported (Jurado *et al.*, 1999). Although a consensus sequence for *O*-mannosylation has not been established, glycosylation does take place in clustered Ser/Thr-rich domains with the latter amino acid serving as the better acceptor (Strahl-Bolsinger *et al.*, 1999) in cell-free studies (Table III).

Since the purification and characterization from Chinese hamster ovary cells of the transferase responsible for the formation of the *Fuc*-α-*Ser/Thr* linkage (Wang and Spellman, 1998), it has been cloned from a human heart cDNA library (Wang *et al.*, 2001). Transcripts of this gene were observed to be expressed in all human tissues examined, and moreover homologs were found in mice, *Drosophila*, and *C. elegans* (Wang *et al.*, 2001). The enzyme was observed to be membrane associated and its type II transmembrane structure was believed to be consistent with a Golgi localization. *O*-fucosylation has been shown on EGF modules of various proteins and a consensus sequence has been identified (Table III) in which the glycosylation site is situated between the second and third conserved cysteine residues (Harris and Spellman, 1993). A recent report on human platelet thrombospondin indicated that the *Fuc*-α-*Ser/Thr* linkages occur outside of the EGF module in a peptide sequence somewhat different from those in other proteins, suggesting that the consensus sequence may be broader than believed or more than one fucosyltransferase may exist (Hofsteenge *et al.*, 2001).

Although the enzymatic formation of the *Glc*-β-*Ser* glycopeptide bond has not as yet been elucidated, it is apparent from structural studies conducted so far that this glucose modification, like fucosylation, is directed toward a consensus sequence on the EGF domain (Table III) and on the basis of studies on human factor IX it would appear that the glucose is attached to Ser located between the first and second conserved cysteine residues of the EGF motif (Harris and Spellman, 1993). More specifically, it has been shown that Ser-52 of human factor VII and Ser-53 of human factor IX as well as human and bovine protein Z are *O*-glucosylated (Nishimura *et al.*, 1989).

The initial step in proteoglycan biosynthesis is mediated by a glycosyltransferase that establishes the *Xyl*-β-*Ser* bond. In rat liver the enzyme appears to be primarily Golgi-situated (Nuwayhid *et al.*, 1986), whereas in chick chondrocytes it has been observed to be present in late ER and early Golgi compartments (Vertel *et al.*, 1993). This xylosyltransferase has

been purified from rat chondrosarcoma (Schwartz and Dorfman, 1975) and rat ear cartilage (Pfeil and Wenzel, 2000). The enzyme has also been isolated from human choriocarcinoma cells (Kuhn *et al.*, 2001) and cloned from this source (Götting *et al.*, 2000). The amino acid sequences around the attachment sites have been documented and shown not to be invariable (Esko and Zhang, 1996). However, in general they are represented by the motif shown in Table III where Ala can substitute for the more common Gly residue; furthermore, one or more acidic amino acids are found in close proximity to the glycopeptide bond.

The formation of the *Glc-Thr* and *GlcNAc-Thr* linkages represent events that are of a pathological nature (Table II). Both linkages are generated in the cytosol of *Clostridium*-infected mammalian cells through the action of the bacterial cytotoxins on the Rho family of small GTPases, including its Rac and Cdc42 members, resulting in an inhibition of their activity. Though *C. difficile* and *C. sordelli* toxins transfer Glc to Thr (Just *et al.*, 1995), the toxin from *C. novyi* adds GlcNAc to this amino acid (Selzer *et al.*, 1996). The specific residues that are modified have been identified (Table III) and cloning of the *C. difficile* (Eichel-Streiber *et al.*, 1992) and *C. novyi* (Hofmann *et al.*, 1995) toxins has been achieved.

The transferase involved in the genesis of the *Gal-Thr* glycopeptide linkage of cuticle collagens has not as yet been identified but structural studies on the hydrothermal vent worm collagen (Mann *et al.*, 1996) have indicated that the glycosylated Thr residues are found in the Gly-X-*Thr* positions (Table III). These substituted Thr constituents are believed to replace Hyp as the primary contributor to triple helix stabilization.

The galactosyltransferase responsible for the synthesis of the *Gal*-β-*Hyl* was found to be widely distributed in the tissues of the rat, including kidney, cartilage, spleen and lung (Spiro and Spiro, 1971b). Its action is directed toward the collagen triplet (Table III) and requires that the ε-amino group of the Hyl to be unsubstituted (Spiro and Spiro, 1971a). Golgi localization of the enzyme was indicated by its association with light membrane fractions (Spiro and Spiro, 1971a) and by *in vivo* studies on the hepatic core specific lectin that contains collagen-like domains in which glycosylated Hyl residues reside (Colley and Baenziger, 1987). This intracellular site is in accord with the fact that a transporter for UDP-Gal is present in the Golgi apparatus and not in the ER (Hirschberg *et al.*, 1998).

The enzyme involved in the formation of the *Ara*-α-*Hyp* bond has not yet been characterized, but it has been determined that in higher plants repetitive Hyp-rich modules (Table III) are the site of arabinogalactan attachment (Kieliszewski *et al.*, 1995).

The cytoplasmic GlcNAc-transferase of *Dictyostelium* involved in the biogenesis of the *GlcNAc-Hyp* sugar–amino acid connection has been purified (Teng-umnuay *et al.*, 1999) and recently cloned (West *et al.*, 2001). Because this novel linkage has so far only been observed in the Skp1 component of the SCF complex, the assay of the glycosylation enzyme employed the Skp1 protein or its peptides. Attachment of the GlcNAc was shown to occur to a Hyp residue at amino acid position 143 and the enzyme works in conjunction with a series of other cytoplasmic glycosyltransferases to form a pentasaccharide carbohydrate unit (Teng-umnuay *et al.*, 1998). The cytoplasmic location of the GlcNAc-transferase suggested to these investigators that a bidirectional flow of the Skp1 protein through the ER membrane must occur because hydroxylation of Pro is believed to take place inside the vesicles (Teng-umnuay *et al.*, 1998).

It has been established that the *Glc*-α-*Tyr* linkage of mammalian glycogenin occurs on Tyr 194 of this protein and, moreover, that the formation of this bond quite uniquely is an autocatalytic cytosolic event (Alonso *et al.*, 1994) occurring between the two subunits of this primer protein (Lin *et al.*, 1999). The enzyme (i.e., glycogenin) has been cloned (Viskupic *et al.*, 1992) and it has been shown that although mutation of Tyr-194 to Phe or Thr results in the loss of the self-glucosylating activity, the glycogenin retains its capacity to transfer glucose to exogenous acceptors (Cao *et al.*, 1995). The recombinant monoglucosylated glycogenin can serve as an acceptor for mammalian glycogen synthase (Viskupic *et al.*, 1992); the K_m for the latter enzyme is 1000-fold greater than for the glucosyltransferase that forms the glycopeptide bond (Pitcher *et al.*, 1988).

Phosphoglycosylation

The enzymatic attachment of a sugar to the polypeptide chain through a phosphodiester bridge, which has been termed phosphoglycosylation (Mehta *et al.*, 1996), has been investigated in *Dictyostelium* and *Leishmania* (Table II). The GlcNAc-1-phosphotransferase was partially purified from *Dictyostelium* and localized to light membranes that are believed to represent the Golgi compartment (Merello *et al.*, 1995). Subsequent studies indicated that the enzyme recognizes Ser-containing peptides of various *Dictyostelium* proteins among which cysteine proteinases are the most prominent (Mehta *et al.*, 1997). Although no single specific motif was observed in the peptide acceptor, it was determined that the transfers occur in Ser-rich domains in which the flanking Ala residues preferentially influence phosphoglycosylation (Table III); Thr residues were not phosphoglycosylated by the enzyme.

Man-1-phosphotransferase has been characterized in *Leishmania mexicana* promastigotes and it is believed to be situated in the *cis*-Golgi compartment (Moss *et al.*, 1999). The enzyme adds Man-α-1-phosphate to Ser residues in domains rich in this amino acid; it does not act on Thr and its action is promoted by flanking Asp and Glu residues.

C-*mannosylation*

The enzyme which links C-1 of mannose to the C-2 atom of the indole ring of Trp has been found to be present in a variety of cultured mammalian cells (Krieg *et al.*, 1997) and has been studied in rat liver microsomes (Doucey *et al.*, 1998). Convincing evidence has been obtained that the glycosyl donor in this reaction is Dol-P-Man, and indeed it was reported that *C*-mannosylation is considerably reduced in Lec15 Chinese hamster ovary cells that are deficient in Dol-P-Man synthase activity (Doucey *et al.*, 1998). Furthermore, it was recently shown that *C*-mannosylation of Trp, along with all previously known classes of Dol-P-monosaccharide-dependent glycosyltransferase reactions, is regulated in hamster by the Lec35 gene, which is required for Dol-P-mannose utilization (Anand *et al.*, 2001). The dependence of the *C*-mannosylation on Dol-P-Man strongly suggests that it takes place in the ER, where all known Dol-P-Man-dependent reactions are localized (Anand *et al.*, 2001). The recognition signal for *C*-mannosylation

has been assigned to a Trp-X-X-Trp sequence (Table III) in which the first Trp becomes glycosylated (Krieg *et al.*, 1998; Doucey *et al.*, 1998; Hartmann and Hofsteenge, 2000); the Trp at position +3 is also important for the glycosylation to take place as the transfer activity was abolished when this amino acid was mutated to Ala and reduced to one-third when replaced by Phe (Krieg *et al.*, 1998). A survey of protein databases has indicated that the Trp-X-X-Trp consensus sequence is present in 336 mammalian proteins, suggesting the possibility that *C*-mannosylation may occur quite frequently in higher eukaryotes (Krieg *et al.*, 1998).

Glypiation

The process of adding GPI to proteins, which has been termed glypiation, is carried out by an ER-situated transamidase that cleaves the C-terminal peptide and concomitantly transfers the preassembled GPI anchor to the newly exposed carboxy-terminal amino acid residue to establish an amide bond between the latter and the ethanolamine moiety of the glycolipid (Kinoshita *et al.*, 1997; Ferguson, 1999). In contrast to the assembly of the oligosaccharide involved in formation of the *N*-glycosidic linkage to Asn, it is believed that GPI assembly takes place entirely on the cytoplasmic side of the ER and is presumably followed by its translocation to the lumenal side, where attachment to the protein takes place. The transamidase reaction has been observed in eukaryotes ranging from yeast to mammals and is believed to be carried out by a multiprotein complex that has as yet not been isolated in its intact form. The genes of two components (GPI8 and GAA1) have been cloned from yeast (Benghezal *et al.*, 1996; Hamburger *et al.*, 1995). Intensive studies have been carried out regarding the carboxy-terminal signal peptide that directs GPI attachment. It has been noted that this peptide, which has to be cleaved prior to binding of the GPI and consists of 15–30 amino acids, has structural similarities to the NH_2-terminal peptide that functions in general to direct nascent chains into the ER lumen (Micanovic *et al.*, 1990; Gerber *et al.*, 1992). The residue to which GPI becomes attached (termed ω) has small side chains (e.g., Gly, Ala, Cys, Ser, Asn) as does the amino acid in the ω+2 position (e.g., Gly, Ala). The latter site is followed by a short hydrophilic domain (5–7 residues) and this is followed by a hydrophobic region (12–20 residues) that extends to the carboxy-terminus of the signal peptide. The ω+1 position apparently can be filled by any amino acid except Pro or Trp.

Enzymatic cleavage of glycopeptide bonds

Although a number of endoglycosidases and glycosidases, usually of bacterial or plant origin, have been effectively employed to split *N*- and *O*-glycosidic bonds in structural investigations (Kobata, 1979; Maley *et al.*, 1989), brief mention will be made only of eukaryotic enzymes active at neutral pH that appear to play an important physiological role.

Cleavage of the GlcNAc-β-Ser/Thr linkage has been shown to take place through the action of a specific cytosolically situated β-*N*-acetylglucosaminidase, which has been purified from rat spleen cytosol (Dong and Hart, 1994) and recently cloned from human brain (Gao *et al.*, 2001). This enzyme is expressed in every human tissue examined and is believed to be a key component of the postulated regulatory cycle in which Ser/Thr residues on various nuclear and cytoplasmic proteins can be modified alternatively by *O*-GlcNAc or *O*-phosphate groups.

The finding that the release of polymannose oligosaccharides from their Asn linkage into the cytosol and ER lumen from newly synthesized glycoproteins occurs as part of the ER-associated quality control of misfolded or improperly oligomerized proteins (Moore and Spiro, 1994; Spiro and Spiro, 2001) led to the finding that peptide-*N*-glycosidases that cleave the *GlcNAc*-β-*Asn* bond occur in the cytosol and ER of liver and other eukaryotic tissues (Suzuki *et al.*, 1997, 1998; Weng and Spiro, 1997).

Influence of enzymatic peptide glycosylation on human disease

It has become evident in recent years that defects in the attachment of carbohydrate to protein have been implicated in a number of human diseases. The congenital disorders of glycosylation represent a group of systemic diseases characterized most prominently by neurological and developmental deficiencies, and these have been well defined at a molecular level (Freeze and Westphal, 2001; Schachter, 2001). At the present time six variants have been described that can be ascribed to specific enzymatic defects responsible for the impairment at different stages of dolichylpyrophosphate oligosaccharide assembly. Because the lipid-linked oligosaccharide is the glycosyl donor in the formation of the *GlcNAc*-β-*Asn* linkage, this group of multisystem diseases ultimately represents a disorder of *N*-glycosylation.

As already indicated, the glycosylation by clostridial cytotoxins of a specific Thr residue on proteins that belong to the Rho family of mammalian GTPases is highly relevant to human disease (Busch and Aktories, 2000). Indeed, toxins produced by members of the *Clostridium* genus have been shown to be responsible for the causation of such pathological states as botulism, gas gangrene, antibiotic-associated diarrhea, and pseudomembranous colitis. Rho proteins act as molecular switches to control cellular processes, such as the organization of the actin cytoskeleton in eukaryotes, by cycling between the active GTP- and inactive GDP-bound states. The highly conserved Thr residue which is the substrate for the glycosylation by the toxins is involved in the nucleotide binding. It is believed that modification of this amino acid by addition of a GlcNAc or Glc residue results in a loss of effector binding by the Rho protein and inhibition of GTPase activity.

Although observations were made some time ago that formation of glucosamine from glucose was strongly favored over the production of glycogen in diabetic rats (Spiro, 1959, 1963), in more recent years this enhanced hexosamine flux has been the subject of intensive investigations that were to a large extent based on the description of the *GlcNAc*-β-*Ser/Thr* carbohydrate unit by Hart and his collaborators (Hart, 1997). Indeed the hexosamine flux hypothesis has been supported by the findings that high levels of GlcNAc brought about by infusion of this sugar (Hawkins *et al.*, 1996) or overexpression of the key enzyme for Glc to GlcNAc conversion, namely glutamine:fructose-6-phosphate amidotransferase (Hebert *et al.*, 1996), promotes the development of insulin resistance in rodents and cultured cells. The mechanism is believed to involve an increase in the presence of Ser/Thr-linked GlcNAc on important regulatory proteins through enhanced UDP-GlcNAc

transfer by the GlcNAc:protein transferase (Tang *et al.*, 2000; Akimoto *et al.*, 2001). Presumably the resulting hyperglycemia would then lead to the microvascular complications, which are the hallmark of the uncontrolled diabetic condition (Spiro, 1976). The observations that the level of GlcNAc transferase transcripts is particularly high in the β-cells of islets of Langerhans and that the experimental diabetes agent, streptozotocin (a structural analog of GlcNAc) selectively inhibits the β-*N*-acetyl-glucosaminidase activity *in vitro* (Hanover *et al.*, 1999) and elevates *GlcNAc*-β-*Ser/Thr* levels in pancreas of diabetic rats (Akimoto *et al.*, 2001) has added to the intriguing puzzle relating to the generation of insulin-resistant diabetes.

Leukocyte adhesion deficiency II, a rare disorder characterized by recurrent infections and severe mental and growth retardation, has been attributed to a lack of GDP-Fuc formation due to impaired activity of GDP-mannose-4,6-dehydratase (Becker and Lowe, 1999). Recently a closely related disorder was described in a patient who manifested a decreased import of GDP-Fuc into the Golgi (Lübke *et al.*, 1999). Symptoms of the disease have been attributed to the absence of fucosylated selectin ligands (Phillips *et al.*, 1995). However, the realization that *Fuc*-α-*Ser/Thr* occurs on the EGF modules of the Notch1 receptor (Moloney *et al.*, 2000), either as an unsubstituted monosaccharide or an oligosaccharide species (Wang *et al.*, 2001) has opened the possibility that some of the developmental defects seen in leukocyte adhesion deficiency syndrome could be the result of impaired formation of the *Fuc*-α-*Ser/Thr* bond due to lack of the GDP-Fuc glycosyl donor.

Paroxysmal nocturnal hemoglobinuria is a disorder characterized by recurrent bouts of complement-mediated intravesicular hemolysis in which there is a defect in the biosynthesis of the GPI anchor of granulocytes and B lymphocytes (Tomita, 1999). Although it has been determined that the gene that is mutated in this condition is PIG-A, which is involved in the addition of GlcNAc to the inositol residue of the phosphatidylinositol (Takeda *et al.*, 1993), this disruption in the multistep GPI assembly ultimately results in an impairment of the *en bloc* attachment of the completed anchor to protein and therefore can be considered to bring about a defect in glypiation.

Concluding remarks

It is apparent from this review that a strikingly large number of diverse carbohydrate–peptide linkages exist in nature and that many of these occur throughout the phylogenetic range, extending from the most primitive microorganisms to highly differentiated multicellular animals and plants. Clearly, the appearance of protein-linked oligosaccharides so early in evolution suggests that this co- or posttranslational event with its elaborate enzymatic machinery plays an essential biological role, because it takes place even in prokaryotic cells without secretory channels. At first consideration it appears surprising that such a high diversity of glycopeptide linkages has evolved. However, appreciation of the critical role that oligosaccharides play in the three-dimensional framework of proteins with diverse amino acid sequences makes glycopeptide bond multiplicity more understandable, since the formation of these crucial linkages determines to a large extent the nature of the final carbohydrate units that are subsequently formed by the numerous processing enzymes. A chronological survey of the description of new glycopeptide linkages suggests that this stream of discoveries will continue for some time into the future. Indeed, the finding in a variety of proteins of mannose linked by a C-C bond to the indole ring of Trp indicates that protein glycosylation does not even require an amino acid functional group and thereby expands the scope of future investigations into novel bonds. Although considerable attention has been given to glycopeptide linkages of eukaryotes, further exploration of archaea and eubacteria glycoproteins promises to yield much new information. The as-yet-incomplete study of the enzymes involved in the attachment of saccharides to protein suggests that in many instances, as exemplified by the multiple α-GalNAc and α-Man:polypeptide transferases, a family of closely related enzymes may be involved in the formation of the same glycopeptide bond with each member of this group having specificity for different proteins or even different regions of the same polypeptide chain. Because glycosylation of proteins appears to be a highly directed process, up to now the difficulty in finding an invariant peptide consensus sequence for a number of sugar–amino acid bonds may be due in part to a lack of resolution of all the members of such enzyme families. It is anticipated moreover that in future years there will be a major expansion in the elucidation of disease states in which attachment of saccharides to protein is altered through genetic or intracellular environmental factors.

Acknowledgments

Work from the author's laboratory was funded by grants DK17325 and DK17477 from the National Institutes of Health.

Abbreviatons

EGF, epidermal growth factor; ER, endoplasmic reticulum; GPI, glycophosphatidylinositol; Hyp, hydroxyproline; Hyl, hydroxylysine; PMT, protein *O*-mannosyltransferase; SCF, Skp1-cullin-F-box; VSG, variant surface glycoprotein; *Dictyostelium* when used in the text refers to the *discoideum* species.

References

Akimoto, Y., Kreppel, L.K., Hirano, H., Hart, G.W. (2000) Increase O-GlcNAc transferase in pancreas of rats with streptozotocin-induced diabetes. *Diabetologia*, **43**, 1239–1247.

Akimoto, Y., Kreppel, L.K., Hirano, H., and Hart, G.W. (2001) Hyperglycemia and the O-GlcNAc transferase in rat aortic smooth muscle cells: elevated expression and altered patterns of O-GlcNAcylation. *Arch. Biochem. Biophys.*, **389**, 166–175.

Allen, A.K., Desai, N.N., Neuberger, A., and Creeth, J.M. (1978) Properties of potato lectin and the nature of its glycoprotein linkages. *Biochem. J.*, **171**, 665–674.

Alonso, M.D., Lomako, J., Lomako, W.M., and Whelan, W.J. (1994) Tyrosine-194 of glycogenin undergoes autocatalytic glucosylation but is not essential for catalytic function and activity. *FEBS Lett.*, **342**, 38–42.

Anand, M., Rush, J.S., Ray, S., Doucey, M.-A., Weik, J., Ware, F.E., Hofsteenge, J., Waechter, C.J. and Lehrman, M.A. (2001) Requirement of the Lec35 gene for all known classes of monosaccharide-P-dolichol-dependent glycosyltransferase reactions in mammals. *Mol. Biol. Cell*, **12**, 487–501.

Apweiler, R., Hermjakob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT data base *Biochim. Biophys. Acta*, **1473**, 4–8.

Bause, E. (1983) Structural requirements of *N*-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.*, **209**, 331–336.

Bause, E. and Legler, G. (1981) The role of the hydroxyamino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the *N*-glycosylation step during glycoprotein biosynthesis. *Biochem. J.*, **195**, 639–644.

Becker, D.J. and Lowe, J.B. (1999) Leukocyte adhesion deficiency type II. *Biochim. Biophys. Acta*, **1455**, 193–204.

Benghezal, M., Benachour, A., Rusconi, S., Aebi, M., and Conzelmann, A. (1996) Yeast Gpi8p is essential for GPI anchor attachment onto proteins. *EMBO J.*, **15**, 6575–6583.

Bjoern, S., Foster, D.C., Thim, L., Wiberg, F.C., Christensen, M., Komiyama, Y., Pederson, A.H., and Kisiel, W. (1991) Human plasma and recombinant factor VII. Characterization of *O*-glycosylation of serine residues 52 and 60 and effect of site-directed mutagenesis of serine 52 to alanine. *J. Biol. Chem.*, **266**, 11051–11057.

Bock, K., Schuster-Kolbe, J., Altman, E., Allmaier, G., Stahl, B., Christian, R., Sleytr, U.B., and Messner, B. (1994) Primary structure of the *O*-glycosidically linked glycan chain of the crystalline surface layer glycoprotein of *Thermoanaeobacter thermohydrosulfuricus* L111-69. Galactosyl tyrosine as a novel linkage unit. *J. Biol. Chem.*, **269**, 7137–7144.

Buko, A.M., Kentzer, E.J., Petros, A., Menon, G., Zuiderweg, E.R., and Sarin, V.K. (1991) Characterization of a posttranslational fucosylation in the growth factor domain of urinary plasminogen activator. *Proc. Natl Acad. Sci. USA*, **88**, 3992–3996.

Burda, P. and Aebi, M. (1999) The dolichol pathway of *N*-linked glycosylation. *Biochim. Biophys. Acta*, **1426**, 239–257.

Busch, C. and Aktories, K. (2000) Microbial toxins and the glycosylation of Rho family GTPases. *Curr. Opin. Struct. Biol.*, **10**, 528–535.

Cao, Y., Steinrauf, L.K., and Roach, P.J. (1995) Mechanism of glycogenin self-glucosylation. *Arch. Biochem. Biophys.*, **319**, 293–298.

Castric, P., Cassels, F.J., and Carlson, R.W. (2001) Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan. *J. Biol. Chem.*, **276**, 26479–26485.

Chalifour, R.J. and Spiro, R.G. (1988) Effects of phospholipids on thyroid oligosaccharyltransferase activity and orientation: Evaluation of structural determinants for stimulation of *N*-glycosylation. *J. Biol. Chem.*, **263**, 15673–15680.

Clausen, H. and Bennett, E.P. (1996) A family of UDP-GalNAc:polypeptide *N*-acetylgalactosaminyl-transferases control the initiation of mucin-type *O*-glycosylation. *Glycobiology*, **6**, 635–646.

Colley, K.J. and Baenziger, J.U. (1987) Identification of the post-translational modifications of the core specific lectin. The core specific lectin contains hydroxyproline, hydroxylysine and glucosylgalactosylhydroxylysine residues. *J. Biol. Chem.*, **262**, 10290–10295.

Collier, E. and Gorden, P. (1991) *O*-linked oligosaccharides on insulin receptor. *Diabetes*, **40**, 197–203.

Comer, F.I. and Hart, G.W. (2000) *O*-glycosylation of nuclear and cytosolic proteins. Dynamic interplay between *O*-GlcNAc and *O*-phosphate. *J. Biol. Chem.*, **275**, 29179–29182.

Cummings, R.D. (1992) Synthesis of asparagine-linked oligosaccharides. Pathways, genetics and metabolic regulation. In Allen, H.J. and Kisalius, E.C. (eds.), *Glyconjugates: composition, structure and function*. Marcel Dekker, New York, pp. 333–360.

de Beer, T., Vliegenthart, J.F., Löffler, A., and Hofsteenge, J. (1995) The hexapyranosyl residue that is *C*-glycosidically linked to the side chain of tryptophan-7 in human RNase U_s, is α-mannopyranose. *Biochemistry*, **34**, 11785–11789.

Dobos, K.M., Khoo, K.-H., Swiderek, K.M., Brennan, P.J., and Belisle, J.T. (1996) Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. *J. Bacteriol.*, **178**, 2498–2506.

Dong, D.L.-Y. and Hart, G.W. (1994) Purification and characterization of an *O*-GlcNAc selective *N*-acetyl-β-D-glucosaminidase from rat spleen cytosol. *J. Biol. Chem.*, **269**, 19321–19330.

Doucey, M.-A., Hess, D., Blommers, M.J.J., and Hofsteenge, J. (1999) Recombinant human interleukin-12 is the second example of *C*-mannosylated protein. *Glycobiology*, **9**, 435–441.

Doucey, M.-A., Hess, D., Cacan, R., and Hofsteenge (1998) Protein *C*-mannosylation is enzyme-catalysed and uses dolichyl-phosphate-mannose as a precursor. *Mol. Biol. Cell*, **9**, 291–300.

Edge, A.S.B. and Spiro, R.G. (1997) Structure of the *O*-linked oligosaccharides from a major thyroid cell surface glycoprotein. *Arch. Biochem. Biophys.*, **343**, 73–80.

Eichel-Streiber von, C., Laufenberg-Feldmann, R., Sartingen, S., Schulze, J., and Sauerborn, M. (1992) Comparative sequence analysis of the *Clostridium difficile* toxins A and B. *Mol. Gen. Genet.*, **233**, 260–268.

Elhammer, A.P., Poorman, R.A., Brown, E., Maggiora, L.L., Hoogerheide, J.G., and Kézdy, F.J. (1993) The specificity of UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase as inferred from a data base of *in vivo* substrates and from *in vitro* glycosylation of proteins and peptides. *J. Biol. Chem.*, **268**, 10029–10038.

Endo, T. (1999) *O*-mannosyl glycans in mammals. *Biochim. Biophys. Acta*, **1473**, 237–246.

Erickson, P.R. and Herzberg, M.C. (1993) Evidence for the covalent linkage of carbohydrate polymers to a glycoprotein from *Streptococcus sanguis*. *J. Biol. Chem.*, **268**, 23780–23783.

Esko, J.D. and Zhang, L. (1996) Influence of core protein sequence on glycosaminoglycan assembly. *Curr. Opin. Struct. Biol.*, **6**, 663–670.

Fanger, M.W. and Smyth, D.G. (1972) The oligosaccharide units of rabbit immunoglobulin G. Multicarbohydrate attachment sites. *Biochem. J.*, **127**, 757–765.

Ferguson, M.A.J. (1999) The structure, biosynthesis and functions of glycosylphosphatidyl-inositol anchors, and the contributions of trypanosome research. *J. Cell Sci.*, **112**, 2799–2809.

Finne, J., Krusius, T., Margolis, R.K., and Margolis, R.U. (1979) Novel mannitol-containing oligosaccharides obtained by mild alkaline borohydride treatment of chondroitin sulfate proteoglycan from brain. *J. Biol. Chem.*, **254**, 10295–10300.

Freeze, H.H. and Westphal, V. (2001) Balancing *N*-linked glycosylation to avoid disease. *Biochimie*, **83**, 791–799.

Gao, Y., Wells, L., Comer, F.I., Parker, G.J., and Hart, G.W. (2001) Dynamic *O*-glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a neutral cytosolic β-*N*-acetylglucosaminidase from human brain. *J. Biol. Chem.*, **276**, 9838–9845.

Gerber, L.D., Kodukula, K., and Udenfriend, S. (1992) Phosphatidylinositol glycan (PI-G) anchored membrane proteins. Amino acid requirements adjacent to the site of cleavage and PI-G attachment in the COOH-terminal signal peptide. *J. Biol. Chem.*, **267**, 12168–12173.

Gerwig, G.J., Kamerling, J.P., Vliegenthart, J.F.G., Morag, E., Lamed, R., and Bayer, E.A. (1993) The nature of the carbohydrate-peptide linkage region in glycoproteins from the cellulosomes of *Clostridium thermocellum* and *Bacteroides cellulosolvens*. *J. Biol. Chem.*, **268**, 26959–26960.

Götting, C., Kuhn, J., Zahn, R., Brinkmann, T., and Kleesiek, K (2000) Molecular cloning and expression of human UDP-D-xylose:proteoglycan core protein B-D-xylosyltransferase and its first isoforms XT-II. *J. Mol. Biol.*, **304**, 517–528.

Guha-Niyogi, A., Sullivan, D.R., and Turco, S.J. (2001) Glycoconjugate structures of parasitic protozoa. *Glycobiology*, **11**, 45R-59R.

Haltiwanger, R.S., Blomberg, M.A., and Hart, G.W. (1992) Glycosylation of nuclear and cytoplasmic proteins. Purification and characterization of a uridine diphospho-*N*-acetylglucosamine:polypeptide β-*N*-acetylglucosaminyltransferase. *J. Biol. Chem.*, **267**, 9005–9013.

Haltiwanger, R.S., Busby, S., Grove, K., Li, S., Mason, D., Medina, L., Moloney, D., Philipsberg, G., and Scartozzi, R. (1997) *O*-glycosylation of nuclear and cytoplasmic proteins: regulation analogous to phosphorylation? *Biochem. Biophys. Res. Commun.*, **231**, 237–242.

Hamburger, D., Egerton, M., and Riezman, H. (1995) Yeast Gaa1p is required for attachment of a completed GPI anchor onto proteins. *J. Cell Biol.*, **129**, 629–639.

Hanover, J.A., Lai, Z., Lee, G., Lubas, W.A., and Sato, S.M. (1999) Elevated *O*-linked *N*-acetylglucosamine metabolism in pancreatic β-cells. *Arch. Biochem. Biophys.*, **362**, 38–45.

Harris, R.J., Ling, V.T., and Spellman, M.W. (1992) *O*-linked fucose is present in the first epidermal growth factor domain of factor XII but not protein C. *J. Biol. Chem.*, **267**, 5102–5107.

Harris, R.J. and Spellman, M.W. (1993) *O*-linked fucose and other posttranslational modifications unique to EGF modules. *Glycobiology*, **3**, 219–224.

Hart, G.W. (1997) Dynamic *O*-linked glycosylation of nuclear and cytoskeletal proteins. *Annu. Rev. Biochem.*, **66**, 315–335.

Hartmann, S. and Hofsteenge, J. (2000) Properdin, the positive regulator of complement, is highly *C*-mannosylated. *J. Biol. Chem.*, **275**, 28569–28574.

Haselbeck, A. and Tanner, W. (1983) *O*-glycosylation in *Saccharomyces cereivisiae* is initiated at the endoplasmic reticulum. *FEBS Lett.*, **158**, 335–338.

Hawkins, M., Barzilai, N., Chen, W., Angelov, I., Hu, M., Cohen, P., Rossetti, L. (1996) Increased hexosamine availability similarly impairs the action of insulin and IGF-I on glucose disposal. *Diabetes*, **45**, 1734–1743.

Haynes, P.A. (1998) Phosphoglycosylation: a new structural class of glycosylation? *Glycobiology*, **8**, 1–5.

Hebert, L.F., Jr., Daniels, M.C., Zhou, J., Crook, E.D., Turner, R.L., Simmons, S.T., Neidigh, J.L., Zhu, J.S., Baron, A.D., and McClain, D.A. (1996) Overexpression of glutamine:fructose-6-phosphate amidotransferase in transgenic mice leads to insulin resistance. *J. Clin. Invest.*, **98**, 930–936.

Herscovics, A. and Orlean, P. (1993) Glycoprotein biosynthesis in yeast. *FASEB J.*, **7**, 540–550.

Hirschberg, C.B., Robbins, P.W., and Abeijon, C. (1998) Transporters of nucleotide sugars, ATP and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus (1998) *Annu. Rev. Biochem.*, **67**, 49–69.

Hofmann, F., Herrmann, A., Habermann, E., and von Eichel-Streiber, C. (1995) Sequencing and analysis of the gene encoding the alpha-toxin of *Clostridium novyi* proves its homology to toxins A and B of Clostridium difficile. *Mol. Gen. Genet.*, **247**, 670–679.

Hofsteenge, J., Huwiler, K.G., Macek, B., Hess, D., Lawler, J., Mosher, D.F., and Peter-Katalinic, J. (2001) *C*-mannosylation and *O*-fucosylation of the thrombospondin Type 1 module. *J. Biol. Chem.*, **276**, 6485–6498.

Holt, G.D., Haltiwanger, R.S., Torres, C.-R., and Hart, G.W. (1987) Erythrocytes contain cytoplasmic glycoproteins. *O*-linked GlcNAc on Band 4.1. *J. Biol. Chem.*, **262**, 14847–14850.

Johansen, P.G., Marshall, R.D., and Neuberger, A. (1961) Carbohydrates in protein. The preparation and some of the properties of a glycopeptide from hen egg ovalbumin. *Biochem. J.*, **78**, 518–527.

Jung, E., Gooley, A.A., Packer, N.H., Karuso, P., and Williams, K.L. (1998) Rules for the addition of *O*-linked *N*-acetylglucosamine to secreted proteins in *Dictyostelium discoideum. Eur. J. Biochem.*, **253**, 517–524.

Jurado, L.A.P., Coloma, A., and Cruces, J. (1999) Identification of a human homolog of the *Drosophila rotated abdomen* gene (POMT1) encoding a putative protein *O*-mannosyltransferase, and assignment to human chromosome 9q34.1. *Genomics*, **58**, 171–180.

Just, I., Wilm, M., Selzer, J., Rex, G., von Eichel-Streiber, C., Mann, M., and Aktories, K. (1995) The enterotoxin from *Clostridium difficile* (ToxA) monoglucosylates the Rho proteins. *J. Biol. Chem.*, **270**, 13932–13936.

Kieliszewski, M.J., O'Neill, M., Leykam, J., and Orlando, R. (1995) Tandem mass spectrometry and structural elucidation of glycopeptides from hydroxyproline-rich plant cell wall glycoprotein indicates that contiguous hydroxyproline residues are the major sites of hydroxyproline *O*-arabinosylation. *J. Biol. Chem.*, **270**, 2541–2549.

Kinoshita, T., Ohishi, K., and Takeda, J. (1997) GPI-anchor synthesis in mammalian cells: genes, their products, and a deficiency. *J. Biochem. (Tokyo)*, **122**, 251–257.

Kjellén, L. and Lindahl, U. (1991) Proteoglycans: structures and interactions. *Annu. Rev. Biochem.*, **60**, 443–475.

Knauer, R. and Lehle, L (1999) The oligosaccharyltransferase complex from *Saccharomyces cerevisiae*. Isolation of the OST6 gene, its synthetic interaction with OST3, and analysis of the native complex. *J. Biol. Chem.*, **274**, 17249–17256.

Kobata, A. (1979) Use of endo- and exoglycosidases for structural studies of glycoconjugates. *Anal. Biochem.*, **100**, 1–14.

Kobayashi, T., Nishizaki, R., and Ikezawa, H. (1997) The presence of GPI-linked protein(s) in an archaeobacterium, *Sulfolobus acidocaldarius*, closely related to eukaryotes. *Biochim. Biophys. Acta*, **1334**, 1–4.

Kreppel, L.K., Blomberg, M.A., and Hart, G.W. (1997) Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique *O*-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.*, **272**, 9308–9315.

Krieg, J., Hartmann, S., Vicentini, A., Gläsner, W., Hess, D., and Hofsteenge, J. (1998) Recognition signal for *C*-mannosylation of Trp-7 in RNase 2 consists of the sequence Trp-x-x-Trp. *Mol. Biol. Cell*, **9**, 301–309.

Krieg, J., Vicentini, A., Gläsner, W., Doucey, M.-A., Löffler, A., Hess, D., and Hofsteenge, J. (1997) *C*-mannosylation of human RNase 2 is an intracellular process performed by a variety of cultured cells. *J. Biol. Chem.*, **272**, 26687–26692.

Kuhn, J., Götting, C., Schnölzer, M., Kempf, T., Brinkmann, T., and Kleesiek, K (2001) First isolation of human UDP-D-xylose:proteoglycan core protein β-D-xylosyltransferase secreted from cultured JAR choriocarcinoma cells. *J. Biol. Chem.*, **276**, 4940–4947.

Lamport, D.T.A., Katona, L., and Roerig, S. (1973) Galactosylserine in extensin. *Biochem. J.*, **133**, 125–132.

Lechner, J. and Wieland, F. (1989) Structure and biosynthesis of prokaryotic glycoproteins. *Annu. Rev. Biochem.*, **58**, 173–194**.**

Lin, A., Mu, J., Yang, J., and Roach, P.J. (1999) Self glucosylation of glycogenin, the initiator of glycogen biosynthesis involves an intersubunit reaction. *Arch. Biochem. Biophys.*, **363**, 163–170.

Lohmander, L.S., DeLuca, S., Nilsson, B., Hascall, V.C. Caputo, C.B., Kimura, J.H., and Heinegard, D. (1980) Oligosaccharides on proteoglycans from swarm rat chondrosarcoma. *J. Biol. Chem.*, **255**, 6084–6091.

Lubas, W.A., Frank, D.W., Krause, M., and Hanover, J.A. (1997) *O*-linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *J. Biol. Chem.*, **272**, 9316–9324.

Lübke, T., Marquardt, T., von Figura, K., and Korner, C. (1999) A new type of carbohydrate-deficient glycoproteins syndrome due to a decreased import of GDP-fucose into the Golgi. *J. Biol. Chem.*, **274**, 25986–25989.

Lussier, M., Gentzsch, M., Sdicu, A.-M., Bussey, H., and Tanner, W. (1995) Protein *O*-glycosylation in yeast. The PMT2 gene specifies a second protein *O*-mannosyltransferase that functions in addition to the PMT1-encoded activity. *J. Biol. Chem.*, **270**, 2770–2775.

Maley, F., Trimble, R.B., Tarentino, A.L., and Plummer, T.H., Jr. (1989) Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal. Biochem.*, **180**, 195–204.

Mann, K., Mechling, D.E., Bächinger, H.P., Eckerskorn, C., Gaill, F., and Timple, R. (1996) Glycosylated threonine but not 4-hydroxyproline dominates the triple helix stabilizing positions in the sequence of a hydrothermal vent worm cuticle collagen. *J. Mol. Biol.*, **261**, 255–266.

Marchesi, V.T., Furthmayr, H., and Tomita, M. (1976) The red cell membrane. *Annu. Rev. Biochem.*, **45**, 667–698.

Marshall, R.D. (1974) The nature and metabolism of carbohydrate peptide linkages of glycoproteins. *Biochem. Soc. Symp.*, **40**, 17–26.

Mehta, D.P., Etchison, J.R., Wu, R., and Freeze, H.H. (1997) UDP-GlcNAc:Ser-protein *N*-acetylglucosamine-1-phosphotransferase from *Dictyostelium discoideum* recognizes serine-containing peptides and eukaryotic cysteine proteinases. *J. Biol. Chem.*, **272**, 28638–28645.

Mehta, D.P., Ichikawa, M., Salimath, P.V., Etchison, J.R., Haak, R., Manzi, A., and Freeze, H.H. (1996) A lysosomal cysteine proteinase from *Dictyostelium discoideum* contains *N*-acetylglucoamine-1-phosphate bound to serine but not mannose-6-phosphate on *N*-linked oligosaccharides. *J. Biol. Chem.*, **271**, 10897–10903.

Merello, S., Parodi, A.J., and Couso, R. (1995) Characterization and partial purification of a novel enzymatic activity. UDP-GlcNAc:Ser-protein *N*-acetylglucosamine-1-phosphotransferase from the cellular slime mold *Dictyostelium discoideum. J. Biol. Chem.*, **270**, 7281–7287.

Mescher, M.F. and Strominger, J.L. (1976) Purification and characterization of a prokaryotic glycoprotein from the cell envelope of *Halobacterium salinarium. J. Biol. Chem.*, **251**, 2005–2014.

Messner, P. (1997) Bacterial glycoproteins. *Glycoconj. J.*, **14**, 3–11.

Messner, P., Christian, R., Kolbe, J., Schulz, G., and Sleytr, U.B. (1992) Analysis of a novel linkage unit of *O*-linked carbohydrates from the crystalline surface layer glycoprotein of *Clostridium thermohydrosulfuricum* S102-70. *J. Bacteriol.*, **174**, 2236–2240.

Micanovic, R., Gerber, L.D., Berger, J., Kodukula, K., and Udenfriend, S. (1990) Selectivity of the cleavage/attachment site of phosphatidylinositol-glycan-anchored membrane proteins determined by site-specific mutagenesis at Asp-484 of placental alkaline phosphatase. *Proc. Natl Acad. Sci. USA*, **87**, 157–161.

Miller, D.H., Lamport, D.T.A., and Miller, M. (1972) Hydroxyproline heterooligosaccharides in *Chlamydomonas. Science*, **176**, 918–920.

Moloney, D.J., Shair, L.H., Lu, F.M., Xia, J., Locke, R., Matta, K.L., and Haltiwanger, R.S. (2000) Mammalian Notch1 is modified with two unusual forms of *O*-linked glycosylation found in epidermal growth factor-like modules. *J. Biol. Chem.*, **275**, 9604–9611.

Montreuil, J. (1980) Primary structure of glycoprotein glycans. *Adv. Carbohydr. Chem. Biochem.*, **37**, 157–223.

Moore, S.E.H. and Spiro, R.G. (1994) Intracellular compartmentalization and degradation of free polymannose oligosaccharides released during glycoprotein biosynthesis. *J. Biol. Chem.*, **269**, 12715–12721.

Moremen, K.W., Trimble, R.B., and Herscovics, A. (1994) Glycosidases of the asparagine-linked processing pathway. *Glycobiology*, **4**, 113–125.

Moss, J.M., Reid, G.E., Mullin, K.A., Zawadski, J.L., Simpson, R.J., and McConville, M.J. (1999) Characterization of a novel GDP-mannose:serine protein mannose-1-phosphotransferase from *Leishmania mexicana. J. Biol. Chem.*, **274**, 6678–6688.

Muir, L. and Lee, Y.C. (1970) Glycopeptides from earthworm cuticle collagen. *J. Biol. Chem.*, **245**, 502–509.

Nayak, B.R. and Spiro, R.G. (1991) Localization and structure of the asparagine-linked oligosaccharides of type IV collagen from glomerular basement membrane and lens capsule. *J. Biol. Chem.*, **266**, 13978–13987.

Nishimura, H., Kawabata, S., Kisiel, W., Hase, S., Ikenaka, T., Takoa, T., Shimonishi, Y., and Iwanaga, S. (1989) Identification of a disaccharide (Xyl-Glc) and trisaccharide (Xyl_2-Glc) *O*-glycosidically linked to a serine residue in the first epidermal growth factor-like domain in human factors VII and IX and protein Z and bovine protein Z. *J. Biol. Chem.*, **264**, 20320–20325.

Nishimura, H., Takao, T., Hase, S., Shimonishi, Y., and Iwanaga, S. (1992a) Human factor IX has a tetrasaccharide *O*-glycosidically linked to serine 61 through the fucose residue. *J. Biol. Chem.*, **267**, 17520–17525.

Nishimura, H., Yamashita, S., Zeng, Z., Walz, D.A., and Iwanaga, S. (1992b) Evidence for the existence of *O*-linked sugar chains consisting of glucose and xylose in bovine thrombospondin. *J. Biochem. (Tokyo)*, **111**, 460–464.

Nuwayhid, N., Glaser, J.H., Johnson, J.C., Conrad, H.E., Hauser, S.C., and Hirschberg, C.B. (1986) Xylosylation and glucuronosylation reactions in rat liver Golgi apparatus and endoplasmic reticulum. *J. Biol. Chem.*, **261**, 12936–12941.

O'Neill, M.A. and Selvendran, R.R. (1980) Glycoproteins from the cell wall of *Phaseolus coccineus. Biochem. J.*, **187**, 53–63.

Parthasarathy, N. and Spiro, R.G. (1984) Isolation and characterization of the heparan sulfate proteoglycan of the bovine glomerular basement membrane. *J. Biol. Chem.*, **259**, 12739–12755.

Perez-Vilar, J., Eckhardt, A.E., and Hill, R.L. (1996) Porcine submaxillary mucin forms disulfide-bonded dimers between its carboxyl-terminal domains. *J. Biol. Chem.*, **271**, 9845–9850.

Pfeil, U. and Wenzel, K.-W. (2000) Purification and some properties of UDP-xylosyltransferase of rat ear cartilage. *Glycobiology*, **10**, 803–807.

Phillips, M.L., Schwartz, B.R., Etzioni, A., Bayer, R., Ochs, H.D., Paulson, J.C., and Harlan, J.M. (1995) Neutrophil adhesion in leukocyte adhesion deficiency syndrome type 2. *J. Clin. Invest.*, **96**, 2898–2906.

Pitcher, J., Smythe, C., and Cohen, P. (1988) Glycogenin is the priming glucosyltransferase required for the initiation of glycogen biogenesis in rabbit skeletal muscle. *Eur. J. Biochem.*, **176**, 391–395.

Plummer, T.H., Jr., Tarentino, A.L., and Hauer, C.R. (1995) Novel, specific *O*-glycosylation of secreted *Flavobacterium meningosepticum* proteins. Asp-Ser and Asp-Thr-Thr consensus sites. *J. Biol. Chem.*, **270**, 13192–13196.

Previato, J.O., Sola-Penna, M., Agrellos, O.A., Jones, C., Oeltmann, T., Travassos, L.R., and Mendonça-Previato, L. (1998) Biosynthesis of *O*-*N*-acetylglucosamine-linked glycans in *Trypanosoma cruzi.* Characterization of the novel uridine diphopho-*N*-acetylglucosamine-polypeptide *N*-acetyl-glucosaminyltransferase catalyzing formation of *N*-acetylglucosamine α-1-*O*-threonine. *J. Biol. Chem.*, **273**, 14982–14988.

Qi, W., Fong, C., and Lamport, D.T.A. (1991) Gum arabic glycoprotein is a twisted hairy rope. A new model on *O*-galactosylhydroxyproline polysaccharide attachment. *Plant Physiol.*, **96**, 848–855.

Roth, J., Wang, Y., Eckhardt, A.E., and Hill, R.L. (1994) Subcellular localization of the UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyl-transferase-mediated *O*-glycosylation reactions in the submaxillary gland. *Proc. Natl Acad. Sci. USA*, **91**, 8935–8939.

Sadler, J.E. (1984) Biosynthesis of glycoproteins: formation of *O*-linked oligosaccharides. In Ginsburg, V. and Robbins, P.W. (eds.), *Biology of carbohydrates*, vol. 2. Wiley, New York, pp. 199–288.

Schachter, H. (2001) Congenital disorders involving defective *N*-glycosylation of proteins. *Cell Mol. Life Sci.*, **58**, 1085–1104.

Schäffer, C., Dietrich, K. Unger, B., Scheberl, A., Rainey, F.A., Köhlig, H., and Messner, P. (2000) A novel type of carbohydrate-protein linkage region of the tyrosine-bound S-layer glycan of *Thermoanaeobacterium thermosaccharolyticum* D120-70. *Eur. J. Biochem.*, **267**, 5482–5492.

Schäffer, C., Müller, N., Christian, R., Graninger, M., Wugeditsch, T., Scheberl, A., and Messner, P. (1999) Complete glycan structure of the S-layer glycoprotein of *Anerurinibacillus thermoaerophilus* GS4-97. *Glycobiology*, **9**, 407–414.

Schreiner, R., Schnabel, E., and Wieland, F. (1994) Novel *N*-glycosylation in eukaryotes: laminin contains the linkage unit β-glucosylasparagine. *J. Cell Biol.*, **124**, 1071–1081.

Schwartz, N.B. and Dorfman, A. (1975) Purification of rat chondrosarcoma xylosyltransferase. *Arch. Biochem. Biophys.*, **171**, 136–144.

Selzer, J., Hofmann, F., Rex, G., Wilm, M., Mann, M., Just, I., and Aktories, K. (1996) *Clostridium novyi* α-toxin-catalyzed incorporation of GlcNAc into Rho subfamily proteins. *J. Biol. Chem.*, **271**, 25173–25177

Shinkai, H. and Yonemasu, K. (1979) Hydroxylysine-linked glycosides of human complement subcomponent C1q and various collagens. *Biochem. J.*, **177**, 847–852.

Silberstein, S. and Gilmore, R. (1996) *Biochemistry*, molecular biology, and genetics of the oligosaccharyltransferase. *FASEB J.*, **10**, 849–858.

Singh, D.G., Lomako, J., Lomako, W.M., Whelan, W.J., Meyer, H.E., Serwe, M., and Metzger, J.W. (1995) β-Glucosylarginine: a new glucose-protein bond in a self-glucosylating protein from sweet corn. *FEBS Lett.*, **376**, 61–64.

Smythe, C. and Cohen, P. (1991) The discovery of glycogenin and the primary mechanism for glycogen biosynthesis. *Eur. J. Biochem.*, **200**, 625–631.

Spiro, M.J. and Spiro, R.G. (2001) Release of polymannose oligosaccharides from vesicular stomatitis virus G Protein during endoplasmic reticulum-associated degradation. *Glycobiology*, **11**, 803–811.

Spiro, M.J. and Spiro, R.G. (1971a) Studies on the biosynthesis of the hydroxylysine-linked disaccharide unit of basement membranes and collagens. II. Kidney galactosyltransferase. *J. Biol. Chem.*, **246**, 4910–4918.

Spiro, R.G. (1959) Role of insulin in two pathways of glucose metabolism: *in vivo* glucosamine and glycogen synthesis. *Ann. NY Acad. Sci.*, **82**, 366–373.

Spiro, R.G. (1963) Glycoproteins and diabetes. *Diabetes*, **12**, 223–230.

Spiro, R.G. (1967) The structure of the disaccharide unit of the renal glomerular basement membrane. *J. Biol. Chem.*, **242**, 4813–4823.

Spiro, R.G. (1969) Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens *J. Biol. Chem.*, **244**, 602–612.

Spiro, R.G. (1972a) Basement membranes and collagens. In Gottschalk, A. (ed.), *Glycoproteins. Their compostion, structure and function*, part B. Elsevier, pp. 964–999.

Spiro, R.G. (1972b) Study of the carbohydrates of glycoproteins. *Methods Enzymol.*, **28**, 3–43.

Spiro, R.G. (1973) Glycoproteins. *Adv. Protein Chem.*, **27**, 349–467.

Spiro, R.G. (1976) Search for a biochemical basis of diabetic microangiopathy. *Diabetologia*, **12**, 1–14.

Spiro, R.G. (2000) Glucose residues as key determinants in the biosynthesis and quality control of glycoproteins with *N*-linked oligosaccharides. *J. Biol. Chem.*, **275**, 35657–35660.

Spiro, R.G. and Bhoyroo, V.D. (1974) Structure of the *O*-glycosidically linked carbohydrate units of fetuin. *J. Biol. Chem.*, **249**, 5704–5717.

Spiro, R.G. and Bhoyroo, V.D. (1980) Studies on the carbohydrate of collagens. Characterization of a glucuronic acid-mannose disaccharide unit from *Nereis* cuticle collagen. *J. Biol. Chem.*, **255**, 5347–5354.

Spiro, R.G. and Spiro, M.J. (1971b) Studies on the biosynthesis of the hydroxylysine-linked disaccharide unit of basement membranes and collagens. III. Tissue and subcellular distribution of glycosyltransferases and the effect of various conditions on the enzyme levels. *J. Biol. Chem.*, **246**, 4919–4925.

Srikrishna, G., Wang, L., and Freeze, H.H. (1998) Fucose β-1-P-Ser is a new type of glycosylation: using antibodies to identify a novel structure in *Dictyostelium discoideum* and study multiple types of fucosylation during growth and development. *Glycobiology*, **8**, 799–811.

Stimson, E., Virji, M., Makepeace, K., Dall, A., Morris, H.R., Payne, G., Saunders, J.R., Jennings, M.P., Barker, S., Panico, M., and others. (1995) Meningococcal pilin: a glycoprotein substituted with digalactosyl 2, 4-diacetamido-2, 4, 6-trideoxyhexose. *Mol. Microbiol.*, **17**, 1201–1214.

Strahl-Bolsinger, S., Gentzsch, M., and Tanner, W. (1999) Protein *O*-mannosylation. *Biochim. Biophys. Acta*, **1426**, 297–307.

Strahm, A., Amado, R., and Neukom, H. (1981) Hydroxyproline-galactosides as protein-polysaccharide linkages in a water soluble arabinogalactan-peptide from wheat endosperm. *Phytochem.*, **20**, 1061–1063.

Suzuki, T., Kitajima, K., Emori, Y., Inoue, Y., and Inoue, S. (1997) Site specific de-*N*-glycosylation of diglycosylated ovalbumin in hen oviduct by endogenous peptide:*N*-glycanase as a quality control system for newly synthesized proteins. *Proc. Natl Acad. Sci. USA*, **94**, 6244–6249.

Suzuki, T., Park, H., Kitajima, K., and Lennarz, W.J. (1998) Peptides glycosylated in the endoplasmic reticulum of yeast are subsequently deglycosylated by a soluble peptide: *N*-glycanase activity. *J. Biol. Chem.*, **273**, 21526–21530.

Takeda, J., Miyata, T., Kawagoe, K., Iida, Y., Endo, Y., Fujita, T., Takahashi, M., Kitani, T., and Kinoshita, T. (1993) Deficiency of the GPI anchor caused by a somatic mutation of the PIG-A gene in paroxysmal nocturnal hemoglobinuria. *Cell*, **73**, 703–711.

Tang, J., Neidigh, J.L., Cooksey, R.C., and McClain, D.A. (2000) Transgenic mice with increased hexosamine flux specifically targeted to β-cells exhibit hyperinsulinemia and peripheral insulin resistance. *Diabetes*, **49**, 1492–1499.

Ten Hagen, K., G., Bedi, G.S., Tetaert, D., Kingsley, P.D., Hagen, F.K., Balys, M.M., Beres, T.M., Degand, P., and Tabak, L.A. (2001) Cloning and characterization of a ninth member of the UPD-GalNAc:polypeptide

N-acetylgalactosaminyltransferase family, ppGalNTase-T9. *J. Biol. Chem.*, **276**, 17395–17404.

Teng-umnuay, P., Morris, H.R., Dell, A., Panico, M., Paxton, T., and West, C.M. (1998) The cytoplasmic F-box binding protein SKP1 contains a novel pentasaccharide linked to hydroxyproline in *Dictyostelium*. *J. Biol. Chem.*, **273**, 18242–18249.

Teng-umnuay, P., van der Wel, H., and West, C.M. (1999) Identification of a UDP-GlcNAc:Skp1-hydroxyproline GlcNAc-transferase in the cytoplasm of *Dictyostelium*. *J. Biol. Chem.*, **274**, 36392–36402.

Thibault, P., Logan, S.M. Kelly, J.F., Brisson, J.-R., Ewing, C.P., Trust, T.J., and Guerry, P. (2001) Identification of the carbohydrate moieties and glycosylation motifs of *Campylobacter jejuni* flagellin. *J. Biol. Chem.*, **276**, 34862–34870.

Tomita, M. (1999) *Biochemica*l background of paroxysmal nocturnal hemoglobinuria. *Biochim. Biophys. Acta*, **1455**, 269–286.

Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, **3**, 97–130.

Vertel, B.M., Walters, L.M., Flay, N., Kearns, A.E., and Schwartz, N.B. (1993) Xylosylation is an endoplasmic reticulum to Golgi event. *J. Biol. Chem.* **268**, 11105–11112.

Viskupic, E., Cao, Y., Zhang, W., Cheng, C., De Paoli-Roach, A., and Roach, P.J. (1992) Rabbit skeletal muscle glycogenin. Molecular cloning and production of fully functional protein in *Escherichia coli*. *J. Biol. Chem.*, **267**, 25759–25763.

Wacker, M., Nita-Lazar, M., and Aebi, M. (2001) Pg1B, an oligosaccharyltransferase in the eubacterium *Campylobacter jejuni*? *Glycobiology*, **11**, 871 (abstract).

Wang, Y. and Spellman, M.W. (1998) Purification and characterization of a GDP-fucose polypeptide fucosyltransferase from Chinese hamster ovary cells. *J. Biol. Chem.*, **273**, 8112–8118.

Wang, Y., Shao, L., Shi, S., Harris, R.J., Spellman, M.W., Stanley, P., and Haltiwanger, R.S. (2001) Modification of epidermal growth factor-like repeats with *O*-fucose. Molecular cloning and expression of a novel GDP-fucose protein *O*-fucosyltransferase. *J. Biol. Chem.*, **276**, 40338–40345.

Weng, S. and Spiro, R.G. (1997) Demonstration of a peptide:*N*-glycosidase in the endoplasmic reticulum of rat liver. *Biochem. J.*, **322**, 655–661.

West, C.M., van der Wel, H., Kaplan, L., Morris, H.R., and Dell, A., (2001) Analysis of a microbial UDP-GlcNAc:hydroxyproline polypeptide GlcNAc-transferase that modifies Skp1 in the cytoplasm of *Dictyostelium*. *Glycobiology*, **11**, 883 (abstract).

Yamagishi, T., Matsuda, K., and Watanabe, Y. (1976) Characterization of the fragments obtained by enzymic and alkaline degradation of rice-bran proteoglycan. *Carbohydr. Res.*, **50**, 63–74.

Yan, Q. and Lennarz, W.J. (1999) Oligosaccharyltransferase: a complex multisubunit enzyme of the endoplasmic reticulum. *Biochem. Biophys. Res. Commun.*, **266**, 684–689.

Yang, L.L. and Haug, A. (1979) Purification and partial characterization of a procaryotic glycoprotein from the plasma membrane of *Thermoplasma acidophilum*. *Biochim. Biophys. Acta*, **556**, 265–277.

Yuen, C.-T., Chai, W., Loveless, R.W., Lawson, A.M., Margolis, R.U., and Feizi, T. (1997) Brain contains HNK-1 immunoreactive *O*-glycans of the sulfoglucuronyl lactosamine series that terminate in 2-linked or 2, 6-linked hexose (mannose). *J. Biol. Chem.*, **272**, 8924–8931.

Zeitler, R., Hochmuth, E., Deutzmann, R., and Sumper, M. (1998) Exchange of Ser-4 for Val, Leu or Asn in the sequon Asn-Ala-Ser does not prevent *N*-glycosylation of the cell surface glycoprotein from *Halobacterium halobium*. *Glycobiology*, **8**, 1157–1164.