Protein Glycosylation, an Overview

Elwira Lisowska, *Ludwik Hirszfeld Institute of Immunology and Experimental Therapy, Wroclaw, Poland*

Introductory article

Article Contents

Online posting date: 30th April 2008

Glycosylation is the most common posttranslational modification of proteins. It is a complex process involving many functional proteins and resulting in a great diversity of structures. Biological role of glycosylation and molecular and genetic basis of glycosylation disorders have recently been extensively explored. The goal of this short article is to signalize the variety of problems of this vast field of research.

Types of Glycosylation

Analysis of the SWISS-PROT database indicated that more than half of all proteins are glycosylated. There are various types of carbohydrate–protein linkage (**Table 1**), involving most known monosaccharides and functional groups of amino acid side chains. The protein-linked monosaccharides are usually extended (exceptions are GlcNAcβ-Ser/Thr and Manα-Trp) by attachment of other monosaccharides that gives multiple oligosaccharide strtuctures.

The most common protein-linked oligosaccharides are *N*-glycosidic chains (linked to Asn via GlcNAcβ) which exist in two major forms: (1) oligomannosidic (or 'high-Man') *N*-glycans with branched or linear oligomannosidic chains attached to both α-mannose residues of the core structure shown in **Table 1** and (2) complex chains containing 2–4 linear or branched antennae composed of one or more LacNAc (Galβ1-4GlcNAcβ) units and linked to α-mannose residues of the core structure. These antennae (and also β-mannose and Asn-linked GlcNAc) are 'decorated' with sialic acid, fucose and other monosaccharides, and also may contain phosphate, sulfate and *O*-acetyl residues that gives in effect a variety of structures. There are also various hybrid *N*-glycans combining features of both forms.

Mucin type *O*-glycans are attached to Ser/Thr by the GalNAcα residue, and the monosaccharides (and their linkages) attached to this GalNAc define various core structures (**Table 1**). Further elongation of these structures yields a large number of different *O*-glycans. The most common *O*-glycans are represented either by the core1 structure substituted with one or two sialic acid residues linked to Gal or/and GalNAc, or by more complex core2 *O*-glycans containing LacNAc-type chains. Generally, the structures (or arrays of structures) of protein-linked glycans are determined by the type of carbohydrate–protein linkage. However, the LacNAc-type chains present in *N*- and *O*-glycans can carry the same terminal nonreducing units, e.g. blood group ABH/Lewis (Le) antigens. **See also**: Blood Groups: Molecular Genetic Basis

In addition to other types of carbohydrate–protein linkage listed in **Table 1**, carbohydrates can be linked to proteins via phosphoester linkage. A distinct form of protein-linked oligosaccharides is a glycosylphosphatidylinositol (GPI, linked to the *C*-terminal group of the protein) which anchors some proteins in the cell membrane lipid bilayer.

The diverse glycan structures can be found in GlycosuiteDB (see http://www.glycosuite.com). This database contains 3238 unique glycan structures (Release 8.0, August 2005), and if known, the proteins to which the glycans are attached are described.

Multiple Proteins Involved in Glycosylation Process

Oligosaccharide chains are indirect products of genes, because many direct gene products (proteins) are involved in their biosynthesis. Protein glycosylation proceeds by the stepwise addition of monosaccharides, first to the protein and then to the growing oligosaccharide chain. Exceptions are GPI anchor, attached to protein in preassembled form, and *N*-glycans (with GlcNAcβ-Asn bond), where a preassembled dolichol-linked triglucosylated nanomannosidic chain is transferred to protein and after partial enzymatic trimming is further extended by addition of monosaccharides. Glycosyltransferases are strictly specific in respect to the donor, the acceptor residue and the type of linkage. Their catalytic efficiency is also more or less dependent on the location of the acceptor residue (type of carrier, underlying sugar chain). For each glycosidic bond, one or several glycosyltransferases exist, and many of them have been cloned. Other multiple enzymes are involved in

ELS subject area: Cell Biology

How to cite:
Lisowska, Elwira (April 2008) Protein Glycosylation, an Overview. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0006211.pub2

Table 1 Protein-linked saccharides: major types of carbohydrate–protein linkage

Glycans	Recognition motif	Occurence
N-glycans, *core structure*	–	Common in cellular and secreted proteins, wide phylogenetic distribution
Manα1,3 Manβ1,4GlcNAcαβ1,4GlcNAcβ-Asn	-N-X-S/T-	
Manα1,6	–	
Glcβ-N-Asn		Laminin, archaebacteria
Mucin type O-glycans, core structures		
Core1: Galαβ1,3GalNAcα-*O*-Ser/Thr Core2: Galβ1,3(GlcNAcβ1,6)GalNAcα-*O*-Ser/Thr Core3: GlcNAcβ1,3GalNAcα-*O*-Ser/Thr Core4: GlcNAcβ1,3(GlcNAcβ1,6)GalNAcα-*O*-Ser/Thr Core5: GalNAcα1,3GalNAcα-*O*-Ser/Thr Core6: GlcNAcβ1,6GalNAcα-*O*-Ser/Thr Core7: GalNAcα1,6GalNAcα-*O*-Ser/Thr Core8: Galα1,3GalNAcα-*O*-Ser/Thr		Mutiple chains (clusters) in mucin-type glycoproteins, glycophorins, leukosialins, single or few chains in some other proteins
O-mannosylation		
OligoMan-Manα-*O*-Ser/Thr		Yeast mannoprotein
O-fucosylation		Epidermal growth factor (EGF)-like domains: coagulation factors, Notch, thrombospondin, properdin, F-spondin
NeuAcα2,6Galβ1,4GlcNAcβ1,3Fucα-*O*-Ser/Thr Glcβ1,3Fucα-*O*-Ser/Thr	EGF module -C-X-X-G-G-S/T-C-	
Xylα1,3Xylα1,3Glcβ-*O*-Ser	-C-X-S-X-P-C-	EGF domains
O-GlcNAc glycosylation		
GlcNAcβ-*O*-Ser/Thr		Nuclear and cytosolic proteins (dynamic glycosylation)
Galα1,2Galβ-*O*-HyLys	Collagen repeats: -G-X-HyK-G-	Collagens, C1q complement, core-specific lectin
Xylβ-*O*-Ser	S-G/A	Proteoglycans
(Ara,Gal)-Araα or β-*O*-HyPro Galαβ-*O*-HyPro		Plant proteins
$(Glc)_n$-Glcα-*O*-Tyr		Glycogenin
C-mannosylation		
Manα1-*C*-Trp	-W-X-X-W, found in over 300 proteins	Ribonuclease 2, IL-12β, properdin, thrombospondin-1 F-spondin, hypertrehalosemic hormone, five complement components

Notes: Abbreviations of sugar names: Ara – arabinose, Fuc – fucose, Gal – galactose, GalNAc – *N*-acetylgalactosamine, Glc – glucose, GlcNAc – *N*-acetylglucosamine, Man – mannose, NeuAc – *N*-acetylneuraminic acid and Xyl – xylose.
Abbreviations of amino acid names (three and one-letter code): Ala (A) – alanine, Asn (N) – asparagine, Cys (C) – cysteine, Gly (G) – glycine, HyLys (HyK) – hydroxylysine, Hypro (HyP) – hydroxyproline, Lys (K) – lysine, Pro (P) – proline, Ser (S) – serine, Thr (T) – threonine, Trp (W) – tryptophan and Tyr (O) – tyrosine.

biosynthesis of 'activated' donors of sugars and other glycan-modifying residues (acetyl, sulfate). Moreover, various hydrolases participate in glycosylation process and catabolism of glycoproteins.

Formation of each glycosidic bond in glycoproteins or lipid-linked precursors requires a nucleotide- or lipid-linked sugar donor, acceptor (containing a monosaccharide or amino acid residue to be glycosylated) and respective transferase, all present in a proper cell compartment. Most glycosylation reactions take place in the lumen of the endoplasmic reticulum (ER) and Golgi compartments where transferases are located in the membranes in a defined order with their functional domains directed into the lumen. Nucleotide–sugar donors are synthesized in cytosol and are transported into the lumen of the Golgi by specific transporters that act as antiporters removing 'used' mononucleotides from the Golgi. It is an example indicating that the glycosylation process is dependent not only on enzymes, but also on many other proteins responsible for intracellular trafficking. **See also**: Protein Degradation and Turnover

The great variability of protein-linked glycan structures is dictated by tissue-specific regulation of genes encoding enzymes involved in glycosylation process, availability of the reaction components in the proper cell compartment, competition between glycosyltransferases for acceptor during glycan elongation, and finally by the structure of the glycosylated polypeptide and microenvironment of the growing glycan chain. The differences in glycan structures exist not only between different glycoproteins, but also between molecules of an individual glycoprotein produced by the same cells and between different glycosylation sites of one molecule, that results in various protein glycoforms. **See also**: Protein: Cotranslational and Posttranslational Modification in Organelles

Biological Role of Glycosylation

A rapidly growing evidence has indicated that protein-linked glycans not only protect proteins against proteolytic degradation and denaturation, but also interact directly with various carbohydrate-specific proteins and these interactions initiate or modulate many important biological events. Moreover, intramolecular interactions between glycan and peptide backbone affect the conformation and flexibility of both glycan and peptide that can modulate protein–protein interactions.

The glycosylation plays a role in 'quality control' of synthesized proteins. The triglucosylated oligomannosidic *N*-glycans, linked to nascent proteins in the ER, undergo trimming of two glucose residues and monoglucosylated glycans bind the lectin-like chaperonins, calnexin and calreticulin, which promote the protein folding. The correctly folded and assembled proteins proceed further to the Golgi compartments (also with the help of mannose-specific lectins, ERGIC-53 and VIP36) and misfolded proteins are reglucosylated and returned to the calnexin–calreticulin cycle or trimmed by the ER-mannosidase I and eliminated by ER-associated degradation (ERAD). Another intracellular process involving glycans is targeting the lysosomal hydrolases to lysosomes with participation of mannose-6-phosphate receptors which recognize Man-6P residues specifically present in oligomannosidic *N*-glycans of hydrolases. **See also**: Protein Folding and Chaperones

Glycans linked to the proteins exported to the plasma membrane participate in the overall architecture of the cell surface. Glycans of cell membrane-bound or soluble glycoproteins react with various carbohydrate-specific proteins that also exist in a membrane-bound or secreted form. The number of discovered and characterized mammalian lectins and information on their roles is rapidly increasing. Apart from mentioned intracellular lectins, there are E-, L- and P-selectins specific for sialyl-Le^x/Le^a and related structures, Siglec family (12 members identified in humans) of sialic acid binding lectins, galectins reactive with β-galactose-terminated glycans, endocytic macrophage and hepatocyte receptors and many others. Interactions of glycans with lectins play various roles in recirculation of cells, cell–cell and cell–matrix adhesion, cell growth and viability, cellular signalling, etc.

An interesting example of a role of glycosylation is glycodelin (Gd). It is the glycoprotein containing two occupied *N*-glycosylation sites. Gd produced by several reproductive tissues has the same polypeptide backbone but represents tissue-specific glycoforms and different biological activities. For example, GdA (from amniotic fluid), which contains sialylated complex *N*-glycans, inhibits sperm–oocyte interaction and shows immunosuppressive activity, while GdS (from seminal plasma) carries oligomannosidic chains and nonsialylated/highly fucosylated complex *N*-glycans and maintains an uncapacitated state in the spermatozoa.

A role of protein *O*-fucosylation (**Table 1**) has been shown in studies on the Notch signalling pathway. Highly conserved Notch cellular receptors have a central role in signal transducing pathways that influence many cell-fate decisions in metazoans and play a role in a variety of developmental processes in higher animals and humans. Fringe proteins have been known as modulators of activation of Notch receptors by Notch ligands. It was reported in 2000 that Fringe proteins have β1,3-*N*-acetylglucosaminyltransferase activity, initiating elongation of fucose residues *O*-linked to epidermal growth factor (EGF)-like repeats of Notch. This puzzling finding initiated further studies which showed various aspects of an important role of *O*-fucosylation in regulating Notch activity, and suggested a quality control function of protein *O*-fucosyltransferase-1 in ER.

Particular roles has *O*-glycosylation of Ser/Thr residues of nuclear and cytoplasmic proteins by β-GlcNAc. The *O*-GlcNAc glycosylation has dynamic character regulated by two enzymes: uridine diphosphate (UDP)-GlcNAc:polypeptide β-*N*-acetylglucosaminyltransferase and β-*N*-acetylglucosaminidase. It is an ubiquitous and essential protein modification. The diversity of proteins modified by *O*-GlcNAc and dynamic interplay between

O-glycosylation and *O*-phosphorylation imply importance of this type glycosylation in signal transduction. Moreover, it has a role in protein expression, degradation and trafficking and in the aetiology of diabetes and neurodegeneration.

Protein glycosylation has a great impact on the immune system. Almost all key molecules involved in the adaptive and innate immunity (receptors, immunoglobulins, cytokines, lectins, etc.) are glycosylated and some specific glycoforms are involved in recognition events. There are significant changes of glycosylation and transient appearing of certain structures on the cell surface, dependently on the cell differentiation or functional state. For example, the disialylated core1 *O*-glycans of resting T cells are shifted during T-cell activation into more complex branched core2 *O*-glycans (carriers of selectin ligands) that modulates intercellular interactions. Glycosylated protein antigens can induce humoral and cellular immune responses against oligosaccharidic, glycopeptidic and glycosylation-oriented peptidic epitopes. The existence of many natural anticarbohydrate antibodies in human sera and immune responses against 'foreign' oligosaccharide structures (of animal or plant origin) have great clinical importance in blood transfusion, transplantation and allergic diseases. Interactions of various viruses, bacteria and parasites with host cell glycans, and interactions of human and animal lectins with glycans of pathogens play a role in infection and host defence processes.

Protein Glycosylation and Disease

Protein glycosylation, which is dependent on so many factors, is altered in multiple diseases. Characteristic alterations are observed in rheumatoid arthritis (decreased galactosylation of immunoglobulin G, IgG), cystic fibrosis (undersialylation and overfucosylation of plasma membrane glycoconjugates), Wiskott–Aldrich syndrome (WAS) and acquired immunodeficiency syndrome (AIDS) (aberrant expression of core2 *O*-glycans on T lymphocytes) and in many other disorders. Altered glycosylation (loss of expression or excessive expression of certain structures) is a universal feature of cancer cells and affects glycoproteins and glycosphingolipids. Most typical alterations include increased branching of complex *N*-glycans, expression of truncated *O*-glycans (Thomsen–Friederich, Tn and sialyl-Tn antigens), overexpression of the sialylated Lewis stuctures (selectin ligands), differentially altered expression of the blood group ABH-related structures, alterations in sialylation. These disease-related changes in protein glycosylation are acquired and are likely to be the effect, and not the primary reason, of the disease. However, some of these altered structures have functional significance, and/or serve as diagnostic or prognostic markers of the disease, or as therapeutic targets.

Nevertheless, there is a group of diseases evidently caused by altered glycosylation. The rare congenital disorders of glycosylation (CDGs) occur due to mutations in genes encoding some key components of the glycosylation 'machinery' that most frequently results in severe clinical symptoms (malformation, psychomotor retardation, dysfunction of some organs and others). These diseases demonstrate the importance of protein glycosylation for the development and functions of the organism.

There are two major groups of CDGs. Group I includes various defects in the assembly of lipid-linked *N*-glycan precursor and its transfer to proteins in the ER that results in decreased number of protein-linked *N*-glycans. The 12 defects identified so far concern the following proteins: phosphomannomutase 2 (type Ia, most frequent, over 300 patients diagnosed worldwide), phosphomannose isomerase (Ib), dolichy(Dol)l-P-Glc:$Man_9GlcNAc_2$-PP-Dol α1,3-glucosyltransferase (Ic), Dol-P-Man:$Man_5GlcNAc_2$-PP-Dol α1,3-mannosyltransferase (Id), Dol-P-Man synthase 1 (Ie), protein faciltating utilization of Dol-P-Man (If), Dol-P-Man:$Man_7GlcNAc_2$-PP-Dol α1,6-mannosyltransferase (Ig), Dol-P-Glc:$Glc_1Man_9GlcNAc_2$-PP-Dol α1,3-glucosyltransferase (Ih), GDP-Man:$Man_1GlcNAc_2$-PP-Dol α1,3-mannosyltransferase (Ii), UDP-GlcNAc:Dol-P GlcNAc-1P transferase (Ij), GDP-Man:$GlcNAc_2$-PP-Dol β1,4-mannosyltransferase (Ik), Dol-P-Man:$Man_{6\ or\ 8}GlcNAc_2$-PP-Dol α1,2-mannosyltransferase (IL).

In CDG group II defects affect the processing of protein-linked *N*-glycans and also biosynthesis of *O*-glycans. It may concern a deficient function of enzymes: β1, 2-GlcNAc-transferase II which initiates elongation of α1,6-Man-linked antenna (IIa), α1,2-glucosidase I removing the first glucose residue from triglucosylated *N*-glycan precursor (IIb) and β1,4-galactosyltransferase (IId). Two types of CDG-II are connected with mutations in genes encoding nucleotide–sugar transporters: GDP-fucose transporter (IIc, known as leukocyte adhesion deficiency, LAD-II) and CMP-sialic acid transporter (IIf). The LAD-II patients, due to lack of fucosylation, do not express the blood group ABH epitopes ('Bombay' phenotype) and selectin ligands. Recently, new types of CDG-II were identified related to defects of genes encoding components of the conserved oligomeric Golgi complex (COG). The COG complex is an eight-subunit (Cog1-8) peripheral Golgi protein required for normal intracellular trafficking and activity of multiple proteins involved in glycosylation machinery. Three types of defects were identified so far, concerning Cog7 (CDG-IIe), Cog1 (CDG-IIg) and Cog8 (CDG-IIh). These data show that a defect in one component destabilizes the complex. **See also**: Cell Adhesion Molecules and Human Disorders

In the congenital dyserythropoietic anaemia type II (CDA-II, known also as HEMPAS, with heterogeneous genetic background) defective *N*- and *O*-glycosylation on erythroid lineage cells is observed. However, the lack of linkage of genetically determined alterations with the activity of proteins involved in glycan synthesis suggests that hypoglycosylation is not the primary defect but a consequence of the dyserythropoiesis. Galactosaemia refers to a group of inherited diseases caused by defects in genes encoding enzymes of galactose metabolism and manifested

by hypogalactosylation of glycoproteins and glycosphingolipids and accumulation of toxic galactose metabolites. Some congenital muscular dystrophies are associated with genetically determined defects in glycosylation of α-dystroglycan.

Normal functions of the organism also depend on the proper catabolism of glycoproteins. There are several 'lysosomal storage diseases' resulting from deficiency of one of lysosomal glycosidases and accumulation of respective undegraded glycans. Deficient GlcNAc-phosphotransferase involved in phosphorylation of mannose residue on lysosomal hydrolases is the reason of the I-cell disease. Hydrolases lacking Man-6P residue are directed to plasma instead to lysosomes and undegraded cellular components accumulate in lysosomes of fibroblasts and macrophages. **See also**: Tay–Sachs Disease

The examples (listed here and elsewhere) of glycosylation-related disorders show that protein glycosylation and catabolism of glycoproteins have profound patophysiological significance, not fully understood yet. Better understanding of these processes can help in finding new therapeutic approaches for treatment of these diseases. **See also**: Glycoproteins; Glycosylation and Disease; Lysosomal Transport Disorders

Further Reading

Davies GJ, Gloster TM and Henrissat B (2005) Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology* **15**: 637–645.

Fraser-Reid B, Tatsuta K and Thiem J (eds) (2001) *Glycosciences*. Heidelberg: Springer.

Freeze HH and Aebi M (2005) Altered glycan structures: the molecular basis of congenital disorders of glycosylation. *Current Opinon in Structural Biology* **15**: 490–498.

Garcia-Vallejo JJ, Gringhuis SI, van Dijk W and van Die I (2006) Gene expression analysis of glycosylation-related genes by real-time polymerase chain reaction. *Methods in Molecular Biology* **347**: 187–209.

Hart GW, Housley MP and Slawson C (2007) Cycling of β-*N*-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**: 1017–1022.

Helenius A and Aebi M (2004) Roles of *N*-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry* **73**: 1019–1049.

Lisowska E (2002) The role of glycosylation in protein antigenic properties. *Cellular and Molecular Life Sciences* **59**: 445–455.

Rampal R, Luther KB and Haltiwanger RS (2007) Notch signaling in normal and disease states: possible therapies related to glycosylation. *Current Molecular Medicine* **7**: 427–445.

Seppälä M, Koistinen H, Koistinen R, Chiu PC and Yeung WS (2007) Glycosylation-related actions of glycodelin: gamete, cumulus cell, immune cell and clinical associations. *Human Reproduction Update* **13**: 275–287.

Spiro R (2002) Protein glycosylation: nature, distribution, enzymatic formation and disease implications of glycopeptide bonds. *Glycobiology* **12**: 43R–56R.

Web Links

GlycosuiteDB http://www.glycosuite.com
Entrez PubMed http://nslij-genetics.org./search_pubmed.html