Glycoproteins

Tony Merry, *University of Manchester, Manchester, UK*

Sviatlana Astrautsova, *Grodno State Medical University, Grodno, Belarus*

Based in part on the previous version of this Encyclopedia of Life Sciences (ELS) article, ''Glycoproteins'' by ''Terry D Butters''.

Advanced article

Online posting date: 15th September 2010

Glycoproteins are proteins that contain covalently bound oligosaccharides (glycans). Although there are structural features of *N*- and *O*-linked glycans common to all eukaryotes, species-specific glycosylation reactions create a wide diversity of structures. This is particularly relevant to cell surface glycoproteins although intracellular and even nuclear proteins may be glycosylated. The factors controlling glycosylation are complex and include the protein sequence and structure, specificity of relevant transferase enzymes, availability of donor sugars and other environmental factors. These may all interact so it is not surprising that there is great variety in glycosylation even on the same protein. Glycosylation of proteins is important in such areas as development and interaction with pathogens and provides a means of modifying the function of proteins that is not directly dependent on their deoxyribonucleic acid (DNA) template. Interest in the field is growing and advances in analytical technology now make the field accessible to a wider community.

Introduction

In all eukaryotic cells, one of the most important and common posttranslational modifications to proteins is the covalent addition of carbohydrate (Apweiler *et al*., 1999). These glycoproteins are an abundant class of molecules that are found in cellular membranes, particularly the plasma membrane, and in extracellular secretions such as plasma. The carbohydrate moiety, or glycan, is responsible for many important physiochemical properties of the protein and plays a role in cell–cell recognition, fertilisation and development. Changes in glycoprotein glycans have been observed for several pathological conditions in humans, including cancer, rheumatoid arthritis and inflammation. **See also**: Protein Glycosylation, an Overview; Proteins: Postsynthetic Modification – Function and Physical Analysis

Types of Glycoproteins and their Occurrence

Two major types of covalent addition of glycan are found. These involve the modification to amino acid side-chains: *N*-glycosylation of asparagine amino groups and *O*-glycosylation of serine or threonine hydroxyl groups (**Figure 1**). *N*-Linked glycans can be further divided into three major classes; the complex type containing *N*-acetylglucosamine, mannose, galactose, fucose and sialic acid; the oligomannose type containing *N*-acetylglucosamine and

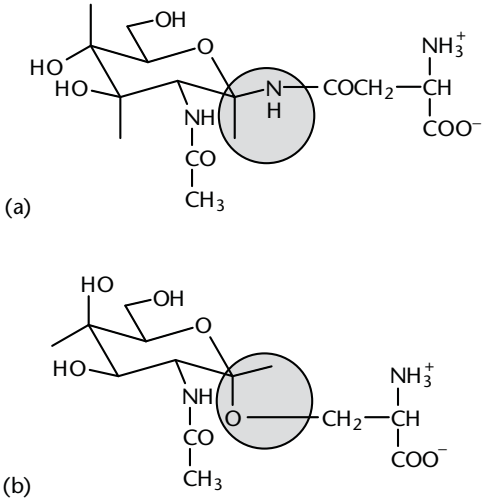

Figure 1 The monosaccharide linkages to amino acids that form *N*- and *O*-linked oligosaccharides. The shaded areas show the atoms involved, the nitrogen of asparagine amino groups and the oxygen of serine hydroxyl groups. (a) *N*-linked D-GlcNAc β1-Asn and (b) *O*-linked D-GalNAc α1-Ser/Thr.

ELS subject area: Biochemistry

How to cite:
Merry, Tony; and Astrautsova, Sviatlana (September 2010) Glycoproteins. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0000705.pub2

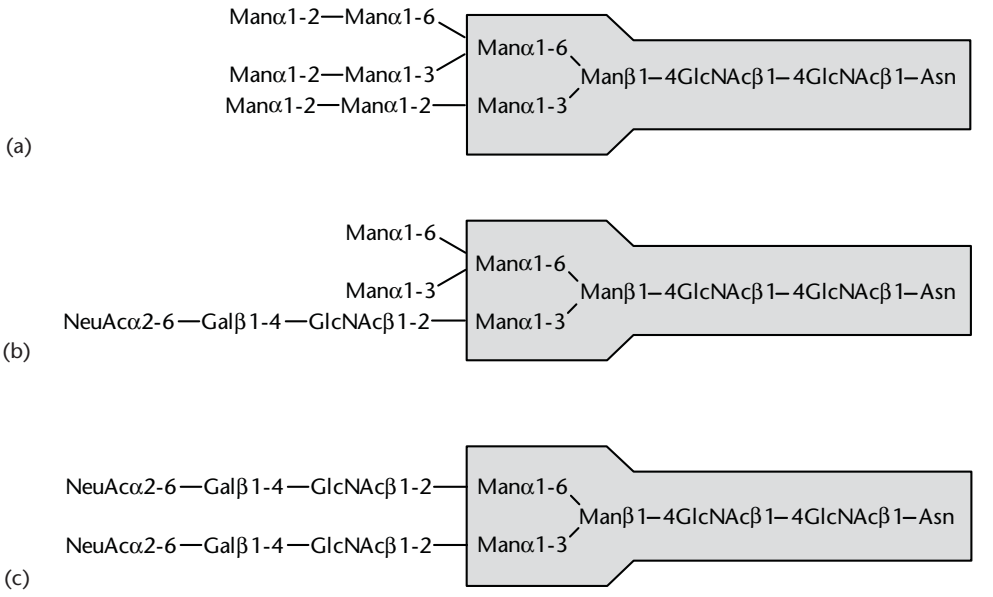

Figure 2 Different types of *N*-linked oligosaccharide structures: (a) oligomannose type, (b) hybrid type and (c) complex type. The shaded area shows the common feature to all classes of *N*-links. This is also conserved in all eukaryotes. Asn, asparagine; GlcNAc, *N*-acetylglucosamine; Man, mannose; Gal, galactose and NeuAc, *N*-acetylneuraminic acid.

mannose only; and the hybrid type that has features common to both complex and oligomannose chains (**Figure 2**). Many glycoproteins will contain both *N*- and *O*-linked glycans and have more than one glycan chain per molecule. Not every asparagine or serine and threonine residue is occupied by carbohydrate. *N*-Linked glycosylation requires a recognition motif, the Asn-X-Ser(Thr) triplet, where the X residue can be any amino acid except proline. In addition, there are constraints by the three-dimensional structure for accessibility to processing enzymes which mean that occupancy of sites may vary. The requirements for *O*-linked glycosylation are less restrictive and no consensus sequence has been identified by glycoprotein analysis. However, studies on the class of enzymes responsible for the transfer of the initial monosaccharide, *N*-acetyl galactosamine (GalNAc) to serine or threonine residues, show that they have precise though varied specificities. Currently, at least 15 members of the family have been well characterised (Tian and Ten Hagen, 2009), and genome analysis suggests there may be more. All isoforms are type II transmembrane proteins with a short *N*-terminal cytoplasmic region, a hydrophobic transmembrane region, a variable length stem region and a conserved catalytic region. Biochemical analyses have revealed unique as well as overlapping substrate preferences amongst isoforms. Some the GalNAc:polypeptide *N*-acetylgalactosaminyltransferases will only transfer GalNAc when other ser/thr residues are already glycosylated, and there appears to be a number of related activities which depend on the tissue in which they are expressed which determine the final pattern of sites of *O*-glycosylation. Several adjacent serine of threonine residues may be glycosylated as is common in the class of heavily *O*-glycosylated proteins known as mucins. Both *N*- and *O*-linked glycans (glycans) comprise similar monosaccharides that can be linked together to form a similar sequence (**Figure 2** and **Figure 3**). This is particularly the case in terminal structures which may be the same for *N*- and *O*-linked glycans and even glycolipids. The products of any particular cell type may therefore carry the same determinant in many different configurations. **See also**: Proteins: Postsynthetic Modification – Function and Physical Analysis; Protein Glycosylation, an Overview; Protein Structure: Unusual Covalent Bonds

Glycoforms

The nature and variability of monosaccharide linkages and sequences in glycoproteins give rise to structural heterogeneity. Although the amino acid sequence of the protein remains constant and is determined by the messenger ribonucleic acid (mRNA) sequence encoding for the protein, the population of glycan structures is variable and not directly determined by RNA/deoxyribonucleic acid (DNA). A number of different factors including metabolic and environmental effects may play a part in determining the precise pattern of the structures found on a particular glycoprotein. Notwithstanding this, the range of structures on any given protein may be remarkably constant and only vary in disease, thus giving the possibility of using changes in glycosylation profile as a disease marker in some cases.

Individual protein molecules will carry a unique set of glycan structures which may modulate protein function

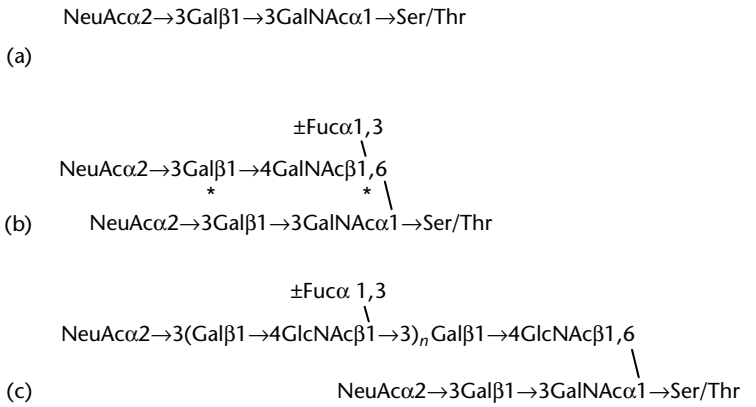

Figure 3 Different types of *O*-linked oligosaccharide structures. Note that these range in complexity from the simple linear structures to repeating and branched types of structure. (a) Core-1-type *O*-link found in red blood cells. (b) GlyCAM-1, the core-2 sulfated (* at 6 position of galactose of *N*-acetylglucosamine) *O*-linked oligosaccharide from endothelial cells. (c) PSGL-1, the core-2 polylactosamine *O*-linked oligosaccharide from neutrophils.

(Arnold *et al*., 2007) and this subset of the population has been termed a glycoform (Rademacher *et al*., 1988). Proteins that are composed of several *N*- and *O*-linked glycans have several glycoforms; those that contain only one *N*-linked glycan may have just a few. For example, human immunodeficiency virus (HIV) gp120, which has 20–25 potential *N*-linked sites, contains over 100 different glycan structures, whereas ribonuclease B, which has a single occupied *N*-linked site, consequently has fewer (~ 5) different structures. The generation of glycoforms is further complicated by cell- and tissue-specific glycosylation, so the potential number of glycoforms generated can be extremely large. Does this have a functional significance for activities associated with the glycoprotein? There are some notable instances where glycoslyation is directly related to function, for example in the trafficking of cell in the immune system; however, it is almost certainly not the case that glycosylation always has a role in protein function. Often, it may be that it is only of importance in a particular context but the type of glycosylation may also exist in many more instances where a defined function cannot be given. On the other hand, there are cases where the differences in glycosylation may indeed have functional significance as in the case of therapeutic antibodies, for example a particular glycoform, one which contains a core fucose, may have significant advantages over others (Jefferis, 2007). **See also**: Protein Glycosylation, an Overview

A general role for glycosylation is in assisting the folding of proteins where a complex interplay between the peptide and glycan portions of the glycoprotein results in a correctly folded molecule (Shental-Bechor and Levy, 2009). This information has only been possible to obtain using accurate methods for glycoprotein analysis (described in the following text) and has allowed computer-generated simulation of the glycan molecular dynamics in solution. In general, the carbohydrate moiety of glycoprotein does not give good diffraction data on X-ray crystallography. Most of the carbohydrate side-chain is flexible and does not adopt a single three-dimensional structure. However, close examination of diffraction data can indicate the positions of sugar residues closest to the polypeptide chain. Care must, however, be taken in interpretation, and many structures in the Protein Data Bank (PDB) structural database have incorrect assignments. For this reason, images of the complete three-dimensional structure of glycoproteins are generally obtained by combining X-ray crystallographic data of the protein and the modelled three-dimensional structure and motion of the glycan. The thermodynamic constraints on the torsion angles of the various linkages may be calculated to give a prediction of the range of possible structures present (**Figure 4**). We, therefore, now have a much clearer idea of what the entire glycoprotein looks like (Petrescu *et al*., 2006). The glycan portion of the glycoprotein occupies a significantly large space, and it is therefore not surprising that the type or presence of glycan can create spatial conflicts with the protein. Consequently, the glycan can mask protein domains and prevent the association with other proteins, for example in an enzyme–substrate binding event. The generation of a heterogeneous population of glycoproteins may be one way the cell can control biological activity: in the next section, we view the variety of *N*-linked glycan structures and understand how these glycoforms are generated. **See also**: Enzyme Activity: Control; Monosaccharides; Protein Stability

N-Linked Glycans: Structure and Biosynthesis

Glycan transfer, trimming and protein folding in the endoplasmic reticulum

N-Linked glycosylation of proteins is initiated by the transfer of Glc3Man9GlcNAc2 glycan from a dolichol lipid intermediate to the nascent peptide chain in the

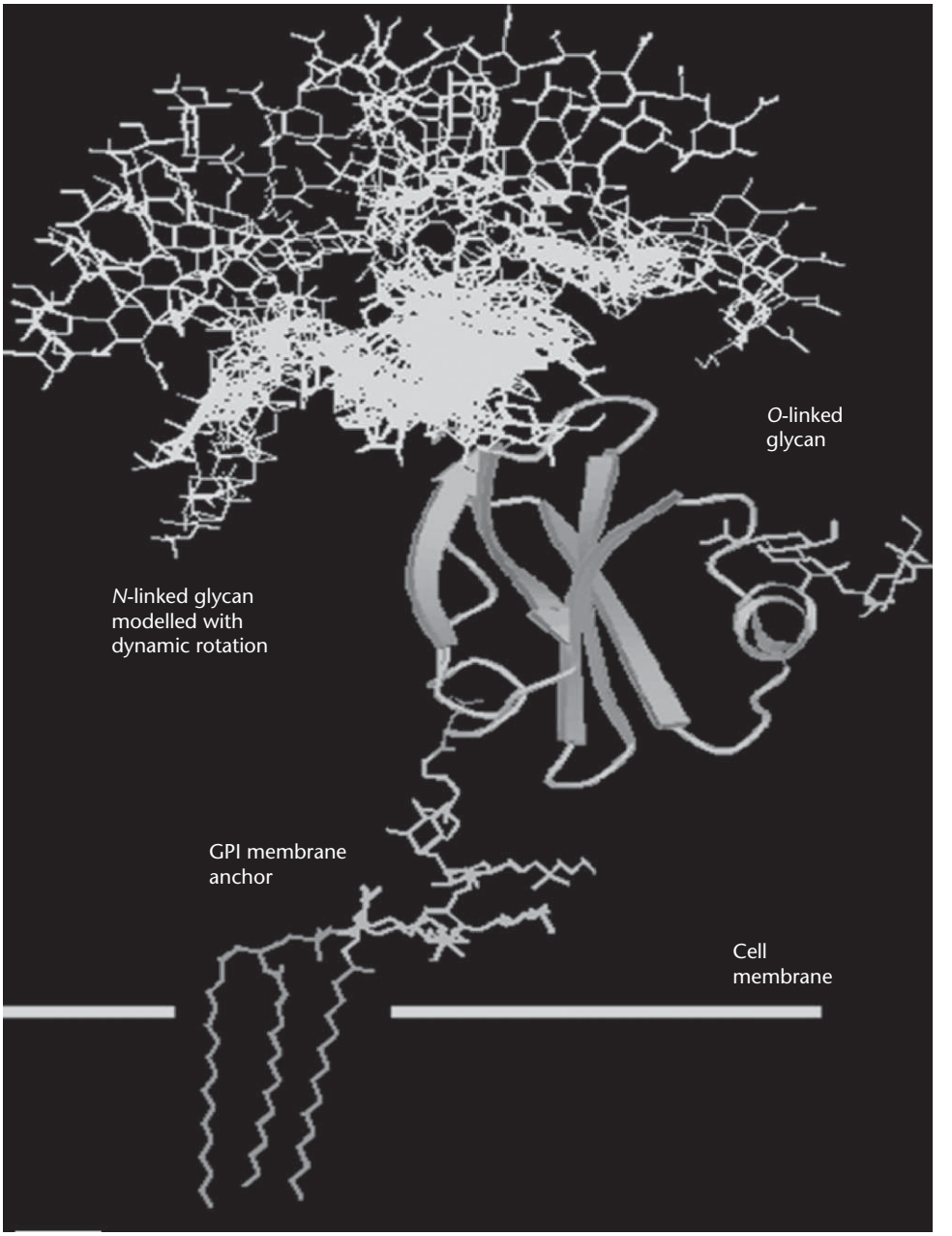

Figure 4 Molecular Model of Human erythrocyte CD59 showing three types of glycosylation – *N*-linked glycans, *O*-linked glycans and GPI anchor. Molecular dynamics simulation of the possible configurations resulting from the allowed torsion angles calculated for the linkage to the peptide and glycosidic linkages. Model courtesy of Mark Wormald, Oxford Glycobiology Institute.

lumen of the endoplasmic reticulum (ER) (**Figure 5**). A comparison of the proteins thought to mediate glycan transfer in yeast and animal cells has revealed a high degree of organisational conservation that suggests a close temporal and spatial relationship exists between polypeptide synthesis, translocation and *N*-glycosylation. Essential subunits of oligosaccharyltransferase, which transfers the glycan from the dolichol lipid to the nascent protein, are members of a much larger protein complex composed of constitutive ribosomal proteins, such as ribophorin. The reaction mechanism of oligosaccharyltransferase has not been completely determined, but some data suggest that the 'Asx-turn' motif enables the carbonyl residue of the asparagine to interact with the threonine/serine NH and hydroxyl to promote its protonation. In studies on the yeast enzyme, the roles of each of nine subunits have been examined, and they all contribute to overall activity (Yan and Lennarz, 2005). In general, the conformation of a glycan is not greatly changed by its attachment to the protein (Wormald *et al.*, 2002).

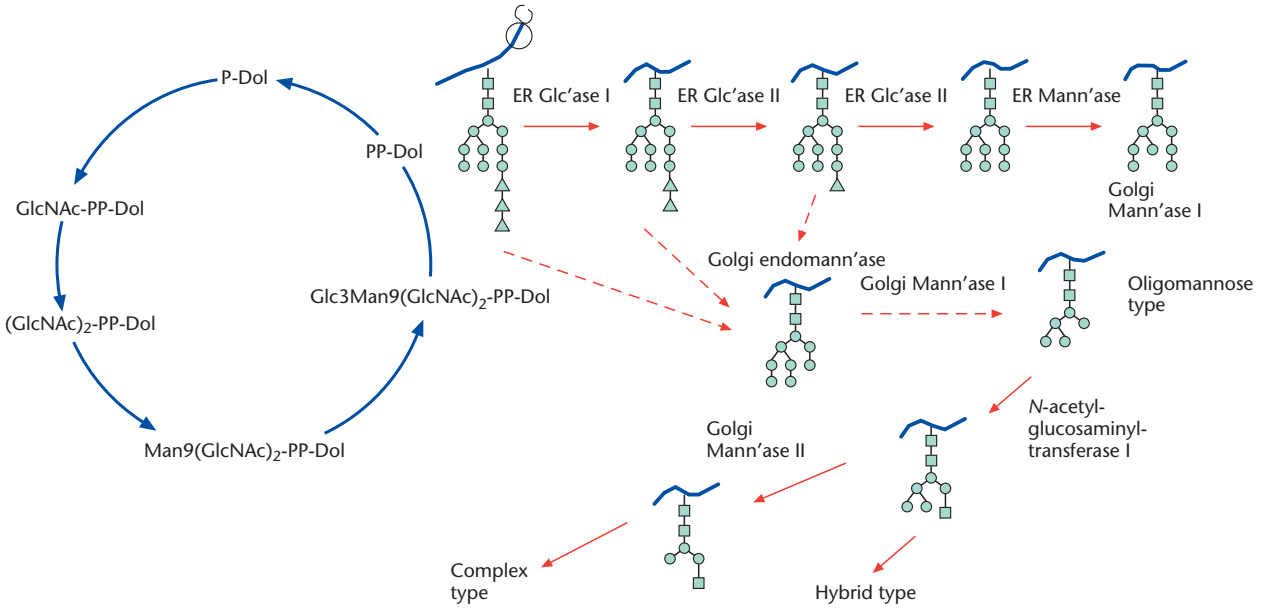

Figure 5 The early events in the glycosylation pathway for *N*-linked oligosaccharides. Dol, dolichol lipid; P, phosphate; Glc, glucose; Man, mannose; GlcNAc, *N*-acetylglucosamine; Glc'ase, glucosidase; Mann'ase, mannosidase and ER, endoplasmic reticulum.

The roles of addition of glycosylated precursors have also been examined, and it has become evident that the modifications that take place in the ER reflect a spectrum of functions related to glycoprotein folding, quality control, sorting, degradation and secretion. The confirmation of the Glc3Man unit in particular may be to balance the chaperone-assisted and chaperone-independent pathways (Mackeen *et al*., 2009). The glycans promote folding not only directly by stabilising polypeptide structures, but also indirectly by serving as recognition 'tags' that allow glycoproteins to interact with a variety of lectins, glycosidases and glycosyltranferases. Some of these (such as glucosidases I and II, calnexin and calreticulin) have a central role in folding and retention, whereas others (such as α-mannosidases and EDEM (ER-degradation enhancing mannosidase-like protein)) target nonsalvageable glycoproteins for ER-associated degradation. Each residue in the core glycan and each step in the modification programme have significance for the fate of newly synthesised glycoproteins (Helenius and Aebi, 2004).

The relative importance of glucose as part of the glycan in assisting with or signalling transfer is species-dependent. In mammalian cells, only those glycans which contain glucose residues are transferred to protein, even in mutant cell lines that synthesise truncated Glc3Man5GlcNAc2 glycans. The recognition and binding of outer-arm glucose residues have been proposed to induce conformational changes in the active site of oligosaccharyltransferase, influencing the association constant of the peptide substrate. The inability of trypanosomatid protozoa to synthesise dolichol-*P*-glucose does not prevent glycosylation but results in the transfer of nonglucosylated glycans. No detectable α-glucosidase I is present in this parasite, but α-glucosidase II is required to participate in the chaperone-mediated reglucosylation pathway. **See also**: Cell Membrane Features; Protein Folding *In Vivo*

Data obtained from independent groups, both studying the initial temporal events that affect glycosylation, support an additional role for glucosidase II activity in the ER. First, classical work from Parodi's group showed that a constitutively active uridine diphosphate (UDP)-glucose:glycoprotein glucosyltransferase is present in microsomes derived from a number of diverse species, including plants, insects, trypanosomatids, fungi and animals (Parodi, 1999). This glucosyltransferase is present as a soluble enzyme in the ER lumen and has been isolated to apparent homogeneity from both rat liver microsomes and the fission yeast *Schizosaccharomyces pombe*. The ability to transiently reglucosylate only denatured glycoproteins suggests that, *in vivo*, only unfolded, partially folded and misfolded glycoproteins are reglucosylated and that this step is prevented by the adoption of the correct conformation – a process that hides the protein determinants (possibly hydrophobic amino acids) from the glucosyltransferase.

A second mechanism that involves transient interactions of newly synthesised glycoproteins in the ER requires molecular chaperons that facilitate the folding and subunit assembly of nascent proteins. These interactions, in particular with integral membrane proteins such as calnexin, or soluble calreticulin, ensure that only properly folded and assembled proteins are exported from the ER. Calnexin shows great specificity for glycoproteins bearing *N*-linked glycans, as tunicamycin prevents glycoprotein–calnexin complexes and pretreatment with α-glucosidase inhibitors,

castanospermine and deoxynojirimycin also blocks these interactions. Calnexin has been shown to have a lectin-like ability to bind to mono-glucosylated glycans as an initial step. This suggests that recognition of unfolded regions of newly synthesised proteins, or transmembrane regions of membrane-bound protein, is also required for calnexin interactions that prevent premature export of incompletely folded proteins from the ER. **See also**: Chaperones, Chaperonin and Heat-Shock Proteins; Lectins; Protein Folding *In Vivo*; Protein Secretory Pathways

One model (Helenius and Aebi, 2004) shows how these two pieces of apparently unrelated data may be involved to provide a general, evolutionary conserved mechanism for preserving the integrity of the protein-folding pathways in all eukaryotes. Newly synthesised proteins are glycosylated and the first α2-linked glucose residue is removed by α-glucosidase I. Rapid removal (half-life $\sim$5 min) of the exposed α3-linked glucose by α-glucosidase II, creates a mono-glucosylated glycan that is also recognised by α-glucosidase II (hydrolysis of this glucose being significantly slower) and calnexin. If transfer to calnexin takes place, via the recognition signal and other protein–protein interactions, the complexed glycoprotein is held in transit long enough to assume a properly folded conformation. The normal route taken for folded proteins would be the dissociation of the complex, removal of the α3-glucose by α-glucosidase II and ultimate export from the cell. For partially folded proteins, those dissociated from calnexin before a correctly folded state has been generated, or those acted on by glucosidases too rapidly for the interaction with calnexin to take place, glycans are reglucosylated and the cycle of events is repeated. If proteins are irreversibly misfolded, the cycle is joined by repetitive reglucosylation, calnexin binding and de-glucosylation reactions. The mechanism prevents protein aggregates forming until their exclusion from the ER to the lysosome. **See also**: Chaperonins; Endoplasmic Reticulum to Golgi Transport: Methods; Lysosomal Degradation of Proteins; Protein Folding: Overview of Pathways

This model, and in particular the recognition of an exquisite processing intermediate glycan by calnexin, has been suggested to confer evolutionary significance to the *en bloc* mode of glycosylation that originated in early eukaryotes. The involvement of α-glucosidase II is pivotal to the success of such quality control mechanisms because the enzyme must act at three times and places during the de-glucosylation and reglucosylation cycle. The solution of the nuclear magnetic resonance (NMR) structure of the glucosylated *N*-glycan has allowed a model to be predicted whereby calnexin and α-glucosidase II approach their common substrate from different sides of the molecule. This explains why α-glucosidase II can hydrolyse the proximal glucose residue from the glycan to promote release of the glycoprotein from the complex.

The properly folded glycoprotein, now containing a truncated glycan, is carried by vesicular transport from the ER to the Golgi apparatus. **See also**: Endoplasmic Reticulum to Golgi Transport: Methods

Glycan processing in the Golgi complex

Glycan processing with glycosidases proceeds through the early elements of the Golgi complex, after the removal of a single mannose residue in the ER. The removal of a further three mannose residues by Golgi α-mannosidase I forms the Man5GlcNAc2 isomer. After the action of *N*-acetylglucosaminyltransferase I (GlcNAcT-I), two terminal mannose residues are excised by Golgi α-mannosidase II.

There is an endo-α-mannosidase in the Golgi that acts to de-glucosylate partially trimmed or incompletely assembled glycans (**Figure 5**). This accounts for the lack of complete inhibition of processing using α-glucosidase I inhibitors and contributes to complex glycan formation in glucosidase-deficient cells. Most mammalian cells contain some measurable endomannosidase activity, but *in vitro* analyses have failed to detect activity in yeast cell extracts, mung bean seedlings and *Drosophila* cells.

The *medial* Golgi apparatus is the cellular site of action for many glycosyltransferases that decide branching patterns of *N*-glycans in eukaryotic cells. It is in this compartment, and the later *trans* elements of the Golgi that glycosyltransferase enzymes regulate synthesis of diverse structures that are found in eukaryotes. Consequently, species-specific glycosylation may depend on the presence or absence of the gene and/or its tissue-specific regulation.

Several factors influence the precise modifications carried out in the medial Golgi. The individual glycosyltransferases may have a number of isoforms which have precise specificities, for example the activity of members of the GlNAc transferase families I–V which act on different arms of the glycan. There is also undoubtedly some influence of the three-dimensional protein structure in determining the availability of the glycan chain for further processing. In this way, glycoforms of the glycoprotein generally have different types of glycosylation at different sites within the protein. When produced in another cell types, the glycosylation at any given site may be different and leads to cell type-specific glycosylation.

The different types of *N*-glycans found in eukaryotic species and their biosynthesis are discussed in the following section. **See also**: Plant Golgi Apparatus

Glycosyltransferase enzymes and genes

A large number of genes which are responsible for addition of sugar residues to glycans (glycosyl transferases) have now been characterised in a large number of species. They clearly fall into families and their evolution may be traced. A surprisingly large fraction of many genomes is dedicated to glycosyl transferase genes and many are highly conserved through evolution hinting at their importance. It has been estimated that 2–3% of the human genome is involved with glycosylation enzymes.

There have also been a large number of knockout mutants of transferase genes made in mice, but it is sometimes difficult to interpret the studies on these in terms of the functional significance of a particular type of

glycosylation. Some are lethal mutants but many appear to have little effect, and when glycosylation of proteins in these mutants is studied, there seems to be only minor changes. This has been explained by the presence of redundancy in the biosynthetic processes and that backup systems exist. When multiple knockouts are made, effects are much more pronounced.

It has been shown that there are some defects in humans where genes involved in glycosylation are defective and these are collectively known as congenital defects in glycosylation. Although rare, they frequently affect children and may be serious. Studies can identify the defect, and in certain cases this may be corrected by additions to the diet. **See also**: Protein: Cotranslational and Posttranslational Modification in Organelles

Species-specific *N*-glycosylation

Prokaryotic glycosylation

Protein *N*-glycosylation is not a modification exclusive to eukaryotes: a number of bacterial species that are able to glycosylate proteins have been identified. A complete chemical structural characterisation of a surface layer (S-layer) glycoprotein from *Halobacterium halobium*, an extreme halophile of the archaebacteria, has revealed the presence of glucose in β linkage to an asparagine residue as part of an Asn-X-Ser/Thr. This class of bacteria also displays a cell surface repeating unit sulfated saccharide, similar in composition to eukaryotic glycosaminoglycans. This structure, together with *N*-glucosyl units, compensates for the lack of peptidoglycan typical of cell walls of eubacteria. The β-GlcAsn modification is not restricted to the archaebacteria and has been identified on laminin isolated from mouse tumours. Is there a novel *N*-glycosylation pathway shared by both archaebacteria and mammalian species that has evolved independently of β-GlcNAcAsn glycosylation? A phylogenetic analysis shows that the archaebacterial transcription factor, TATA-binding protein (TBP), has over 40% sequence homology to the human protein and is able to interact with human proteins that regulate transcription. These data suggest that the transcriptional apparatus must have arisen in a common lineage that had already diversified from the eubacteria (all other prokaryotes). The Archaea may share an ancestral lineage with eukaryotes that explains the novel *N*-glycosylation data. **See also**: Glycosaminoglycans: Structure and Biological Functions; Peptidoglycan

Eukaryotic glycosylation

A fundamental characteristic of eukaryotes is the presence of membrane-bound compartments and membrane-transport pathways in which the Golgi apparatus plays a central role in the processing and secretion of glycoproteins. The parasitic protozoan *Giardia lamblia*, a primitive eukaryote, does not require an identifiable Golgi complex to secrete simple, nonglycosylated proteins, but induction of both Golgi enzyme activities and Golgi structure occurs during encystations of trophozoites. As the parasite's ribosomal RNA shares more sequence homology with prokaryotes than any other eukaryote, yet it can developmentally regulate the biogenesis of biosynthetically competent complexes similar in morphology to higher eukaryotes, this species may represent an evolutionary 'bridge' between the prokaryotes and eukaryotes. The ability to parasitise emerging life forms may have been the selective pressure required to modify single-gene products in this unique organism. Comparative studies on organisms such as *Drosophila* and *Caenorhabditis elegans* have also thrown light on the development of glycosylation processes (ten Hagen *et al*., 2009). **See also**: Eukaryotes and Multicells: Origin

Yeast glycosylation

Analysis of the structural features of *N*-glycans from a range of eukaryotic species reveals Golgi-mediated glycosyltransferase-dependent changes. Yeast glycans are typically of the oligomannosidic type and are terminated by further mannose chains to form a homologous family of Man9–14GlcNAc2 cores. Some are more extensively processed by the addition of an outer chain consisting of an α1,6-linked mannose backbone with an α1,2- and α1,3-linked side-chain, all of which may consist of 50–200 residues per chain. The conclusion from studies is that yeasts must have diversified their basic core oligomannose structures by the action of additional and often unique sets of glycosyltransferases. To synthesise the most complex mannans, for example perhaps 10–12 mannosyltransferases may be involved (Gemmill and Trimble, 1999). **See also**: Gene Expression in Yeast

Insect glycosylation

The majority of glycoproteins analysed from insect tissues, for example cultured mosquito and lepidopteran cells, and *Drosophila* larvae, are of the oligomannosidic type. Fucosylation of the innermost GlcNAc residue has been observed in glycoproteins derived from *Drosophila* and in *Apis* (honeybee), and in the latter case are often bi-fucosylated in α1,3- and α1,6-linkages. The use of lepidopteran cell lines for the expression of heterologous proteins using baculovirus vectors has aroused further interest in the glycosylation events in the cells of insects. Using *in vitro* glycosyltransferase assays, the presence of GlcNAc-transferase I and II activities has been measured in lepidopteran cell extracts (Vadaie and Jarvis, 2004). Several lines of evidence have been presented to demonstrate the developmentally regulated expression of sialylated proteins in *Drosophila* embryos and these are used for model organisms in unravelling the complexities of glycobiology across many species (ten Hagen *et al*., 2009). **See also**: Baculoviruses; *Drosophila* Evolutionary Genetics

Protozoan parasite glycosylation

Sialic acid is incorporated into trypanosomatid protozoan glycoproteins, not via a sialyltransferase (as these parasites

are unable to synthesise cytidine monophosphate (CMP)-sialic acid from *N*-acetylmannosamine) but by a surface expressed trans-sialidase. Addition of sialic acid is found on complex-type *N*-linked glycans that also contain some αl,3-linked galactose residues and branched polylactosamine. In addition to complex-type glycans, oligomannose structures similar to those from vertebrates and invertebrates have been detected and unusual galactofuranose, rhamnose, ribose and xylose monosaccharides substituents have been reported. The complex life cycles of parasites, particularly those that have both mammalian and insect hosts, are characterised by a number of biochemical processes, of which *N*-glycosylation is a good example, that are unique. **See also**: *Trypanosoma*

Not all parasites have a strict requirement for *N*-glycosylation during their life cycle. The asexual intererythrocytic stage of the malarial parasite *Plasmodium falciparum* shows no detectable incorporation of radiolabelled precursors into *N*-linked glycoproteins and no oligosaccharyltransferase activity can be measured. The parasite apparently synthesises *O*-linked glycans and it is not known if glycosylation at other stages in the life cycle, for example in the sexual stage using insect vectors, occurs. **See also**: Gene Expression in Yeast

Plant glycosylation

Unlike in animal cells, in plants there appears to be no rigid spatial relationship between the ER and adjacent Golgi cisternae. This quite different organisation of internal membranes reflects the high proportion, as much as 80%, of synthetic capacity of the Golgi apparatus that is devoted to the assembly of complex polysaccharides, a unique plant Golgi function. Vacuolar and cell surface glycoproteins are glycosylated by similar mechanisms to those described for other eukaryotic cells, leading to the completion of both oligomannose and complex-type *N*-glycans. Studies of several plant lectins and other proteins have confirmed the lack of sialic acid in the complex types and the presence of core α1,3-linked fucose residues and β1,2-linked xylose to the β-mannose residue. The β1,2-linked xylose residue in *N*-glycans appears to increase the immunogenicity of the proteins to which they are attached. Antibodies raised against this epitope detect a common antigenic determinant on plants and on a number of species of the Mollusca and Insecta. **See also**: Antigens: Carbohydrates; Lectins; Plant Golgi Apparatus; *Trypanosoma*

Other lower eukaryotes

The popular use of recombinant baculoviruses to express heterologous proteins in lepidopteran (Sf9) cells for therapeutic purposes and for structural/functional studies has increased the awareness of glycosylation in insect cells. Few data are available of *N*-glycosylation in other invertebrates for comparative purposes. *N*-glycans participating in supramolecular structures formed by the annelid *Perinereis aibuhitensis* have been shown to contain mannose, fucose, galactose and amino sugars; however, no structural data are available and this is one of very few studies of lower invertebrates. C-reactive proteins from the horseshoe crab, *Limulus polyphemus* (Arthropoda taxa), are exclusively of the oligomannose type. Cell surface glycans from the slime mould *Dictyostelium discoideum* are thought to play a role in the formation of cell aggregates and have been shown to be developmentally regulated. The glycans isolated from this primitive eukaryote are largely of the oligomannose type that contain small amounts of sulfated and phosphorylated residues contributing to charge, but no sialic acid has been detected. The presence of β1,2-linked xylose in *N*-linked glycans from *D. discoideum* has been demonstrated and may account for the antigenicity of slime mould glycoproteins. **See also**: Antigens: Carbohydrates; Baculovirus Insect Cell Expression System

Good evidence exits for β1,2-xylosylation of *N*-glycans, from other than plant species, from the analysis of connective tissue glycoproteins of the snail *Lymnaea stagnalis*.

Mammalian complex *N*-linked glycosylation

The action of GlcNAc-transferases I and II in the *medial* Golgi signals the processing to complex glycoproteins by glycosyltransferases in the *trans*-Golgi elements in mammals and avian species (**Figure 5**). The action of α- and β-galactosyltransferases, α-fucosyltransferases, β-*N*-acetylgalactosaminyltransferases and α-sialyltransferases are responsible for generating the diversity of *N*-glycan structure in these species. Although each glycosyltransferase catalyses the synthesis of a single glycosidic linkage (an exception is some fucosyltransferases), several distinct enzymes are able to synthesise the same linkage. Each of the glycosyltransferase genes or complementary DNAs cloned so far predict a common type II transmembrane topology, consisting of a short *N*-terminal cytoplasmic domain, a single membrane-spanning region and a larger *C*-terminal catalytic domain. Molecular cloning efforts have substantiated distinct families of glycosyltransferase that have considerable homology at the primary amino acid level but are unrelated to members of other families, suggesting independent evolution. Members of a family will often include glycosyltransferases from widely divergent species, as demonstrated by the similarity of plant and human GlcNAc-transferase I. Members of the α1,3-fucosyltransferase family share considerable amounts of sequence similarity. The sialyltransferase family contains over 12 members, each having different acceptor specificities and primary sequence, apart from a short peptide region of 48–49 amino acids, the 'sialyl motif', a sugar–nucleotide-binding region.

Differences in the specificity of sialyl transferases may have important consequences and some have occurred very late in humanoid development from great apes prompting speculation that they may be related to brain development. Certainly, sialylation plays an important role in the brain and probably in forming neural connections. It is also important in susceptibility to infection. The influenza virus infecting birds has a specificity for a form of sialic acid

which is in an α2–3 linkage, whereas those infecting humans prefer a α2–6 linkage.

How have these transferase enzymes evolved to generate the diversity of glycan structures that occur in nature, yet are apparently quite conserved? Glycosyltransferases share the same topographical and structural organisation, but this similarity is not reflected at the genomic level as there is no common pattern of intron–exon structure between the transferase genes. Not all glycosyltransferases have evolved by shuffling exon-encoded domains. The fucosyltransferases, for example, have single exonic coding sequences, whereas the galactosyltransferase and sialyltransferase genes are multiexonic. The adaptation of organisms to create divergent glycans that meet the needs for a more sophisticated method of cell–cell interactions would be the driving force for gene duplication and/or exon shuffling. An ancestral gene containing regions encoding the catalytic domain could be used several times to create a number of distinct glycosyltransferase genes, for which tentative evidence exists.

The structures of several glycosyltransferases have now been determined and details of the reaction mechanism of some of these are known such the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase-T1 (Davies *et al*., 2005). **See also**: Evolutionary Developmental Biology: Gene Duplication, Divergence and Co-option; Multigene Families: Evolution

O-Linked Glycans: Structure and Biosynthesis

A second modification is a covalent attachment of glycans to the α-hydroxyl group of serine or threonine residues to form an *O*-linked glycan (**Figure 1**). The most common form involves the hydroxyl group at C1 of *N*-acetylgalactosamine, although xylose, mannose and other monosaccharides are occasionally found.

O-Glycans exhibit either a simple structural design, including those that are commonly found on many glycoproteins containing *N*-linked glycans, or those that have glycan chains as complex as the *N*-glycans, such as GlyCAM-1, the endothelial-associated ligand for L-selectin and the neutrophil ligand for P-selectin (**Figure 3**). Mucins, the major constituent of mucosal secretions, are rather large molecules, containing 50–80% carbohydrate by weight, where the glycan contributes greatly to the physicochemical properties of mucin glycoproteins. Unfortunately, insufficient data are generally available for assigning 'informational' properties to precise carbohydrate sequences, primarily due to difficulties in obtaining unambiguous structural information. One exception to this is the *O*-glycosylation of the muscle α dystroglycan. This is an unusual type of *O*-glycan based on *O*-linked mannose, but it has been demonstrated that defects in enzymes involved in its biosynthesis are linked to muscular dystrophy (Hewitt, 2009). **See also**: Cell Surface Glycoconjugates; Inherited Disorders of Glycosylation; Muscular Dystrophies

O-Linked glycoprotein biosynthesis

Unlike the cotranslational addition of *en bloc* synthesised glycan to asparagine residues, modification to serine and threonine amino acids is an event that requires the sequential action of several different glycosyltransferases. Consequently, trimming glycosidases do not play a role in the control of biosynthesis, and further elaboration of each glycan is defined by the competitive forces between transferases for common acceptor sites as the protein migrates through the secretory pathway of the cell. In addition, there appears no consensus amino acid sequence to direct the first event, the addition of *N*-acetylgalactosamine to serine and threonine by UDP-*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosaminyl transferases. There is a hierarchy of action within this family, with certain members acting as 'initiating' transferases, transferring GalNAc to unmodified substrates, whereas other members act only on previously glycosylated substrates (glycopeptide transferases; Tian and Ten Hagen, 2009). The specificities of the GalNAc transferase undoubtedly contribute to specificity of sites of glycosylation and not all ser/thr residues are glycosylated. There is insufficient data yet available to assess the degree of heterogeneity of the sites for *O*-glycosylation, although advances in mass spectrometry (MS) techniques for glycopeptides should help in the future elucidation of this.

As *O*-glycans appear to be clustered at the *N*-terminus of most glycoproteins and have an increased preference for solvent-exposed β turns, the specificity of the transferase may be determined by the amino acid sequence that is able to adopt an exposed β turn. It is also apparent that there is also some involvement of chaperone proteins in *O*-glycan biosynthesis. In particular, one such chaperone COSMC (core 1 β3–Gal-T specific molecular chaperone; Ju *et al*., 2008) was found to be the cause of failure of extension of *O*-glycans by addition of Gal in the Tn syndrome. This is now known to be involved in the folding of the enzyme β3-galactosyltransferase (Aryal *et al*., 2010).

Further elaboration of the mucin-type *O*-glycans can be best described by examination of their structural contents. The simplest glycans contain serine- or threonine-substituted *N*-acetylgalactosamine, followed by those containing two monosaccharides, for example Gal βl, 3GalNAc, and those containing three or more. The latter class of glycan contains structural features that are recognisable as belonging to separate groups, or 'core' glycans, of which eight have been described (Jensen *et al*., 2010).

Elongation of these core sequences by backbone structures, of which five are commonly found, is extremely diverse and the number of branch points has made complete structural elucidation of many of the larger chain glycans difficult. Understanding how glycosylation is controlled by the cell remains a goal for which little success has been achieved in providing a consensus framework

analogous to the *N*-glycans. The subcellular site for initiation of *O*-glycosylation is probably protein-, species- and tissue-specific and is a posttranslational event that takes place within the smooth membrane complex. Maturation of the glycan in the later elements of the Golgi apparatus will most likely depend on environmental conditions and on the complement of glycosyltransferases that compete for acceptor substrates.

The relatively high abundance and clustering of *O*-glycans may mean that they allow clustering of epitopes to allow multivalent attachment for such epitopes as sialyl Lewis X (Karlsson and Thomsson, 2009). It is interesting to note that the distribution of fucosyltransferases in particular may be both spatially and temporally regulated so that some structures are only present in certain cells at certain stages in development. This is believed to play an important role in developmental processes, although in most cases the exact mechanism is not yet clear. The ability to modify the cell surface structures by glycosyltransferase expression is probably an important process in cell–cell interaction.

Many of the cDNAs for transferases involved in biosynthesis have been cloned, including those that synthesise the blood group antigens, and their study will undoubtedly increase our understanding of the control mechanisms for the synthesis of this complex class of glycan. **See also**: Antigens: Carbohydrates

Species-specific *O*-glycosylation

Other species are also able to *O*-glycosidically link carbohydrate moieties to serine and threonine. In higher plants, a major modification to serine residues of lectin and cell wall proteins is the addition of galactose, and the potato lectin carries a further modification of hydroxyproline residues with arabinose. Although thought to be widely distributed in plants, these modifications have not been identified in animal species, where hydroxyprolines are confined to the collagens. Similarly, the addition of xylose in an *O*-glycosidic link to threonine in maize root slime cap and in red algae resembles the initial glycanation step in proteoglycan synthesis. **See also**: Glycosyltransferases in Plant Cell Wall Synthesis

Yeasts and fungal cells *O*-glycosylate serine and threonine residues with mannose, a process that is initiated in the ER via two, possibly three, mannosyltransferases that use dolicholphosphomannose as a sugar donor. Additional mannose residues are added in either the ER or the Golgi to form a glycan that is structurally very different to that seen in higher eukaryotes. **See also**: Fungal Physiology

Much less is known about *O*-linked glycosylation in invertebrate cells. Analysis of glycoproteins from cultured insect cells shows that they synthesise predominately a single Ser/Thr-linked GalNAc residue, only few of which are substituted by galactose and are never substituted further with sialic acid. This type of truncated structure is analogous to the *N*-glycan structures found in these cells and represents the product of restricted biosynthesis.

Other *O*-linked glycans

O-Linked *N*-acetylgalactosamine is not the only modification to serine and threonine residues, and a number of proteins with repeating epidermal growth factor (EGF) motifs have been shown to contain fucose and xylose–glucose units similarly linked. Recent data indicate that the activity of a number of nuclear and cytoplasmic proteins is critically regulated by the acquisition of one or more *O*-linked residues, particularly *N*-acetylglucosamine (Zachara and Hart, 2006). This has now been shown to be widely distributed in intracellular proteins and may well be dynamic in a reciprocal relationship with phosphorylation and thus forms an important means of controlling activity of intracellular proteins especially those in the nucleus (Jensen *et al*., 2010).

Is the presence of so few carbohydrate residues enough to modulate the function of proteins? Compared to the amount of energy spent by the cell in decorating proteins with complex, highly branched types of glycan, why have these biosynthetic pathways been conserved if similar functions can be performed with a much smaller pool of cellular enzymes? Clues to the strategy adopted by cells can be found by examination of the compartmentalised function of proteins. The semipermeable barrier that separates all cells from the external milieu is populated with proteins whose functional moieties are placed to interact with both the internal and external environments. The majority of *N*-glycosylated proteins resides at the extracellular surface or are secreted into the extracellular space. Better methods for the unambiguous analysis of glycan modifications have identified several intracellular resident proteins that are *N*-glycosylated, including those from the ER (e.g. α-glucosidase II), and from nuclear membranes and the lysosome. Very few appear to be cytoplasmic in origin, unlike the majority of *O*-GlcNAc-modified proteins. Is this because they must have different functions? *O*-GlcNAc modification has been compared to phosphorylation, largely an intracellular event, for its ubiquitousness in regulating certain protein functions. There is now good evidence for the dynamic nature of *O*-GlcNAc modification (Copeland *et al*., 2008) but we are still a long way from understanding how this control process works. **See also**: Protein: Cotranslational and Posttranslational Modification in Organelles

Recombinant Glycoproteins and Remodelling

Biotechnology companies are increasingly looking to the use of recombinant glycoproteins for therapy. For production reasons, these are usually expressed in nonhuman cell lines or even in plant cells (Chiba and Akeboshi, 2009). One consequence of this is that their glycosylation will not be the same as in the human glycoprotein. This may have several effects but that of most concern is that some

nonhuman types of glycosylation such as α-linked galactose may be immunogenic and may result in very short half-life of the recombinant protein or even a severe immune response (Galili, 2005). This has caused a lot of concern but it is now possible to modify glycosylation properties by manipulation of glycosyltransferases in the cell line or by treating the product with glycosidases and glycosyltransferases. There may even be reasons to manipulate glycosylation to improve the therapeutic efficacy of the product as in the case of erythropoietin α where additional sialic acid carrying glycan were added and greatly increased the biological half-life of the drug (Elliott *et al*., 2003).

Finally, the possibility now exists to completely synthesise glycoproteins with any desired form of glycosylation by chemical or chemico-enzymatic routes. This has been aided by significant advances in carbohydrate chemistry and has been demonstrated as proof of principle (Gamblin *et al*., 2009). **See also**: Protein Production for Biotechnology

Analysis of Glycoprotein Glycans

SDS-PAGE analysis of glycoproteins

A simple and readily available method for analysing glycoproteins is sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Historically, chemical detection methods for the carbohydrate moiety, such as periodate oxidation and detection *in situ* by Schiff reagent were used. More specific and sensitive fluorescent techniques are now available which can detect glycoproteins at the nanogram (fmol) level (Steinberg, 2009). More information on the type of glycosylation present in a glycoprotein band may also be obtained by use of labelled lectins. Lectins, plant and animal proteins that are able to bind sugars specifically either to terminal residues or as part of an extended sequence, can detect glycan chains of glycoproteins by Western blotting. This technique can identify the type and even portions of the glycan sequence in *N*- and *O*-linked glycoproteins. The denaturation by SDS also enhances the activity of the *N*-glycan cleaving enzyme peptide-N4-(*N*-acetyl-β-glucosaminyl)asparagine amidase (PNGaseF), and the released *N*-glycans may be conveniently recovered in solution by digestion of gel bands and subjected to further analysis so the glycosylation of the protein may be readily studied without its purification (Domann *et al*., 2007). **See also**: Gel Electrophoresis: One-dimensional

The contribution of carbohydrate to the mass of the polypeptide chain can be assessed after de-glycosylation using trifluoromethanesulfonic acid. This procedure efficiently cleaves *N*- and *O*-linked glycan chains, leaving the protein intact. Selective cleavage of *N*-linked glycans using enzymes, such as PNGaseF or endo-β-*N*-acetylglucosaminidase, cleave glycans nondestructively from most glycoproteins. Estimates of the polypeptide mass by SDS-PAGE before and after de-glycosylation can be used to determine the number of *N*-linked glycan chains. Unfortunately, there are no commercially available enzymes that are able to cleave all *O*-linked glycans from the protein but these may be released by chemical procedures following de-*N*-glycosylation allowing the analysis of *N*- and *O*-glycosylation from the same protein band.

Analysis of glycoproteins

Glycan release and labelling

Quantisation of the types of monosaccharide linkage and number of glycan structures has frequently been performed following release of the intact glycan, followed by qualitative analysis. Traditionally this was performed by use of anhydrous hydrazine; but now this is generally performed by enzymatic release with PNGaseF. The terminal *N*-acetylhexosamine residue can be labelled by reductive methods using fluorescent molecules, such as 2-aminobenzamide (2-AB), that provide a more sensitive reporter group, to aid detection.

Fractionation of glycans

High-performance liquid chromatography (HPLC) for glycan analysis increases resolution of mixtures, and elution times, calibrated to a hydrolysed dextran to give glucose units which can be predictive for glycan structure (Domann *et al*., 2007). Using a combination of chromatography separation methods, such as hydrophilicity and charge, a complete profile of the structures present on a glycoprotein can be analysed. The use of a fluorescent tag such as 2-AB allows a profile to be obtained showing the relative distribution of glycans at a high degree of sensitivity. This system can now be fully automated and can be used to release glycans from glycoprotein bands separated by SDS-PAGE. **See also**: Liquid Chromatography

Mass spectrometric analysis

Developments in mass spectrometry instrumentation have aided the rapid analysis of glycans by this technique. A variety of systems may be employed and some give better results for certain types of glycosylation. The labelling of the glycan may aid its analysis by MS. The techniques may be applied to both the released glycans or analysis of glycopeptides produced by cleavage of the peptide. Methods to ionise glycans and detection by mass spectrometry require relatively little material (low pmol range). The matrix-assisted laser desorption/ionisation (MALDI) technique involves irradiation of the sample mixed with an ultraviolet light-absorbing matrix followed by mass separation of the ions. This provides only mass values and not linkage information, unless some fragmentation of the glycan is introduced. Alternatively, electrospray instrumentation allows ionisation in either positive or negative ion mode which can be separated by time-of-flight in multidimensional modes and this can now be directly coupled to fragmentation (Harvey, 2005). MALDI mass spectrometry provides a rapid method for profiling neutral

N-linked glycans as their [M + Na](+) ions which can be fragmented by collision-induced decomposition to give spectra containing both glycosidic and cross-ring fragments. Electrospray ionisation mass spectrometry is more versatile in that it is relatively easy to change the type of ion that was formed and, furthermore, unlike MALDI, electrospray does not cause extensive loss of sialic acids from sialylated glycans. Negative ions formed by addition of anions such as chloride and, particularly, nitrate, to the electrospray solvent are stable and enable singly charged ions to be obtained from larger glycans than was possible in positive ion mode. Fragmentation of negative ions follows specific pathways that define structural details of the glycans that are difficult to obtain by classical methods such as exoglycosidase digestion (described in the following text).

The improved instrumentation now available coupled with bioinformatic tools such as the open source GlycoWorkBench (Ceroni *et al*., 2008) is allowing much greater use of such fragmentation techniques. Recent technical developments have included improvements in tandem mass spectrometry (MS/MS and MS(n)) sequencing methodologies, more sensitive methods for analysing sulfated and polysialylated glycans and better procedures for defining the sites of *O*-glycosylation (North *et al*., 2009). The analysis of glycans in solution by electrospray mass spectrometry has the additional advantage that direct coupling to HPLC liquid phase chromatographic separation is possible. **See also**: Mass Spectrometry in Protein Characterization

NMR analysis usually requires much larger amounts of free glycan to confirm the identity of structure and linkage, but can do so unambiguously and nondestructively at the micromol level. Multidimensional techniques may be employed and the availability of data from bioinformatic sources can assist in the interpretation of data. However, this remains the method of choice for determining the precise conformation of glycans and can be particularly useful in looking at interactions of glycans with proteins (Jimenez-Barbero *et al*., 2008). **See also**: Macromolecular Structure Determination: Comparison of Crystallography and NMR

Enzymatic methods

The high degree of specificity of glycosidase enzymes means that they can very usefully be employed in gaining further structural information especially in terms of the type of linkages present between residues in the glycan. Structural assignment using sequential exoglycosidase digestion relies on the known specificity and purity of glycosidase-catalysed hydrolysis (Campbell *et al*., 2008). Exoglycosidases are usually catalytically specific for the monosaccharide that is cleaved and its anomeric configuration; for example β-galactosidase only hydrolyses glycans containing terminal β-galactose residues, and α-fucosidase those containing terminal α-fucose residues. Coupled with analysis of the reaction products after glycan digestion at the nonreducing terminus, using HPLC or mass spectrometry, the total structure can be inferred (Abd Hamid *et al*., 2008; Domann *et al*., 2007). Software (GlycoBase and AutoGU) which allows calculation of glucose units and the products of enzyme action is now available to aid the interpretation of the glycosidase digestion data (Campbell *et al*., 2008). **See also**: Bioinformatics; Glycosidases: Functions, Families and Folds

References

Abd Hamid UM, Royle L, Saldova R *et al*. (2008) A strategy to reveal potential glycan markers from serum glycoproteins associated with breast cancer progression. *Glycobiology* **18**: 1105–1118.

Apweiler R, Hermjakob H and Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta* **1473**: 4–8.

Arnold JN, Wormald MR, Sim RB, Rudd PM and Dwek RA (2007) The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annual Review of Immunology* **25**: 21–50.

Aryal RP, Ju T and Cummings RD (2010) The endoplasmic reticulum chaperone Cosmc directly promotes in vitro folding of T-synthase. *Journal of Biological Chemistry* **285**: 2456–2462.

Campbell MP, Royle L, Radcliffe CM, Dwek RA and Rudd PM (2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics* **24**: 1214–1216.

Ceroni A, Maass K, Geyer H *et al*. (2008) GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *Journal of Proteome Research* **7**: 1650–1659.

Chiba Y and Akeboshi H (2009) Glycan engineering and production of 'humanized' glycoprotein in yeast cells. *Biological & Pharmaceutical Bulletin* **32**: 786–795.

Copeland RJ, Bullen JW and Hart GW (2008) Cross-talk between GlcNAcylation and phosphorylation: roles in insulin resistance and glucose toxicity. *American Journal of Physiology. Endocrinology and Metabolism* **295**: E17–E28.

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.

Domann PJ, Pardos-Pardos AC, Fernandes DL *et al*. (2007) Separation-based glycoprofiling approaches using fluorescent labels. *Proteomics* **7**(suppl. 1): 70–76.

Elliott S, Lorenzini T, Asher S *et al*. (2003) Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nature Biotechnology* **21**: 414–421.

Galili U (2005) The alpha-gal epitope and the anti-Gal antibody in xenotransplantation and in cancer immunotherapy. *Immunology and Cell Biology* **83**: 674–686.

Gamblin DP, Scanlan EM and Davis BG (2009) Glycoprotein synthesis: an update. *Chemical Reviews* **109**: 131–163.

Gemmill TR and Trimble RB (1999) Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochimica et Biophysica Acta* **1426**(2): 227–237.

ten Hagen KG, Zhang L, Tian E and Zhang Y (2009) Glycobiology on the fly: developmental and mechanistic insights from *Drosophila*. *Glycobiology* **19**: 102–111.

Harvey DJ (2005) Structural determination of *N*-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. *Proteomics* **5**: 1774–1786.

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

Hewitt JE (2009) Abnormal glycosylation of dystroglycan in human genetic disease. *Biochimica et Biophysica Acta* **1792**: 853–861.

Jefferis R (2007) Antibody therapeutics: isotype and glycoform selection. *Expert Opinion on Biological Therapy* **7**: 1401–1413.

Jensen PH, Kolarich D and Packer NH (2010) Mucin-type *O*-glycosylation – putting the pieces together. *FEBS Journal* **277**: 81–94.

Jimenez-Barbero J, Diaz MD and Nieto PM (2008) NMR structural studies of oligosaccharides related to cancer processes. *Anti-Cancer Agents in Medicinal Chemistry* **8**: 52–63.

Ju T, Aryal RP, Stowell CJ and Cummings RD (2008) Regulation of protein *O*-glycosylation by the endoplasmic reticulum-localized molecular chaperone Cosmc. *Journal of Cell Biology* **182**: 531–542.

Karlsson NG and Thomsson KA (2009) Salivary MUC7 is a major carrier of blood group I type *O*-linked oligosaccharides serving as the scaffold for sialyl Lewis x. *Glycobiology* **19**: 288–300.

Mackeen MM, Almond A, Deschamps M *et al*. (2009) The conformational properties of the Glc3Man unit suggest conformational biasing within the chaperone-assisted glycoprotein folding pathway. *Journal of Molecular Biology* **387**: 335–347.

North SJ, Hitchen PG, Haslam SM and Dell A (2009) Mass spectrometry in the analysis of *N*-linked and *O*-linked glycans. *Current Opinion in Structural Biology* **19**: 498–506.

Parodi AJ (1999) Reglucosylation of glycoproteins and quality control of glycoprotein folding in the endoplasmic reticulum of yeast cells. *Biochimica et Biophysica Acta* **1426**: 287–295.

Petrescu A, Wormald MR and Dwek RA (2006) Structural aspects of glycomes with a focus on *N*-glycosylation and glycoprotein folding. *Current Opinion in Structural Biology* **16**: 600–607.

Rademacher TW, Parekh RB and Dwek RA (1988) Glycobiology. *Annual Review of Biochemistry* **57**: 785–838.

Shental-Bechor D and Levy Y (2009) Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Current Opinion in Structural Biology* **19**: 524–533.

Steinberg TH (2009) Protein gel staining methods: an introduction and overview. *Methods in Enzymology* **463**: 541–563.

Tian E and Ten Hagen KG (2009) Recent insights into the biological roles of mucin-type *O*-glycosylation. *Glycoconjugate Journal* **26**: 325–334.

Vadaie N and Jarvis DL (2004) Molecular cloning and functional characterization of a Lepidopteran insect beta4-*N*-acetylgalactosaminyltransferase with broad substrate specificity, a functional role in glycoprotein biosynthesis, and a potential functional role in glycolipid biosynthesis. *Journal of Biological Chemistry* **279**: 33501–33518.

Wormald MR, Petrescu AJ, Pao Y *et al*. (2002) Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modelling. *Chemical Reviews* **102**: 371–386.

Yan A and Lennarz WJ (2005) Unraveling the mechanism of protein *N*-glycosylation. *Journal of Biological Chemistry* **280**: 3121–3124.

Zachara NE and Hart GW (2006) Cell signaling, the essential role of *O*-GlcNAc!. *Biochimica et Biophysica Acta* **1761**: 599–617.

Further Reading

Fraser-Reid BO, Tatsuta K, Thiem J *et al*. (eds) (2008) *Glycoscience Chemistry and Chemical Biology*, 2nd edn, vol. XCVI, 2874 p. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-540-30429-6; ISBN: 978-3-540-36154-1.

Kamerling JP, Boons G-J, Lee YC *et al*. (eds) (2007) *Comprehensive Glycoscience*, *Volume 3: Biochemistry of Glycoconjugate Glycans; Carbohydrate-Mediated Interactions*. Elsevier Science Title ISBN: 978-0-444-52746-2.

von der Lieth CW, Luetteke T and Frank M (eds) (2009) *Bioinformatics for Glycobiology and Glycomics: An Introduction* (Hardcover). Chichester, UK: Wiley, Vol. 1, 471pp.

Taniguchi N, Suzuki A, ItoY *et al*. (eds) (2008) *Experimental Glycoscience*, XVIII. 978-4-431-77921-6.

Taylor ME and Drickamer K (2003) *Introduction to Glycobiology*, 2nd edn. Oxford: Oxford University Press.

Varkii A, Cummings RD, Esko JD *et al*. (eds) (2009) *Essentials of Glycobiology*. 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. ISBN 978-087969770-9.