Review TRENDS in Biotechnology Vol.22 No.10 October 2004

Full text provided by www.sciencedirect.com

Sugar-mediated ligand–receptor interactions in the immune system

Pauline M. Rudd, Mark R. Wormald and Raymond A. Dwek

Oxford Glycobiology Institute, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK

Most molecules involved in the recognition and elimination of pathogens by the immune system are glycoproteins. Oligosaccharides attached to glycoproteins initiate biological functions through mechanisms that involve multiple interactions of the monosaccharide residues with receptors. For example, calreticulin, a quality-control lectin-like chaperone, interacts with glucosylated mannose glycans presented by empty major histocompatibility complex (MHC) class I molecules, retaining them in the endoplasmic reticulum (ER) until antigenic peptide is loaded. Clusters of specific IgG glycoforms, present in increased amounts in rheumatoid arthritis, bind mannose-binding lectin (MBL), providing a potential route to inflammation through activation of the complement pathway. Secretory IgA glycans bind gut bacteria, and an unusual cluster of mannose residues on gp120, the surface coat protein of the HIV virus, is recognized by the novel 'domain-swapped' IgG 2G12 serum antibody.

Glycosylation is essential for life. Almost all organisms including fungi, yeast, plants, insects, fish, birds and mammals glycosylate most of their secreted and cell-surface proteins. In addition, viruses, which have no glycosylation machinery of their own, attach sugars to their envelope proteins by exploiting the biosynthetic pathways of their hosts. The essential role of glycosylation in viability has been demonstrated in knockout mice. The *Mgat1* gene encodes *N*-acetylglucosamine transferase I, an enzyme required in the processing of complex type sugars (Box 1). When *Mgat1* was deleted in mice, the offspring died after 9 days in embryo as a result of defects in vascularization and neural tube closure [1]. Similarly, deletion of the *Mgat2* gene, encoding *N*-acetylglucosamine transferase II, which is also necessary for processing complex type sugars, caused frequent postnatal lethality and 99% of the mice died within the first week of birth [2].

In the immune system, oligosaccharides attached to both host and pathogenic proteins have many roles that are both structural and functional (Box 2). In this review we discuss some of these roles, focusing in particular on recognition events in which oligosaccharides generally initiate biological functions through mechanisms that involve multiple interactions of the monosaccharide residues with receptors. This requirement for multivalency is a vital means of ensuring that physiological consequences generated by the activation of signalling pathways are not triggered inadvertently by single weak interactions, but only in response to a strong stimulus. Although individual monosaccharides have low affinity for protein receptors – usually in the millimolar to micromolar range – two mechanisms enable oligosaccharide ligands to interact with their receptors with high affinity or avidity.

The first mechanism requires the presence of subsites in the receptor, each of which can accommodate a monosaccharide residue with very different affinities. An example of such a receptor is the quality-control lectin-like chaperone calnexin, which can interact with a tetrasaccharide epitope presented by an unfolded glycoprotein, thereby retaining the glycoprotein in the ER (see below). The second mechanism depends on the presentation of multiple oligosaccharides in such a way that the sugars can interact with several carbohydrate recognition sites on the receptor. An example of this mechanism is the functional binding of the carbohydrate recognition domains of MBL to specific aggregated IgG glycoforms that present arrays of terminal *N*-acetylglucosamine residues (see below). Insights into some of the ligand–receptor interactions that involve glycoproteins in the immune system have come from combining glycan analysis, protein structural data and biochemical data with molecular modelling.

In general, oligosaccharides confer stability on glycoproteins and protect them from proteases and nonspecific protein–protein interactions, but, as we review here, some oligosaccharides form specific recognition epitopes that have functional consequences on receptor binding. Roughly 80% of secreted and cell-surface proteins are glycosylated. In the ER, *N*-linked sugars are usually attached co-translationally to partially folded proteins through the side chains of asparagine residues contained in the consensus sequence Asn-Xaa-Ser/Thr (where Xaa is any amino acid except proline). The fully folded proteins are transported to the Golgi, where *N*-linked glycans are further processed to complex type oligosaccharides and *O*-glycans can be added to accessible side chains of serine or threonine residues [3,4].

Notably, molecular modelling of both proteins and their attached oligosaccharides provides graphic illustrations of the finding that the sugars frequently occupy a larger three-dimensional space than do the protein domains to which they are attached. Although glycoproteins usually consist of an ensemble of glycosylated

Corresponding author: Pauline M. Rudd (pmr@glycob.ox.ac.uk).

Available online 14 August 2004

www.sciencedirect.com 0167-7799/$ - see front matter © 2004 Elsevier Ltd. All rights reserved. doi:10.1016/j.tibtech.2004.07.012

Box 1. Glycoproteins comprise a mixture of glycosylated variants (glycoforms)

The covalent attachment of sugars (glycosylation) is a common posttranslational modification of both secreted and cell-surface proteins and one on which the cell expends a significant amount of energy. About 30 different enzymes are required to attach and to process the oligosaccharides linked to asparagine (*N*-linked glycans) and to serine or threonine (*O*-linked glycans) residues (Figure I). *N*-linked sugars are attached to proteins in the endoplasmic reticulum (ER), where a preformed oligosaccharide precursor attached to dolichol phosphate is transferred to some asparagine residues that are part of the motif, Asn-Xaa-Ser/Thr (where Xaa is any amino acid except proline). Further processing of the sugars on the nascent glycoprotein takes place first in the ER and then in the Golgi. An important intermediate in the ER is the monoglucosylated glycan $Glc_1Man_{9-7}GlcNAc_2$, which provides the folding protein with the means of entering the calnexin/calreticulin quality-control pathway. *O*-linked sugars are added to the fully folded protein in the Golgi. Attachment of an initial *N*-acetylgalactosamine or fucose residue is followed by sequential additions of monosaccharides by transferases that branch, elongate and terminate the glycans.

Glycoproteins usually consist of numerous glycoforms in which the same amino acid sequence is diversified by a range of different sugars at each of the glycosylation sites. For example, CD59 expressed on human erythrocytes has more than 150 glycoforms [32], which arise when the glycan processing pathways terminate at various stages on different copies of a glycoprotein. Glycosylation processing is controlled by the amounts of processing enzymes expressed in any particular cell, the levels of the sugar-nucleotide donors and the three-dimensional structure of the specific protein around the glycosylation sites [3].

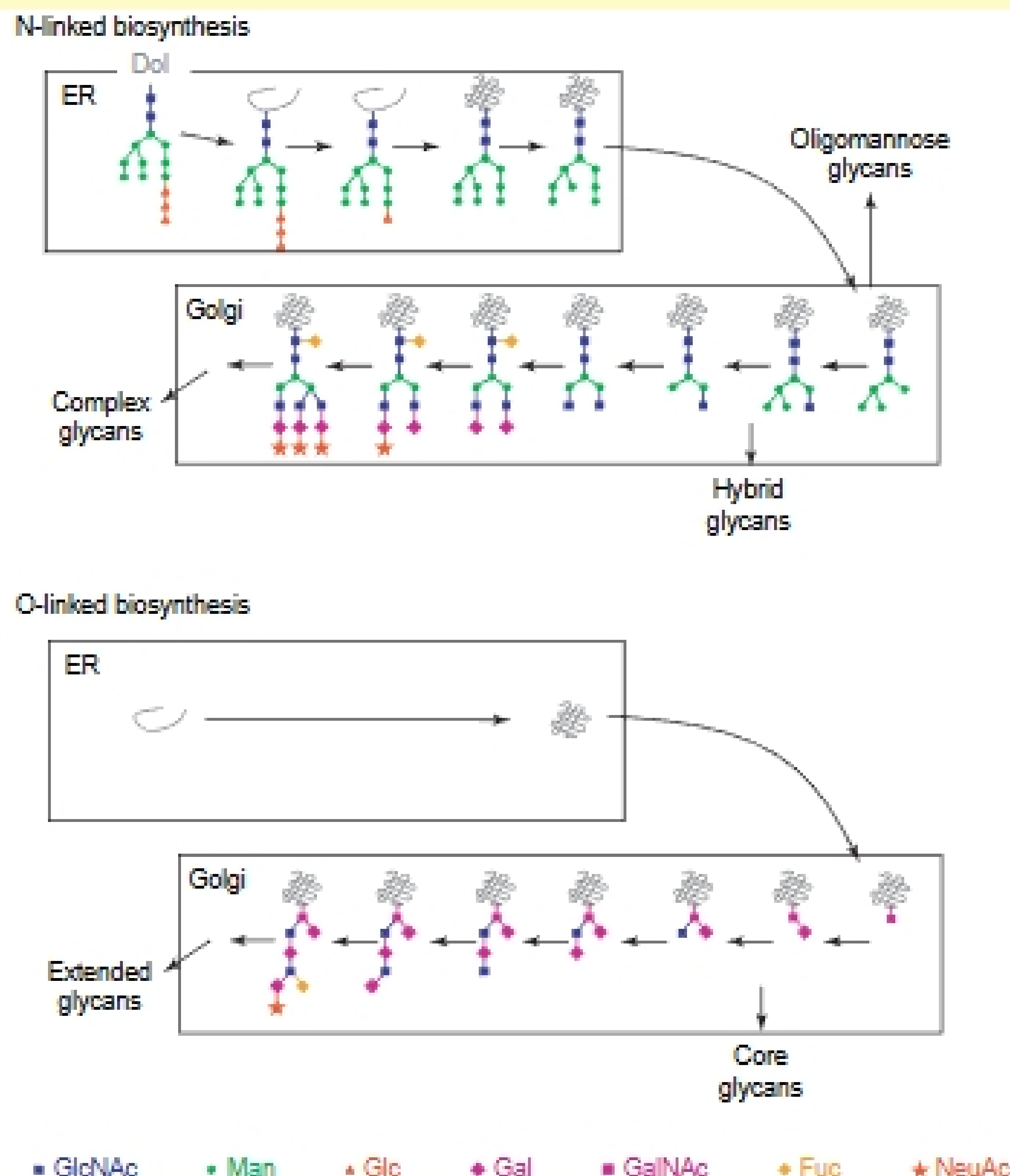

Figure I. Biosynthetic pathways for *N*- and *O*-linked glycans.

variants ('glycoforms'), specific sugar epitopes required for recognition might be a feature of many glycoforms, and thus a complete glycan analysis is required if such structures are to be identified.

Glycan analysis

State-of-the-art glycan analysis of scarce biological glycoproteins is based on a combination of high-performance liquid chromatography (HPLC) analysis and mass spectrometry. Exoglycosidase array digestions provide monosaccharide and linkage information and, if sufficient material is available, mass spectrometry fragmentation data can also provide composition and linkage information.

The analytical strategy used in the Oxford Glycobiology Institute is shown in Figure 1. Glycans are released either enzymatically from SDS–PAGE gel bands (for *N*-linked

Box 2. Oligosaccharides on proteins have both structural and functional roles

Oligosaccharides have many different functions, which can be broadly subdivided into two categories. First, the glycans attached to proteins have structural roles [3,4]. The sugars contribute to thermodynamic stability throughout the protein and not only in the region of the glycosylation site [33]. The sugars, which are very flexible around the *N*-glycosidic linkage in particular, occupy a conformational space that is often as large as a protein domain. The large sizes and dynamics of the sugars enable them to protect extensive regions of the protein surface from nonspecific protein–protein interactions and importantly from proteases. On membrane glycoproteins, which are usually heavily glycosylated, the sugars can also function to orient the protein with respect to the membrane and can prevent the protein from interacting with cell-surface lipids. Many proteins are attached to the cell surfaces by glycosylphosphatidylinositol (GPI) anchors, which contain a glycan core that links the protein to the lipid chains that insert into the top leaflet of the cell membrane. CD59 is an example of a protein that has *N*- and *O*-linked sugars, as well as a GPI anchor (Figure I).

Second, oligosaccharides attached to proteins take part in many highly specific recognition events. For example, through a Ca^{2+} ion embedded in their structure, the C-type lectins can ligate to two hydroxyl groups that are presented in a di-equatorial orientation on the terminal monosaccharide residue, such as *N*-acetylglucosamine, mannose and fucose, of oligosaccharides [34]. The affinity for a single monosaccharide residue of a protein receptor is usually low (in the millimolar range). In general, multiple interactions (involving either the recognition of several residues on a single glycan or the presentation of arrays of glycans to multivalent receptors) are required if recognition is to trigger biological function.

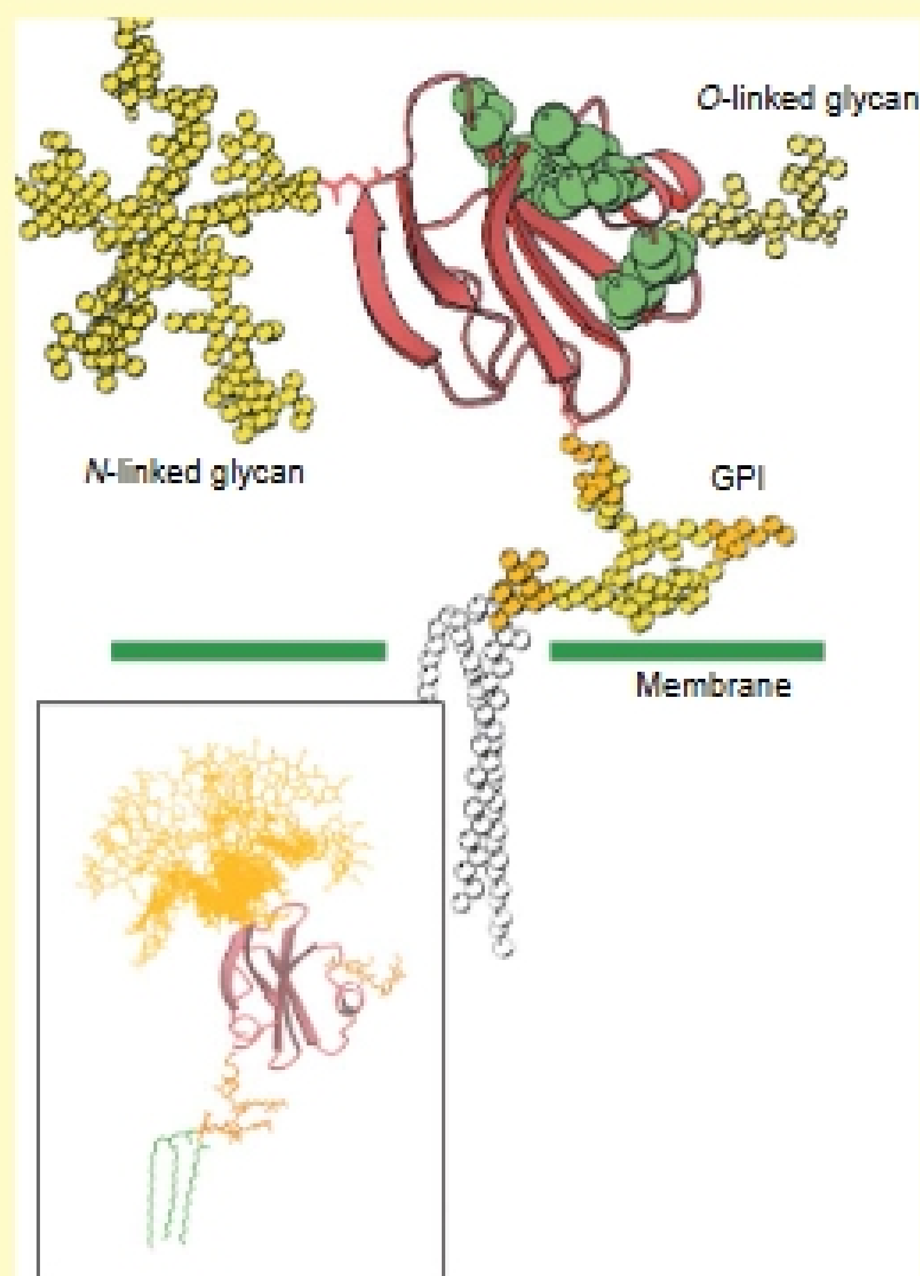

Figure I. Molecular model of CD59. A single conformer of CD59 is shown together with a series of conformational overlays of the *N*-glycan, which emphasizes the large molecular volume of the glycans as well as their dynamic mobility. The binding site residues are shown in green.

glycans) or by hydrazinolysis (for *N*- and *O*-linked glycans). The glycan pool is then labelled with 2-aminobenzamide. Aliquots of the labelled glycan pool are treated with exoglycosidase arrays and, together with the original pool, are analysed by normal-phase HPLC. *N*- and *O*-glycan structures are assigned to the peaks automatically from entries in the Oxford Glycobiology Institute database by using elution positions, measured in 'glucose units' derived from a standard dextran hydrolysate ladder, and exoglycosidase digestion patterns [5,6].

$Glc_1Man_{9-7}GlcNAc_2$ oligosaccharides bind subsites on calnexin

Peptides enter the secretory pathway by passing through the heterotrimeric Sec61 channel into the ER, where the oligosaccharyl transferase complex is located in close proximity to the luminal exit. The *N*-linked oligosaccharide precursor ($Glc_3Man_9GlcNAc_2$, where Glc is glucose, Man is mannose and GlcNAc is *N*-acetylglucosamine) is attached to appropriate asparagine residues in the nascent peptide by the oligosaccharyl transferase complex and rapidly processed to $Glc_1Man_{9-7}GlcNAc_2$ sugars (Figure 2a). These monoglucosylated glycans confer on every nascent glycoprotein the possibility of entering the calnexin/calreticulin folding and quality-control pathway. In a dynamic cycle in which the terminal glucose residue is alternately removed and replaced, these chaperones release and rebind unfolded proteins, retaining them in the ER until they achieve their properly folded conformation (Figure 2a).

The proposed glycan recognition epitope in the $Glc_1Man_{9-7}GlcNAc_2$ chain is the tetrasaccharide Glc_1Man_3 on the α1,3-arm (Figure 2b) [7,8]. The free tetrasaccharide binds to calreticulin with micromolar affinity [9]. Calreticulin has a single binding site for glucosylated glycans [9], and calnexin is also thought to have a single binding site with similar affinity, because of its high sequence homology with calreticulin. Studies on the calnexin-binding protein tyrosinase have shown, however, that this protein does not co-precipitate with anti-calnexin antibodies unless at least two of its four *N*-glycosylation sites are occupied [10], suggesting that more than one calnexin molecule can bind to glycoproteins presenting more than one oligosaccharide or that calnexin functions as a dimer in a larger chaperone complex [11–13].

Monoglucosylated glycoforms of the human MHC class I molecule (MHCI), which is *N*-glycosylated at Asn86 (Figure 2b), bind both of the lectin-like chaperones calnexin and calreticulin but at different stages of protein maturation (Box 3). Initial binding to calnexin facilitates the efficient folding and assembly of MHCI heavy chains with β2-microglobulin. The glycan on MHCI then binds to soluble calreticulin, which forms part of the peptide-loading complex. The interaction of calreticulin with MHCI stabilizes the multimolecular complex and retains