Role of protein glycosylation in immune regulation

Elizabeth F Hounsell, Michael J Davies

Most cell surface and secreted proteins are glycosylated – that is, they have one or more oligosaccharide chains covalently attached to specific amino acids. These oligosaccharides usually make up a significant amount of the hydrodynamic mass of the molecules and also have effects on protein conformation, surface expression, stability, circulating half-life, activity, and antigenicity. In addition, oligosaccharide sequences are themselves recognised as antigens and as ligands for carbohydrate binding proteins (lectins). The immune system exemplifies the diversity of oligosaccharide structure and function, which this review will aim to communicate. In particular, the relevance of oligosaccharide heterogeneity, antigenicity, and immune regulatory activity in autoimmunity and microbial pathogenesis will be considered. Firstly, a general description of the structure and three dimensional arrangement of protein glycosylation will be given. Then, we describe the structure of the major glycoprotein effector molecules involved in immune regulation and concentrate on the role of T cell γ δ cells in tuberculosis and rheumatoid arthritis (RA).

General description of protein glycosylation (also see Glossary)

Protein glycosylation is of two major types called N- or O-linked depending on the linkage of the oligosaccharide chain through asparagine (NH₂) or serine/threonine (OH), respectively. N-linked chains typically have a pentasaccharide core of mannose (Man) and N-acetylglucosamine (GlcNAc) residues with additional mannose residues (high mannose chains) or backbone galactose-N-acetylglucosamine sequences with and without chain terminating sialic acid residues (complex chains). Hybrid chains are defined as those having some high mannose branches and some complex-type sequences. The core and backbone residues of complex chains are invariably decorated with blood group and related antigenic sequences involving substitution with fucose, (Fuc), N-acetylgalactosamine (GalNAc), and galactose (Gal). The presence of the different chains can be distinguished not only by chemical composition but also by their susceptibility to enzyme digestion – all are released by peptide-N-glycosidase F, whereas endoglycosidase H only releases high mannose chains.

O-linked glycosylation discussed in this article is characterised by core regions based on N-acetylgalactosamine linked to protein, with galactose, N-acetylglucosamine, or N-acetylgalactosamine attachment. Extension of the chains is by substitution patterns similar to those of the outer branches of N-linked glycosylation (although there are differences). In both chains the diversity of oligosaccharide sequences is manifest in the large number of ways the monosaccharides can be linked together (that is, through 1→4 hydroxyl groups, with and without branching, in α or β configuration, etc). Each linkage has a particular set of allowed conformations in solution which provide specific orientations of functional groups for molecular recognition. Such epitopes usually extend over a tri- to heptasaccharide sequence, but topographical epitopes caused by folding back of long carbohydrate chains onto themselves or onto protein are possible. In addition, the oligosaccharide cores of O-linked chains are in close association with the protein backbone, which strongly influences protein conformation and oligosaccharide-protein antigenicity. Often multiple O-glycosylation sites are clustered in one region of the protein to accentuate these conformational effects (for example, CD8 discussed below).

N-linked chains are in general less stereochromatically restricted around the protein-oligosaccharide linkage than their O-linked counterparts, and, in addition, have considerable flexibility at branch points within the chain. The oligosaccharide moieties, therefore, sample a large amount of conformational space, but in some cases they may be restrained by interaction with the protein backbone or adjacent oligosaccharides attached to the same protein. The paradigm for protein-to-N-glycosylation interaction is the Fe region of IgG where electron density for oligosaccharides has been shown by x ray crystallography which can be interpreted as constraints by the protein to impose a single conformational state. In other glycoproteins that have been characterised by x ray crystallography the oligosaccharide conformation could not be discerned at high resolution – for example, studies of the major histocompatibility complex (MHC) glycoprotein molecule HLA-A2. For still further glycoproteins, crystals of sufficiently high quality for x ray crystallography are not formed unless some of the glycosylation is modified. This can be for the reasons already mentioned – that is, large hydrodynamic volume (in particular when multiple glycosylation is present); effects on protein conformation; or, inherent oligosaccharide flexibility. An additional factor, discussed next, is the microheterogeneity of oligosaccharide structure at each glycosylation site within a protein. Because of the difficulties of their visualisation by x ray studies the characterisation of protein
glycosylation has largely relied on enzymatic, gel, and chromatographic methods of analysis with, when possible, high resolution physicochemical techniques (in particular, mass spectrometry and nuclear magnetic resonance spectroscopy).2-4 7

**Microheterogeneity of oligosaccharide sequences of glycoproteins**

All glycoproteins exhibit heterogeneity such that they exist as a population having a spectrum of oligosaccharide sequences (glycoforms). Variation occurs in the number and length of branches (or antennae) leading from the glycan core and from the protein backbone, as well as the amino acid sequence through which the glycans are linked. The length and number of such antennae can vary substantially, as can the type of saccharides present, giving rise to cores, short chains, and to highly branched glycosaccharides.8 The presence of these chains, termed Gal, is suggested to be a marker of RA that may be functional by reducing the intramolecular oligosaccharide-to-protein interaction discussed above or by effecting binding to GlcNAcβ specific endogenous lectins, or both. This binding would have to be specific to the GlcNAcβ1-6/4/2Man sequence as GlcNAcβ1-3Gal and GlcNAcα1-4Gal are found quite commonly as chain terminating groups in mucus.3

The studies quoted in reference 8 were carried out on IgG pooled from human serum. From other studies it is known that glycoproteins from single donors can also show considerable heterogeneity — for example, more than 250 different glycoforms from the Tamm-Horsfall proteins of the urine of one donor.9 There is also considerable variation in cell-cell glycosylation — for example, in glycoproteins from single cell lines.10 The structural changes in IgG are also mirrored in cell studies from patients with RA11 or from the lpr autoimmune mouse model.12 These studies provide support for the contention that particular clones of B cells can secrete stimulated to express immunoglobulins of a restricted phenotype.13 Since the original observations of Gal in rheumatoid patients14 several other immune disorders, granulomatus-type diseases such as tuberculosis, and age matched controls have shown the prevalence of this type of IgG glycosylation.11 Although, in general, there is a large variability in the glycosylation of glycoproteins in normal and diseased states, the Gal phenotype remains relatively restricted to IgG. The relevance of this observation within the context of the complexity of autoimmunity remains obscure. This review explores other areas where protein glycosylation may play a part in immune regulation.

**Direct role of protein glycosylation**

In addition to the heterogeneity in the types of glycoprotein chains discussed above, there are further glycosylation patterns found in, for example, parasite and mammalian glycosylphosphatidylinositol lipid anchors14 and bacterial glycosylation conjugates, including cell wall peptidoglycans, capsular polysaccharides, and O-antigen-type specific lipopolysaccharides.15 16 These molecules are highly antigenic, but in the case of lipopolysaccharides and peptidoglycan are more likely to exert their arthritogenic effect through mechanisms involving direct mitogenicity.17 The mammalian proteoglycans are an additional type of glycosylation which needs to be discussed in this context. One of the initial responses in the rheumatoid joint is breakdown of the proteoglycan matrix which, along with collagen, make up cartilage. Hyperimmunisation with host proteoglycans will lead to experimental arthritis.17 Proteoglycans are classically very highly molecular weight molecules of different uronic acid-hexosamine repeating disaccharide sequences having multiple variable sulphation.18 Relatively short regions of proteoglycan sequences, however, can also be found on membrane-type glycoproteins — for example, the invariant chain of class II MHC19 and lymphocyte antigen CD44.20 Therefore besides direct damage to the joint, the enzymes induced in arthritis which degrade proteoglycans may also directly affect T cell function. Proteoglycans are also common constituents of secretory granules of a variety of haemoepoietic cells.21 In mast cells evidence is accumulating for their function during degranulation in controlling the release of mast cell proteases,22 which are the mediators of allergic inflammatory reactions. In addition, proteoglycans at endothelial cell membranes are necessary for an intact microvasculature, the breakdown of which is an essential component of the rheumatoid process.23

Other glycoproteins are being increasingly implicated in the inflammatory process — in particular, the recruiting of neutrophils to areas of endothelial cell damage. Studies have shown that particular oligosaccharide structures on a subset of neutrophil glycoproteins are believed to be recognised by select WELA to induce an activated endothelial cells.24 The major oligosaccharide structure involved, called sialyl Lewis X, belongs to a family of previously characterised oligosaccharide differentiation antigens25 which have now found a role in cellular interactions. These interactions are influenced by cytokines such as tumour necrosis factor.26 Other cytokines, the interleukins IL-1 and IL-2, have been shown to have lectin properties and respectively bind to Tamm-Horsfall glycoprotein27 (also called uromodulin when found in the urine of pregnant females with more glycosylation than the Tamm-Horsfall glycoprotein discussed above) and high mannose oligosaccharides.28 This may explain the well known immunosuppressive effect of the oligosaccharides of Tamm-Horsfall glycoprotein/uromodulin and also of the high mannose chains of yeast. A second mechanism for the immunosuppressive effects of polysaccharides of microbial origin is
their uptake by macrophages and disruption of the normal paths of antigen presentation.\textsuperscript{26} Mention should also be made here of (a) the lectin properties of IgD\textsuperscript{36}; (b) the strong homologies between a rat IgE binding protein and galactose binding proteins such as CBP35 and Mac-2 macrophage antigen\textsuperscript{31}; and (c) the glycosylation inhibiting factor, which alters the glycosylation of IgE binding proteins causing selective suppression of IgE synthesis.\textsuperscript{32}

Effect of glycosylation on immune protein function

In addition to the specific functions of oligosaccharides in immune regulation, glycosylation has several regulatory effects on protein function and recognition. The oligosaccharide chains of glycoproteins tend to cover a large surface area inhibiting immune recognition and processing of the underlying protein. Many of the effector molecules in immune regulation, including the immunoglobulins mentioned above, are glycosylated. The cytokine receptors so far characterised are glycoproteins, as are some of the cytokines themselves.\textsuperscript{33} Oligosaccharides are known to influence the conformation of the protein to which they are attached and this is particularly true of the multiple O-glycosylation found, for example, on the receptor for IL-2\textsuperscript{34} and on CD8. The first evidence for the importance of O-glycosylation in lymphocyte antigens was reported for CD45.\textsuperscript{35} For CD8 it has been shown\textsuperscript{36} that O-glycosylation must be removed before crystallisation can take place for x-ray crystallography studies. Interestingly, the related molecule CD4, which has a role equivalent to that of CD8 in their respective binding to class I and class II MHC molecules,\textsuperscript{37} has a very different glycosylation pattern with no reported O-glycosylation but having four possible N-linked sites, the occupancy of which is a prerequisite for correct folding and membrane expression.\textsuperscript{38} MHC molecules are themselves glycosylated and this can influence antigen presentation.\textsuperscript{39} In fig 1 we have chosen to illustrate the relative sizes and orientation of a single oligosaccharide chain as present on the human class I MHC molecule HLA-A2. It depicts the minimum sized N-linked complex chain which in the native molecule could have additional branches and peripheral substitution. As discussed, most glycoproteins have more than one glycosylation site. Figure 2 presents a

![Figure 1](http://ard.bmj.com/Ann Rheum Dis; first published as 10.1136/ard.52.Suppl_1.S22 on 1 March 1993. Downloaded from http://ard.bmj.com/)  

\[\text{Antigen binding pocket} \rightarrow \text{Oligosaccharide} \]

\[\alpha_1 \quad \alpha_2 \quad \alpha_3\]

**Figure 1** A computer graphics molecular model of MHC class I HLA-A2 α-α domains (270 amino acids) taken from Brookhaven files of crystallographic studies carried out by Bjorkman et al.\textsuperscript{9} The protein carbon backbone is accentuated with a ribbon. The \(\alpha_1\) and \(\alpha_2\) region \(\alpha\) helices surround the antigen binding pocket. A disialyated biantennary nonasaccharide is attached at the consensus N-linked glycosylation site, Asn 86 (top right), at the end of the \(\alpha_3\) helix. This represents one of many possible solution conformers of the oligosaccharide chain which will explore space within a dome of the dimensions shown. Modelling was carried out using Biosym software on a Silicon graphics IRIS workstation.
Role of protein glycosylation in immune regulation

synopsis of the oligosaccharide chains of CD44, CD4, CD8 αβ, and the T cell receptor (TCR) αβ chains showing their various glycosylation patterns taken from references 20, 36, 38, and 40–43.

A more complicated pattern of cell specific glycosylation of T cell γδ receptors has emerged (fig 3), which we discuss here in full to illustrate the type of diversity encountered, together with the evidence implicating this T

Figure 2 The glycosylation patterns of the human T cell glycoproteins CD44, CD4, CD8 (αβ heterodimer), and the T cell receptor (HPB-MLT α chain, HPB-ALL β chain). Consensus N-glycosylation sites are denoted by ▶, characterised O-glycosylation sites by ●, and potential proteoglycan attachment sites by ■. SH represents free cysteines capable of forming interchain disulphide bridges and -S-S- represents interchain disulphide bridges.

Figure 3 The diversity of γδ T cell receptor structures; glycosylation sites are classified as in fig 2. ▶ represents the Cy1 C2 exon; ▲ represents the Cy2 C2 exon copy a; ■ represents the Cy2 C2 exon copy b; □ represents the Cy2 C2 exon copy c.
cell subset in the aetiology of RA and tuberculosis.

**T cell γδ receptor structure**

The TCR δ chain is a polypeptide of 40 kilodalton size with two N-linked glycosylation sites (one high mannose oligosaccharide and one complex oligosaccharide, as determined by endoglycosidase H and peptide N-glycanase F digests). In addition, there are two intramolecular disulphide bridges and potential for one intermolecular disulphide bridge. In T cell lines γ chains exhibit much more structural diversity, being expressed as one of three forms. All three forms have two intramolecular disulphide bridges and at least two potential N-glycosylation sites. The type 1 receptor was first isolated from a peripheral blood lymphocyte cell line. This form expresses two glycoforms of the γ chain, a 40 and a 36 kilodalton form, both forms containing four N-linked glycosylation sites. Treatment with endoglycosidase H reduced both glycoforms to a core peptide of 31 kilodaltons. This indicates differential glycosylation of the two peptides, notably in the distribution of high mannose and hybrid oligosaccharides. The type 1 receptor uses the Cγ1 gene. Within this, one copy of the C2 exon is expressed, containing two consensus N-glycosylation sites and a cysteine residue. This cysteine is able to crosslink with the cysteine of the δ chain forming a disulphide bridge. It is this feature which is the main distinguishing characteristic of the type 1 receptor.

In addition to the type 1 form, two non-intermolecular disulphide linked forms of the receptor have been described. Both forms use the Cγ2 gene for their constant region. The 2abc form expresses three copies of the C2 exon (copies a, b, and c). The resulting polypeptide has a size of 55–60 kilodaltons, which reduces to 40 kilodaltons on endoglycosidase H treatment, again showing differential glycosylation and the probable use of all five consensus N-glycosylation sites. The second non-disulphide linked form of the receptor (type 2bc) expresses two copies of the C2 exon (copies b and c). This results in a glycosylated peptide of 40 kilodaltons, which reduces to 35 kilodaltons, on endoglycosidase H treatment. The amino acid sequence gives a theoretical non-glycosylated size of 34–8 kilodaltons, which together with tunicamycin data (an inhibitor of N-linked glycosylation) shows the glycosylation to be essentially all N-linked. The 5 kilodalton loss in molecular weight indicates that unlike the type 2abc receptor only two of the six possible glycosylation sites are used. The differences in the number of glycans added can be explained by the fact that the protein encoded by the C2α exon copy has an altered secondary or tertiary structure making more of the 2abc γ chain consensus N-glycosylation sites accessible.

Expression of the type 1 receptor is predominant in peripheral blood lymphocytes, although if the type 2 receptors are expressed, types 1 and 2 are usually found in equal proportions. This diversity of T cell receptors is unique to γδ T cells and as yet no distinct functional roles have been attributed to either TCR type, though a report has suggested different cytoskeletal organisation and morphology between the Cγ1 and Cγ2 forms.

Like αβ TCRs, γδ TCRs are associated with the CD3 glycoprotein to facilitate cell signalling and activation. The CD3 γδ peptide cores are the same for both T cell subsets (αβ and γδ). However, glycosylation of the CD3α subunit is different. In αβ T cells the CD3α subunit carries one high mannose and one complex oligosaccharide, whereas in γδ T cells both oligosaccharides are fully processed to the complex form. γδ T cells seems to increase intracellular calcium on γδ TCR recognition as a substrate for protein kinase C activation. Thus if the different glycosylation is important it may function in the regulation of T cell signalling via the γδ TCR/CD3 pathway.

**Role of TCR γδ cells in antigen and superantigen recognition**

Heat shock proteins are the major antigens described to be recognised by γδ T cells. In both a live and heat killed Mycobacterium tuberculosis infection model, γδ T cells produced interferon γ, granulocyte-monocyte colony stimulating factor, IL-3, and tumour necrosis factor α. The nature of the dominant antigen or antigens recognised remains to be characterised fully as up to 400 different components may be involved, but the mycobacterial 65 kilodalton heat shock protein (hsp65) is a major contender and, in particular, a peptide corresponding to residues 181–195. Heat shock proteins are a highly conserved set of proteins with both bacterial and mammalian homologues. The 60 kilodalton family (including M tuberculosis hsp65) is thought to influence the folding and trafficking of mitochondrial proteins.

The 65 kilodalton heat shock protein has been described as a possible antigen in several autoimmune disorders. It has been suggested to be a putative γδ T cell ligand on Daudi cells and a ligand on B cells in lupus nephritis. Rheumatoid arthritis synovial fluid γδ T cells reactive to M tuberculosis were first isolated in 1989, with one clone reacting to the hsp65 of M bovis-BCG. An earlier rat model of adjuvant arthritis had shown the antigenicity of hsp65 peptide 180–188, though this epitope is only present in the bacterial protein and not its mammalian homologue. More recently, both γδ and αβ T cell clones have been isolated, which are stimulated by both mycobacterial and human hsp65, supporting the possibility of hsp65 involvement in autoimmune disorders.

Data for hsp65 as a major antigen of RA are not entirely conclusive. A recent review has shown that recombiant hsp65 from Escherichia coli also contained small quantities of E coli derived antigens. Thus stimulation
Role of protein glycosylation in immune regulation

The initial discovery of molecules including mycobacterial antigens has found unresolved. The role of hsp65 in RA is thus not clear cut and it may be that hsp65 continues the degenerative disease process after stimulation by a different antigen. It is of interest here that heat shock protein may mimic the MHC in a similar manner to that proposed for HIV gp120. No alternative mycobacterial antigens of possible relevance for human RA have yet been characterised. Mycoplasmal mitogens have been proposed in murine models. Superantigen stimulation has been used to explain the increased usage of V814 of αβ T cells in RA. In addition, the superantigen staphylococcal enterotoxin D has been shown to stimulate CD4 Vβ6 αβ T cells to produce rheumatoid factors. There is also evidence that superantigens may stimulate γδ T cells in RA: Vγ9δ2 T cell clones from rheumatoid arthritic synovial fluid have been shown to recognise a short tetanus toxin peptide in the context of MHC class II; this Vγ9δ2 subset may be stimulated by the acetone precipitable fraction of M tuberculosis. There is also evidence for clonal expansion of Vγ9 T cells in the synovial fluid of patients with RA. Similarly, in M tuberculosis infection there is increasing evidence that hsp65 is not a major γδ T cell antigen; and indeed there is some evidence that it may be a poor M tuberculosis antigen in general. The major subset of γδ T cells which expands in response to M tuberculosis is Vγ9, which predominantly express the Cγ1 form of the TCR discussed above. The same Vγ9 subset has been shown to be stimulated by the superantigen, staphylococcal enterotoxin A, which requires the presence of the MHC class II. This is directly analogous to the activation of αβ cells by superantigens. A second Vγ9-seeking superantigen has been isolated from M tuberculosis, again requiring MHC class II. The antigen seems to be a low molecular weight, protease resistant ligand which loses its antigenicity on incubation with the plant lectins UEA1, SBA, and DBA. Thus it has been postulated that the active components are glycosylated determinants having chain terminating α-linked fucose and GalNAcα1-3GalNAc for which these lectins are specific. This is important not only from the point of view of M tuberculosis infection but is the first time a possible non-protein superantigen has been described.

The mechanism by which antigens are presented to γδ T cells remains largely unresolved. Conflicting reports exist on the requirement for β2 microglobulin and hence class I restriction of γδ T cells, though the requirement may depend on the T cell type studied. Restriction by non-classical class I molecules including Qa-1, has been shown.

The initial discovery of γδ T cells in rheumatoid synovial fluid noted the requirement of antigen presenting cells for stimulation by mycobacterial antigens, but this was suggested not to be class I or class II restricted. Later a class II restricted Vγ9δ2 T cell response was characterised against a tetanus toxin peptide. However, only a small percentage of γδ T cells are CD4+, the majority express either CD8α chains only, or are CD4+CD8−, and it has been suggested that CD4+ γδ T cells are more likely to be a fetal T cell subset which has persisted postnatally. In general, γδ and αβ T cells produce a similar range of cytokines, but the proposed minor subset of CD4+ γδ T cells shows production of IL-2 and granulocyte-monocyte colony stimulating factor and low levels of cytototoxic activity, whereas the CD4− γδ T cells exhibit lower levels of cytokine production and a higher cytotoxic activity. Based on this γδ T cells can be subdivided into a CD4+ ‘helper’ subset and a CD4− ‘cytotoxic’ subset.

Conclusion

Study of the possible role of superantigens in autoimmune disorders has recently become the vogue. Care should be exerted in interpretation of the data, however, owing to the complexity of the interactions of the effector molecules involved in immune regulation. An important aspect of these interactions is the role of glycosylation in regulating glycoprotein activity, antigenicity, and recognition.

Protein glycosylation may be involved in several stages in the aetiology of RA: (a) initiating antigen; (b) control of T cell interactions; (c) control of antibody function; (d) microvasculature breakdown; (e) breakdown of synovial gel/water structure; and (f) specific signals to inflammatory cells.

Of crucial importance is the role that proteoglycan sequences have in maintaining the endothelial cell surface barrier, the physicochemical properties of the synovium, and cartilage structure. The release of enzymes which breakdown these proteoglycans could also directly affect lymphocyte immune function by altering cell surface molecules and intracellular trafficking.

Glossary

CD: Clustered Determinant lymphocyte antigens.

Consensus N-glycosylation site: The nitrogen (N) of the Asn amino acid in the sequence Asn-X-Ser/Thr has the potential to be glycosylated, where X is any amino acid except proline.

N-linked protein chains: high mannose complex hybrid

O-linked glycoprotein chains: The hydroxyl group of any Ser or Thr amino acid usually in high Ser/Thr/Pro-containing regions is glycosylated.

Peptidoglycan: Repeating bacterial cell wall polymer of amino acids, muramic acid, and N-acetylglucosamine.

Polysaccharides: In addition to plant (e.g. cellulose) and...
mammalian (e.g. glycoconjugate) polymers, bacteria have very long chain molecules of often complex repeats defined as: (a) capsular antigens (b) O-antigen lipopolysaccharide (LPS).

Proteoglycan: Components of mammalian extracellular matrix, membrane cell surface molecules, secretory granules, and cartilage.

Selectin: Mammalian carbohydrate binding protein also called lecanns (lectin cell adhesion molecules).

Role of protein glycosylation in immune regulation


70 Kaufmann S H E. Heat shock proteins and autoimmunity: fact or fiction? *The evidence from animal models suggesting the involvement of heat shock protein I autoimmune disease is supported by some, but not all, data from autoimmune disease patients.* *Current Biology* 1991; 1: 539–61.


