

Glycosylation and the Immune System

Pauline M. Rudd,^{1*} Tim Elliott,² Peter Cresswell,³ Ian A. Wilson,⁴ Raymond A. Dwek^{1*}

Almost all of the key molecules involved in the innate and adaptive immune response are glycoproteins. In the cellular immune system, specific glycoforms are involved in the folding, quality control, and assembly of peptide-loaded major histocompatibility complex (MHC) antigens and the T cell receptor complex. Although some glycopeptide antigens are presented by the MHC, the generation of peptide antigens from glycoproteins may require enzymatic removal of sugars before the protein can be cleaved. Oligosaccharides attached to glycoproteins in the junction between T cells and antigen-presenting cells help to orient binding faces, provide protease protection, and restrict nonspecific lateral protein-protein interactions. In the humoral immune system, all of the immunoglobulins and most of the complement components are glycosylated. Although a major function for sugars is to contribute to the stability of the proteins to which they are attached, specific glycoforms are involved in recognition events. For example, in rheumatoid arthritis, an autoimmune disease, agalactosylated glycoforms of aggregated immunoglobulin G may induce association with the mannose-binding lectin and contribute to the pathology.

A current model of the immune system classifies the processes by which foreign antigens are eliminated from their hosts into three interrelated systems, depending on whether the initial event takes place in the cytosol, the endocytic pathway or the extracellular space. In the cytosol, antigenic peptides generated by the proteasome are transported by membrane-bound transporters associated with antigen processing (TAP1 and 2) into the ER where they bind to MHC class I molecules (1). In the endocytic pathway, protein antigens are degraded in specialized acidic compartments [MHC class II-containing compartments (MIIC)] (2) into peptides prior to loading onto MHC class II molecules (3, 4). Also in the endocytic pathway, glycosylphosphatidylinositol (GPI) anchors and glycolipid antigens associate with CD1 (2, 5–7). MHC class I and class II molecules, as well as CD1, are recognized by T lymphocytes. In the extracellular space, intact antigens are recognized either by antibody molecules or by the mannose-binding lectin (MBL), both of which subsequently activate the complement system.

The diversity of protein glycosylation plays

an important role in the biosynthesis and biological activity of the glycoproteins involved in antigen recognition. During transport of the glycoproteins through the secretory pathway, the sugar chains undergo successive modifications which are regulated by the glycosylation processing enzymes. This process ensures that each glycoprotein displays the relevant glycan structures for its range of essential functions. The sugars play a role in protein folding and assembly, quality control, ER-associated retrograde transport of misfolded proteins, the generation and loading of antigenic peptides into MHC class I, and also influence the range of antigenic peptides generated in the endosomal pathway for presentation by MHC class II.

By virtue of their size, glycans can shield large regions of the protein surfaces, providing protease protection for immune molecules in serum and secretions and in the T cell synapse, where they also limit nonspecific lateral protein-protein interactions. Glycans located close to the cell membrane, or in heavily glycosylated O-linked domains, can also determine the orientation and location of the binding faces of the proteins to which they are attached (8). In addition, the local three-dimensional structure of the individual protein directs its final glycoform pattern (9) such that a diverse repertoire of oligosaccharides will be displayed on the surface of a single cell. This makes it unlikely that the cell will bind lectins involved in the innate immune system, since these generally require repetitive arrays of specific sugars for functional recognition.

Some of these advantages conferred by glycosylation are exploited by viruses, which use the host glycosylation machinery to assemble their own envelope glycoproteins such that they help avoid immune detection (10). In autoimmune diseases, such as rheumatoid arthritis

(RA) and systemic lupus erythematosus (11, 12), significant changes in the populations of immunoglobulin G (IgG) glycoforms have been noted. In RA, aggregated agalactosyl glycoforms of IgG are specifically recognized by the MBL such that inappropriate activation of the innate immune system may occur (13). Similarly on tumor cells, aberrant glycosylation can expose new epitopes for recognition by the immune system, such as for MUC1 (14), whereas mice deficient in β 1,6 *N*-acetylglucosaminyltransferase V (Mgat5) showed kidney autoimmune disease and enhanced delayed-type hypersensitivity and increased susceptibility to experimental autoimmune encephalomyelitis (15).

Glycosylation and Protein Folding

Almost all of the key molecules involved in antigen recognition and the orchestration of the subsequent events are glycoproteins consisting of two or more subunits. Some, such as the T cell receptor (TCR) and CD3, are assembled in the ER into multimolecular complexes. The proper folding and controlled assembly of many newly synthesized glycoproteins requires them to engage in a series of coordinated interactions with chaperones and enzymes, through the attachment of a common oligosaccharide precursor, GlcNAc₂Man₉Glc₃ (Fig. 1A), to N-linked glycosylation sites. This sugar precursor is rapidly processed to GlcNAc₂Man₉Glc₁ (Fig. 1B), which can bind two chaperones, the membrane-bound calnexin (Clx) and soluble calreticulin (Clr) (Fig. 1B) (16, 17).

The lectin-like interactions of Clx and/or Clr with nascent glycoproteins provide access to a folding pathway (18), allow the recruitment of the thiol oxidoreductase, ERp57, and assist the assembly of subunits. Clx and Clr are also involved in the loading of antigenic peptides onto MHC class I from complexes of TAP and tapasin. In their role as quality control factors, Clx and Clr retain unfolded glycoproteins in the ER until they are correctly folded and assembled, an event which is signaled by the permanent removal of the terminal glucose residue by glucosidase II (Fig. 2). The folded glycoprotein, or the assembled multimolecular complex, is then transferred to the Golgi apparatus where the oligomannose sugars may be further processed. Misfolded or unassembled subunits are reglucosylated by uridine 5'-diphosphate-glucose:glycoprotein glucosyltransferase. Reglucosylation of unfolded proteins allows them to rebound Clx and enter a cyclical pathway until they either achieve their correctly folded structure and are released or they are targeted

¹The Glycobiology Institute, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. ²Cancer Sciences Division, University of Southampton Medical School, CRC Medical Oncology Unit, Level F, Centre Block, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK. ³Howard Hughes Medical Institute, Section of Immunobiology, Yale University School of Medicine, 310 Cedar Street, New Haven, CT 06510, USA. ⁴Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

*To whom correspondence should be addressed. E-mail: pmr@glycob.ox.ac.uk (P.M.R.) and raymond.dwek@exeter.oxford.ac.uk (R.A.D.)

for retrograde transport and degradation (Fig. 2) (19, 20).

Glycosylation and ER-Associated Degradation

The TCR complex is composed of at least six protein subunit components (α , β , γ , δ , ϵ , ζ); neither partial complexes nor individual subunits can be expressed by themselves on the protein surface. Misfolded or unassembled glycoproteins can be eliminated directly from the ER in a process known as ER-associated degradation (ERAD) that is initiated by the action of mannosidase I. This enzyme removes a single mannose residue from the oligosaccharide precursor so that it carries the glycan motif, GlcNAc₂Man₈ (Fig. 2) (19, 20). If the glycoprotein is still misfolded or unassembled, the GlcNAc₂Man₈ glycan is reglycosylated. Compared with GlcNAc₂Man₈Glc₁, GlcNAc₂Man₈Glc₁ is a poor substrate for glucosidase II and the glycoprotein readily rebinds Clx (21–23). The misfolded or unassembled glycoprotein subunit, attached to Clx, may then be released from the ER through the Sec61 translocation channel into the cytosol where it is degraded by the proteasome. The release of the polypeptide into the cytoplasm may require ubiquitination of the cytosolic tail (24) or that of Clx (25). Denatured polypeptides in the cytosol are degraded by the 20S proteasome while protein substrates, such as unassembled subunits, are ubiquitinated, unfolded, and processed by the 26S proteasome. Antigenic peptides from internally synthesized viral proteins are generated by the protease activity of the proteasome (26).

Glycosylation and Antiviral Strategies

Enveloped viruses, such as human immunodeficiency virus (HIV), can evade immune recognition by exploiting the host glycosylation machinery to protect potential protein antigenic epitopes (10). Enveloped viruses also use the host secretory pathway to fold and assemble their often heavily glycosylated coat proteins. A possible antiviral strategy involves the use of glycosylation inhibitors to interfere with the folding of viral envelope proteins, such as for hepatitis B virus (HBV) and the HIV.

HBV has three envelope glycoproteins (L, M, and S) which are derived from alternative translations of the same open reading frame. All three surface antigens contain one common N-linked glycosylation site, Asn¹⁴⁶ (about 50% occupied), but the M protein contains a second at Asn⁴ (100% occupied) that is responsible for the interaction of M protein with Clx. When glucosidases I and II are inhibited by *n*-butyl deoxynojirimycin (*n*-BuDNJ), M protein not only remains hyperglycosylated, but is also misfolded (27). The inability of the virus to incorporate misfolded M protein into the viral coat has a destabilizing effect, and secretion of the virus is pre-

vented. However, subviral particles (which do not contain DNA) are secreted and, in the presence of the inhibitor, consist only of S glycoproteins, rather than M and S (27).

The major component of the HIV-1 viral envelope is gp160, a noncovalently linked complex of gp120 and gp41. When gp120 is expressed in the presence of *n*-BuDNJ, the V1/V2 loops (residues 128 through 199) are misfolded. In contrast to HBV, where misfolding of the M-protein affects the assembly of the viral coat, misfolding of the V1 and V2 loops of gp120 does not affect the assembly of the viral envelope of HIV. This is because these loops do not form part of the surfaces which contact each other in the gp120 trimeric complex (28). Binding of gp160 to CD4 is also unaffected by the presence of *n*-BuDNJ in the culture medium; however, the conformational change which follows the binding event involves the rearrangement of the V1/V2 loops to release gp120 from gp160. This rearrangement cannot take place (29), and, as a result, gp41 cannot mediate the fusion event, precluding entry of the virus into the cell.

Glycosylation and the Assembly of MHC I: Peptide Complexes

Mature MHC class I (MHC I) consists of three subunits: the heavy chain, which is a transmembrane glycoprotein; the small soluble nonglycosylated protein, β_2 -microglobulin (β_2 M); and an antigenic peptide that is required for the subsequent transport of MHC I to the cell membrane. Class I assembly [Fig. 2, reviewed in (1)] requires multiple coordinated intra- and intermolecular events to ensure the continuous reporting of cellular contents to cytotoxic T lymphocytes. In the ER, unassembled heavy chains interact initially with membrane bound Clx through the Man₆GlcNAc₂Glc₁ sugar which is attached (in human HLA) to Asn⁸⁶ in the early stages of glycan processing (Fig. 1). Concomitant with β_2 M association with heavy chain, the sugar is released from Clx and binds to Clr. One or both of these chaperones recruits the thiol

oxidoreductase (Erp57) into the complex, facilitating the formation of the intrachain disulfide bonds of the class I heavy chain (30–32). The assembled class I- β_2 M dimer, together with associated Clr and Erp57, are associated with a larger complex that also contains the TAP transporter and the transmembrane glycoprotein tapasin (33). Tapasin provides the bridge that links MHC I, Clr, and Erp57 to the TAP heterodimer (34).

Peptides that are destined for MHC I binding are generated by proteolysis of cytoplasmic proteins by the proteasome. The peptides are then translocated in an ATP-dependent fashion from the cytosol into the ER by the TAP heterodimer. Those peptides with the appropriate sequence may undergo further NH₂-terminal trimming prior to binding to TAP-associated MHC I. Peptide binding induces dissociation of the class I- β_2 M dimer from the tapasin-TAP complex, as well as from Clr and Erp57, allowing glycan maturation and transport of the assembled MHC I from the ER (Fig. 2). The precise roles of tapasin, Clr, and Erp57 in the late stages of MHC I assembly are poorly understood.

Generation of Peptides from Glycosylated T Cell Antigens

Class I MHC-restricted peptide antigens are generated primarily in the cytosol. Proteins are targeted for degradation by ubiquitination and degraded by the 26S proteasome and/or further broken down into smaller peptides by the 20S proteasome modified by the 11S regulator or PA28 complex (35) and by other downstream enzymes in the cytosol, such as leucine aminopeptidase (36), puromycin-sensitive aminopeptidase and bleomycin hydrolase as well as unidentified sugar-trimming enzymes within the ER (37).

The 20S proteasome is a cylindrical assembly approximately 150 Å in height and 110 Å in diameter. The four rings of the α and β subunits of the proteasome encase a channel that widens

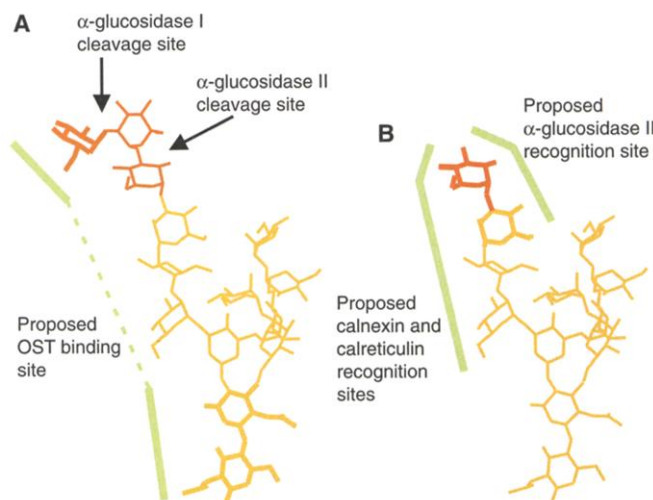


Fig. 1. Carbohydrate structure and its role in protein folding. Nuclear magnetic resonance (NMR) solution structures of (A) GlcNAc₂Man₉Glc₃ and (B) GlcNAc₂Man₉Glc₁ showing the cleavage sites for α -glucosidase I and II and the proposed recognition sites of the oligosaccharyl transferase (OST), Clx, Clr, and of α -glucosidase II for GlcNAc₂Man₉Glc₁.

into three large cavities in which two constrictions that are only 13 Å in diameter form a passage from the cytosol to the inside of the cylinder. A typical sugar, such as a complex biantennary sugar containing 12 monosaccharide residues, is about 30 Å by 10 Å by 10 Å. Thus prior to entry into the proteasome, the sugars are removed from glycosylated polypeptides by a proximal glycanase, cytosolic peptide *N*-glycoamidase (38). A role for a cytosolic *N*-glycanase is supported by the identification of a peptide derived from tyrosinase which is transported into the ER bound to TAP (39). Tyrosinase contains six *N*-glycosylation sites. A tyrosinase peptide, YMNGTMSQV that includes the glycosylation site Asn³⁷¹, is presented to MHC I as the converted peptide, YMDGTMSQV (40). The action of peptide *N*-glycanase, which removes *N*-linked glycans by cleaving the glycan from Asn at the *N*-glycosidic linkage, also converts the Asn residue to Asp (41).

Glycosylation Protects Cleavage Sites in Class II Antigen Processing

Class II MHC molecules draw on a different pool of peptides that are generated in the endocytic pathway from proteins that have either

been diverted to endosomes from the secretory pathway or internalized by receptor-dependent or -independent endocytosis. Several endosomal and lysosomal proteases have been implicated in generating class II restricted peptides, including the endopeptidases cathepsin D, L, and S and the exopeptidases cathepsin A, B, and H (3). A carbohydrate-dependent, HLA class II-restricted human T cell response to the bee venom allergen phospholipase A2 has been described in allergic patients (42). For microbial tetanus toxin, the dominant processing enzyme is an asparagine-specific cysteine endopeptidase (AEP) (4). *N*-glycosylation blocks the action of AEP, indicating that glycans can protect self proteins and some viral glycoproteins from degradation. In contrast, microbial organisms which do not contain *N*-linked sugars can be fully processed to antigenic peptides.

Glycosylation and T Cell Recognition of Antigen Presenting Cells

Multiple events and stages are involved in the T cell recognition of antigen presenting cells (APCs). These include the formation of a junction between the two cells known as the immu-

nological synapse (43, 44), recognition of antigenic peptide-loaded MHC molecules by the TCRs (Fig. 3) and signal transduction. Sugars are likely to be important in many of these processes. For example, receptor molecules of a similar length are recruited into a specialized junction between the T cell and APC (45). Oligosaccharides located on the membrane proximal domains tend to be conserved across species and seem likely to restrict the orientations of the cell adhesion molecules CD2 and CD48 contributing to the alignment of the opposing cell surfaces (46) [(Fig. 3)].

In addition, oligosaccharides may play a role in the transport of peptide-loaded MHCs (pMHCs) into the center of the junction where they are recruited by TCRs (Fig. 3). In vivo, this movement takes place over a 5-min time scale and may involve serial engagements of pMHC with the TCR, mediated by actin-based transport mechanisms (44). The size and location of the glycans prevent nonspecific protein-protein interactions, such as the aggregation of TCRs on the membrane, as well as possibly limiting the geometry of the interactions of the proteins in the central clusters (47). Extensive glycosylation makes it unlikely that specific interactions of the protein components of $\alpha\beta$ TCR-pMHC

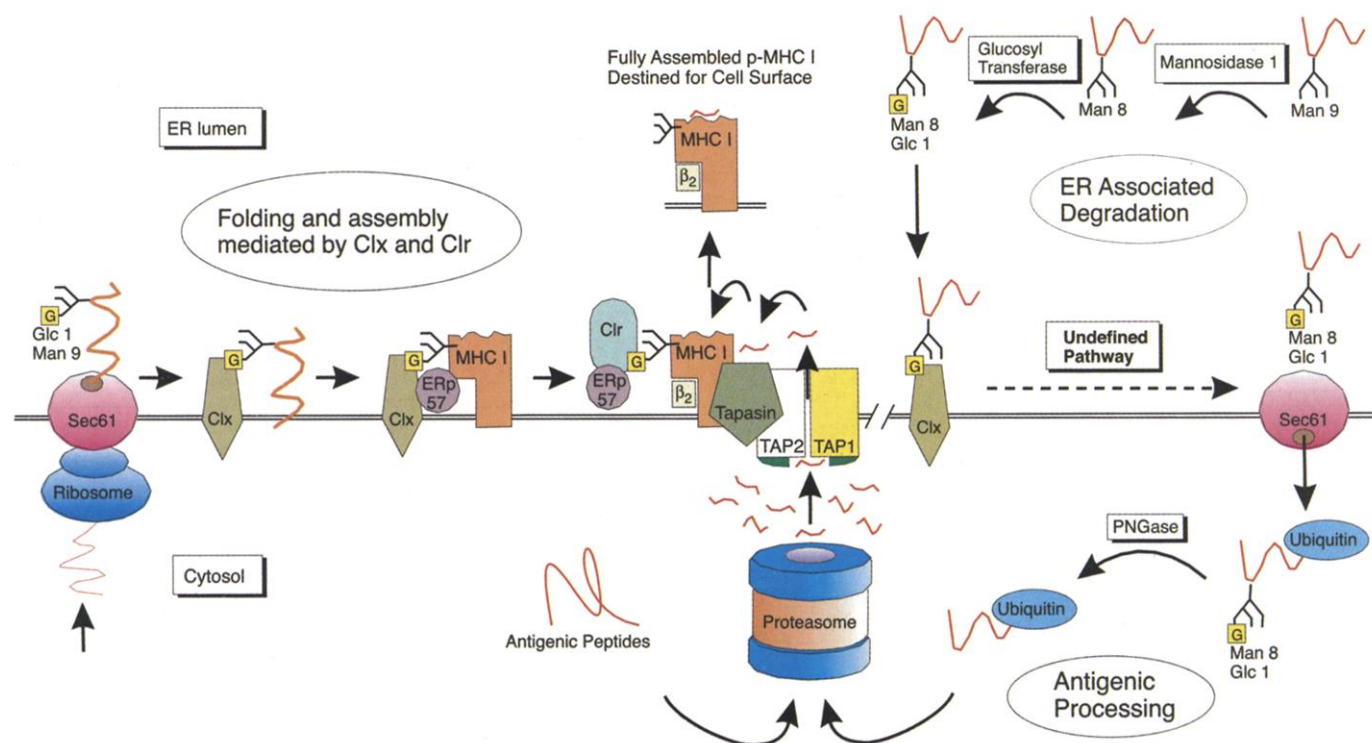


Fig. 2. A proposed model for the folding, and peptide loading of MHC I and the degradation of unfolded glycoproteins from the ER. Nascent glycoproteins are translated across the lumen of the ER and where they may become immediately associated with membrane-bound calnexin (Clx). Association of glycoproteins with Clx is dependent upon the presence of a single terminal glucose residue bound to the GlcNAc₂Man₉ glycan precursor. In this example, interaction of Clx is shown with newly synthesized MHC class I glycoproteins. The thiol oxidoreductase ERp57 also binds and facilitates disulfide bond formation, which facilitates the interaction with β_2 M. This induces dissociation of class I from Clx, which

is replaced by the soluble chaperone calreticulin (Cir). MHC class I- β_2 M dimers become loaded with peptide while associated with the TAP transporter, an interaction mediated by tapasin. The stable peptide-MHC complex is released and transported to the cell surface. Misfolded glycoproteins are recognized as such by enzymes, which convert the GlcNAc₂Man₉ structure to GlcNAc₂Man₈Glc₁ in a stepwise manner. The GlcNAc₂Man₈Glc₁ structure is recognized by Clx, and the glycoprotein is subsequently targeted for proteolytic degradation through the Sec 61 pathway. It is not yet proven whether the terminal Glc remains attached during this process.

can form laterally on the cell surface. In support of this notion, no dimers or specific higher order oligomers of soluble TCR-pMHC molecules have been found in the crystal lattices of mouse (48, 49) and human (50, 51) class I TCR-pMHC complexes. Within the synapse, specific protein-protein contacts must arise either from interaction of the CD4 or CD8 co-accessory molecules with the TCR-pMHC or from lateral interactions of the CD3 components.

During signal transduction, the sugars can also play a role in stabilizing the individual molecules (52) in the complexes in the synapse by protecting them from the action of proteases during T cell engagement in a process which, for class II APCs, may take several hours.

MHC I interacts with TCRs on CD8⁺ T cells, whereas MHC II is recognized by CD4⁺ T cells. Although CD4 and CD8 are structurally very distinct molecules, they interact with structurally homologous sites on their MHC ligands. The binding site of CD4 is held at the appropriate distance from the T cell membrane by protein domains. In CD8, the globular head containing the pMHC binding site is tethered to the T cell membrane by an extended polypeptide stalk that contains four O-linked sugars. These sugars are expected to confer some rigidity to this extended peptide chain, to protect it from digestion by proteases, and to inhibit nonspecific interactions of the stalk with the TCR.

Recognition of Glycopeptide Epitopes by T Cells

Around 0.5% of peptides that are bound to MHC class I molecules on the surface of normal cells carry O-linked *N*-acetylglucosamine (GlcNAc) residues (53), which are presumably derived from O-glycosylated cytosolic and nuclear proteins (54). In contrast, peptides carrying N-linked GlcNAc residues are probably not presented by MHC class I (55), although *N*-GlcNAc-substituted peptides are naturally presented on the surface of cells bound to MHC class II molecules (56).

Both CD8⁺ and CD4⁺ T cells can recognize glycopeptides carrying mono- (55, 57–59) and disaccharide substitutions (58, 60) in a classical MHC-restricted way. In these cases, specificity of the T cell depends on contacts with both the glycan and the peptide moieties of the antigenic glycopeptide (61, 62). In the context of the MHC molecule, the crystal structures of three MHC class I–glycopeptide complexes (61, 63) clearly show that recognition of the additional carbohydrate moieties is within the physicochemical capability of an $\alpha\beta$ TCR, and certain motifs in the TCR CDR3 β region recur in T cells recognizing *O*-GlcNAc-substituted glycopeptides (64, 65). In addition, titration of sugars of varying lengths has shown that the $\alpha\beta$ TCRs can recognize up to two sugars plus an

extensive homocysteine linker that itself would correspond in length to approximately two sugars (63). Longer sugar chains can elicit $\gamma\delta$ TCR responses which are not MHC restricted (63). That such large antigens can be presented to T cells has been confirmed by the structure of a rat MHC class I molecule that is bound to a 13-residue peptide oligomer (13-mer) (66). The peptide bulges out extensively up to 13 Å above the peptide-binding groove on the MHC class I molecule. The immunodominant arthritogenic MHC class II-restricted T cell epitope in collagen-induced arthritis in mice is a glycopeptide with one or two hydroxylysine residues substituted with galactose (65), and T cells recognize carbohydrates on type II collagen (67). The presence of larger glycans has been shown to lead to the loss of MHC-restricted recognition of glycan by glycopeptide-specific T cells, and the tendency to induce $\gamma\delta$ T cells which recognize the nominal glycan antigen, regardless of whether it is tethered to the APC surface via an MHC-binding peptide or via a lipid tail (58).

Other studies have also indicated that glycopeptides containing more complex glycans can be recognized by classical $\alpha\beta$ T cells in an MHC class I (68) and class II-restricted

manner. Moreover, in a detailed study of the CD1-restricted recognition by both CD8⁺ and CD4⁺ $\alpha\beta$ T cells of the ganglioside GM1, the pentasaccharide moiety was shown to be the minimal structure for forming the epitope. Whether the entire pentasaccharide can be accommodated by the central TCR antigen combining region is unknown, but the crystal structure of the disaccharide-substituted glycopeptide:MHC complex (63) suggested that the longer oligosaccharides may be accommodated outside the normal "footprint" of the TCR.

Glycolipids and GPI Anchors Are Presented by CD1

CD1 [reviewed in (2, 5–7)] is a nonpolymorphic, β_2 M-associated cell surface glycoprotein with antigen-presenting properties. In contrast to MHC class I and II molecules, the five isoforms of CD1 (CD1a, CD1b, CD1c, CD1d, CD1e), do not present peptides to TCR but lipids or glycolipids which are acquired within the acidic endosomal MHC compartments (2). While CD1a, CD1b, and CD1c present glycolipids derived from the cell walls of

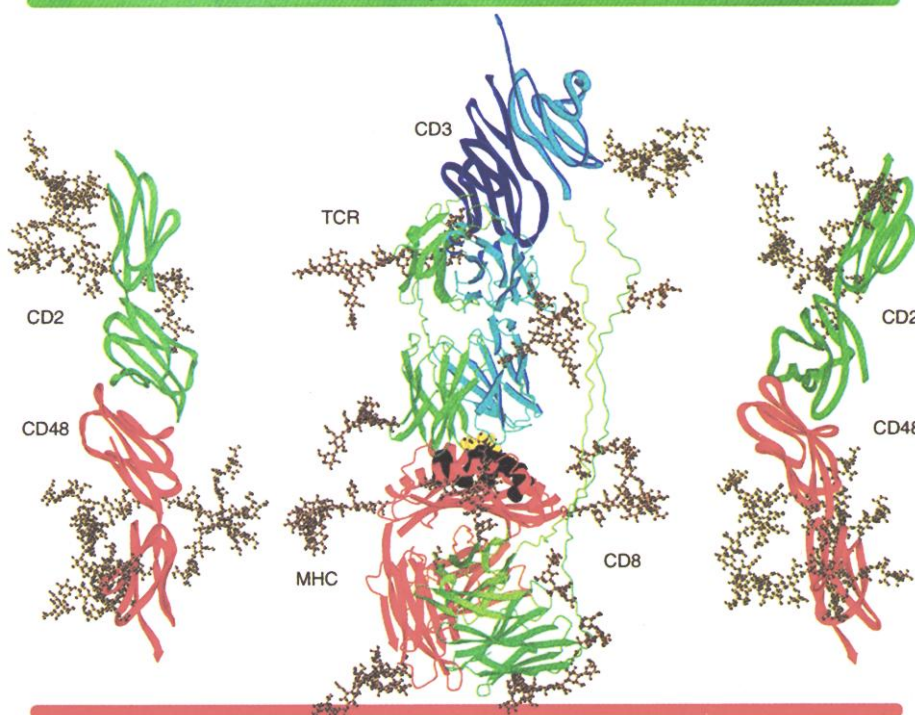


Fig. 3. The immunological synapse. A hypothetical model of the TCR-CD3-CD8 complex with pMHC and the CD2-CD48 (green and red) cell adhesion pair. The protein NMR or x-ray structures are all derived from the recombinant soluble forms of the proteins and currently no information is available on which to model the membrane proximal regions of the proteins. The cell membranes are represented by the green (TCR) and red (APC) lines. The $\alpha\beta$ TCR (green and blue), pMHC, and CD8 (green and red) structures are taken from crystal structures of the 2C TCR-H-2K^b complex (48) and CD8 $\alpha\alpha$ -HLA-A2 complex (85). A theoretical model (86) of the CD3 ϵ (purple) and CD3 γ postulated dimer (blue) has been placed at the base of the $\alpha\beta$ TCR (87) to emphasize its size and location. The location of the corresponding CD3 γ - ϵ complex is also not known, but can be modeled onto the $\alpha\beta$ TCR [see figure 11 in (47)]. CD48 (88) and CD2 (89) are modeled with glycans analyzed in (90). CD8 (85) contains *N*- and *O*-glycans analyzed by (91) Mouse MHC class I (H-2D^b) (red) is complexed with a nine-residue glycopeptide (black and yellow) (61). Oligosaccharides are colored black and yellow.

mycobacteria, such as *M. leprae* and *M. tuberculosis* (5), murine CD1d has been shown to present GPI anchors (69) and human CD1e, isoprenoids (70). The exact details of the molecular interactions which allow bulky glycolipids to be accommodated in the presentation groove of CD1 are not known. However, significant insights have been gained from the crystal structure of murine CD1d that revealed a deep, voluminous binding groove consisting of two very hydrophobic pockets (71). On the basis of this structure, it is predicted that the long fatty acid tails of the glycolipid bind in these pockets with the hydrophilic glycan moiety oriented so that it protrudes from the binding groove making contact with the TCR (5).

Antibody-Antigen Recognition by Immunoglobulins in the Extracellular Space

In the extracellular space, two forms of antigen-binding are currently recognized. The first is the recognition of repetitive arrays of sugars, for example on bacteria or yeast, by the MBL. The second is the classical recognition of antigen by immunoglobulins.

Structural roles for glycosylation in IgG are well documented. In human IgG Fc (Fig. 4A), crystallographic studies (72) have shown that the two C_H2 domains do not interact by protein-protein contacts, but instead through an interstitial region that is formed by oligosaccharides, attached at Asn²⁹⁷ on each heavy chain. Protein-oligo-

saccharide and oligosaccharide-oligosaccharide interactions play a role in maintaining the relative geometry of the C_H2 domains (73), consistent with less efficient binding of both aglycosylated and degalactosylated IgG to the Fc γ receptor (72–74).

By contrast, IgA1 (Fig. 4A) does not contain an interstitial space large enough to accommodate the Fc sugars and the Asn²⁵⁸ side chains are fully exposed and point out into solution (74). Their ER and Golgi processing is unhindered by interaction with the protein surface, and most serum IgA1 N-linked sugars, which include tri- and biantennary structures, are fully galactosylated and sialylated. The hinge region of IgA1, which contains five O-glycosylation sites, is resistant to many com-

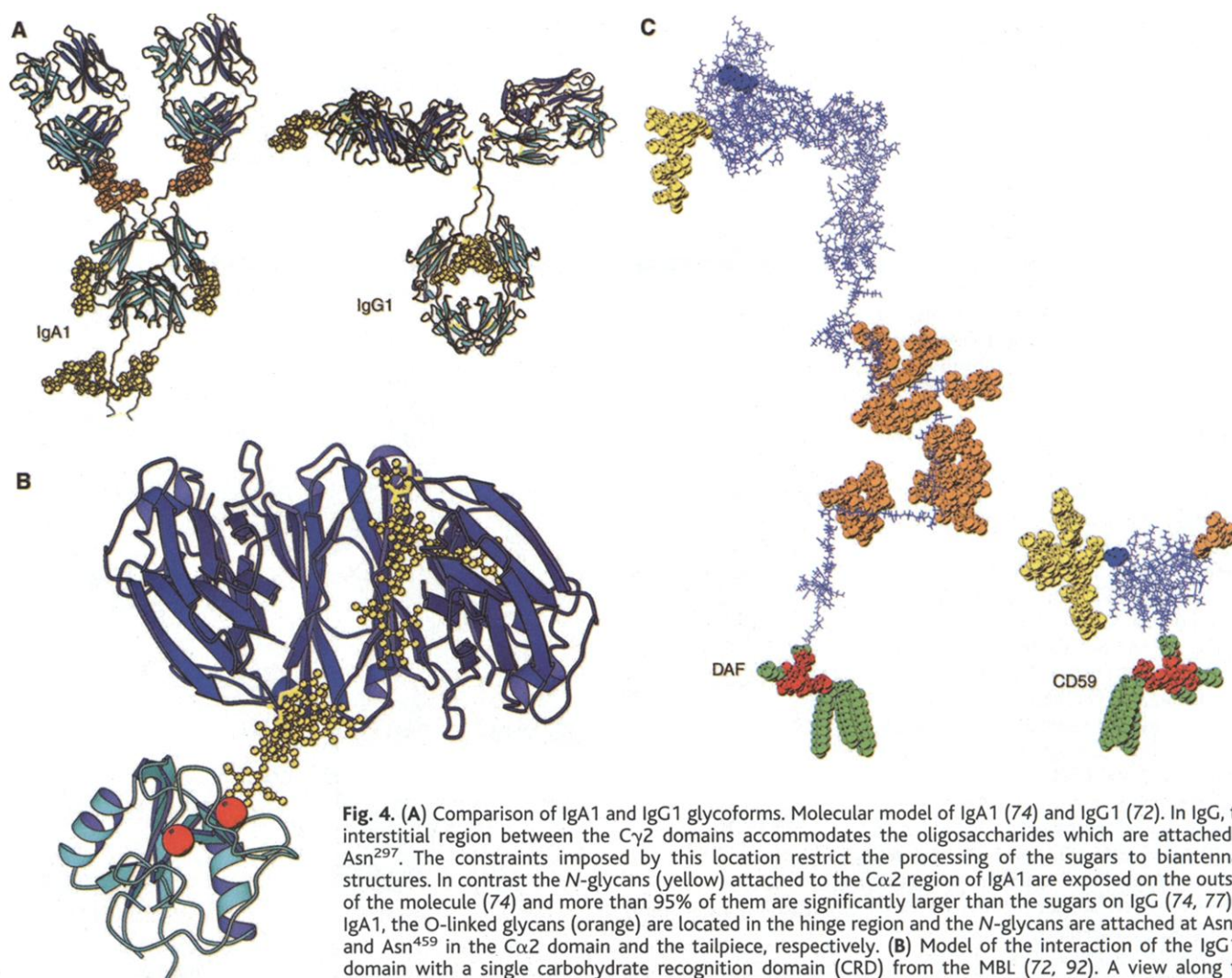


Fig. 4. (A) Comparison of IgA1 and IgG1 glycoforms. Molecular model of IgA1 (74) and IgG1 (72). In IgG, the interstitial region between the C_H2 domains accommodates the oligosaccharides which are attached at Asn²⁹⁷. The constraints imposed by this location restrict the processing of the sugars to biantennary structures. In contrast the N-glycans (yellow) attached to the $C\alpha2$ region of IgA1 are exposed on the outside of the molecule (74) and more than 95% of them are significantly larger than the sugars on IgG (74, 77). In IgA1, the O-linked glycans (orange) are located in the hinge region and the N-glycans are attached at Asn²⁵⁸ and Asn⁴⁵⁹ in the $C\alpha2$ domain and the tailpiece, respectively. (B) Model of the interaction of the IgG1Fc domain with a single carbohydrate recognition domain (CRD) from the MBL (72, 92). A view along the pseudo-C2 axis of the Fc domain (from the hinge toward the COOH-terminus). The nonreducing terminal galactose residues of the left-hand oligosaccharide chain have been removed and the oligosaccharide has been allowed to move relative to the protein by free rotation around the Asn²⁹⁷ $C\alpha$ - $C\beta$ and $C\beta$ - $C\gamma$ bonds. The 6-arm nonreducing terminal GlcNAc residue has been docked to the CRD saccharide binding site (72, 92). The principal interaction involves the chelation of the two *cis* 3-4- hydroxyl groups of the GlcNAc residue from the Fc to the Ca^{2+} ion of the CRD. (C) GPI-linked cell surface glycoproteins that inhibit cell lysis by the membrane attack complex. Decay accelerating factor (DAF) is based on the model of (8) with a biantennary sialylated glycan on domain 1. The COOH-terminal tail, containing 12 O-linked glycans and a GPI anchor, is modeled in an extended conformation. CD59 is based on the NMR-solution structure (93). The glycan anchor is modeled with a trimannosyl core, an ethanolamine bridge at Man3 and additional ethanolamine groups at Man1 and Man2. Two lipids are attached to inositol via phosphate and the third is attached directly to the inositol ring through an ester linkage. A trisialylated, tetra-antennary complex N-glycan is shown attached to Asn¹⁸ and the O-glycan, NeuNAc α 2-3Gal β 1-3GalNAc, is attached to Thr⁵¹ (84).

mon proteases (75). The sugars, which on serum IgA1 mainly consist of sialylated GalGalNAc (74, 76, 77), shield large areas of the hinge region (Fig. 4A) and produce an extended rigid structure (78).

In IgG, the constraint imposed by the inter-chain location of the sugars within the C_H2 domains restricts them to complex biantennary structures. A further constraint arises when a galactose residue is added to the terminal GlcNAc residue on the α 1,6 arm during the biosynthetic process of elongating the sugar chain. This galactose residue forms a tight interaction with a lectin-like binding pocket on the antibody (Lys²⁴⁶, Glu²⁵⁸, and Thr²⁶⁰), which results in further immobilization of the glycan chain. However, sugars that are first galactosylated on the α 1,3 arm (G1 α 1,3) can more readily become digalactosylated, as reflected in the final ratio of G1 α 1,3 to G1 α 1,6 glycans, which is about 1:4 (79). Another factor that significantly influences the distribution of glycoforms in IgG is the level of galactosyl transferase that transfers galactose to terminal N-acetylglucosamine residues; its levels determine the relative proportions of zero-, mono-, and digalactosylated IgG (IgG0, IgG1, and IgG2 respectively). The arm-specific distribution of monogalactosylated IgG between the G1 α 1,6 and G1 α 1,3 glycoforms continues to be determined by the protein-sugar interactions.

In rheumatoid arthritis, galactosyl transferase levels are decreased such that IgG0 is increased. The lack of terminal galactose residues on the α 1,6 arm result in the exposure of the terminal GlcNAc residues to the MBL.

Recognition of Repetitive Arrays of Sugars by the MBL

Sugar-protein interactions that involve single monosaccharide residues have low binding affinities. In general, to trigger biological events, multiply presented sugar ligands are required to interact with multivalent receptors. Thus, in carbohydrate-lectin-type interactions, or in carbohydrate antigen-antibody interactions, several monosaccharide moieties must normally be presented in the correct conformation in order to bind to the receptor with high affinity (80). Alternatively, in C-type lectins, clustering of multiple copies of the same sugar epitope on a surface in a particular geometry, can provide a multivalent surface for recognition by the multiple carbohydrate recognition domains (81). In the extracellular space, many bacteria and yeasts present arrays of mannose residues on their surfaces to the MBL (81) that leads to activation of the complement pathway and cell lysis.

IgG Glycoforms Can Activate Complement Through Binding MBL

IgG glycans may provide an additional route to inflammation in rheumatoid arthritis by activating

the classical complement pathway through binding to MBL (Fig. 4B) (13). Specific IgG0 glycoforms become ligands for MBL and are prevalent in the serum, synovial fluid, and synovial tissue of patients with rheumatoid arthritis; levels of activity of MBL and IgG0 correlate with time of disease onset (82). The mechanism by which specific IgG sugars initiate complement activation in rheumatoid arthritis is expected to be particularly important in the synovial cavity where IgG0 and MBL levels are elevated and IgG is clustered so that the G0 glycans are multiply presented. Interestingly, IgG0 glycoforms are continuously taken up by macrophages through the mannose receptor (83) and that may have implications for tolerance, as MHC II peptide antigens are generated in the endosomal pathway.

Glycosylation and Complement-Mediated Cell Lysis

Finally, both classical and alternative complement pathways terminate in formation of the membrane attack complex (MAC) on the cell surfaces of bacteria and other pathogens that leads to cell lysis. Host cells are normally protected from destruction through inhibitors of the complement pathway, such as the GPI-anchored glycoproteins decay accelerating factor (DAF; CD55), which destabilizes the C3 convertase components C3b/Bb and C4b2a, and CD59 (Fig. 4C), which binds C8 and/or C9 preventing formation of a fully-assembled MAC complex. The rigid O-glycosylated domain of DAF positions the binding site at an appropriate distance from the membrane for interaction with C3b, while the heterogeneous array of CD59 glycoforms (more than 130) (84) make it unlikely that the mobility of this GPI-anchored glycoprotein will be restricted by clustering into regular arrays on the cell surface. Excessive stimulation of the complement pathway in the absence of pathogens, such as may occur in RA, may saturate the inhibitor system, leading to inappropriate cell lysis and inflammation.

Further Perspectives

Glycoproteins are key components of the immune system effectors. Their oligosaccharides are important in synthesis, stability, recognition and regulation of the proteins themselves and in many of their diverse interactions. Establishing the extent and variability of glycosylation in the context of the three-dimensional structure of the proteins to which they are attached will lead to further insights into immune processes in health and disease. It is already becoming evident that the manner in which oligosaccharides are presented provides a mechanism for distinguishing self from non-self. Pathogens which present specific oligosaccharides as repetitive arrays can, if these arrays have a suitable geometry, activate multivalent receptors. One example

is the MBL, which is involved in the recognition and subsequent elimination of pathogens. Others may yet be discovered. Sugars attached to mammalian glycoproteins (self) will not normally be presented in an appropriate homogeneous geometrical array to activate the immune response. The ubiquity and diversity of protein glycosylation is not a paradox, but consistent with functional rules.

References and Notes

1. P. Cresswell et al., *Immunol. Rev.* **172**, 21 (1999).
2. M. Sugita et al., *Clin. Immunol. Immunopathol.* **87**, 8 (1998).
3. H. A. Chapman, *Curr. Opin. Immunol.* **10**, 93 (1998).
4. B. Manoury et al., *Nature* **396**, 695 (1998).
5. S. A. Porcelli et al., *Immunol. Today* **19**, 362 (1998).
6. R. S. Blumberg et al., *Immunol. Rev.* **147**, 5 (1995).
7. A. Melian et al., *Curr. Opin. Immunol.* **8**, 82 (1996).
8. L. Kuttner Kondo et al., *Protein Eng.* **9**, 1143 (1996).
9. P. M. Rudd et al., *Glycobiology* **9**, 443 (1999).
10. J. N. Reitter, R. E. Means, R. C. Desrosiers, *Nature Med.* **4**, 679 (1998).
11. R. B. Parekh et al., *Nature* **316**, 452 (1985).
12. M. Watson, P. M. Rudd, M. Bland, R. A. Dwek, J. S. Axford, *Arthritis Rheum.* **42**, 1682 (1999).
13. R. Malhotra et al., *Nature Med.* **1**, 237 (1995).
14. V. Apostolopoulos et al., *Cancer Lett.* **90**, 21 (1995).
15. M. Demetriou, M. Granovsky, S. Quaggin, J. Dennis, *Nature* **409**, 733 (2001).
16. J. J. Bergeron et al., *Adv. Exp. Med. Biol.* **435**, 105 (1998).
17. Y. Saito et al., *EMBO J.* **18**, 6718 (1999).
18. E. S. Trombetta, A. Helenius, *Curr. Opin. Struct. Biol.* **8**, 587 (1998).
19. J. L. Brodsky et al., *J. Biol. Chem.* **274**, 3453 (1999).
20. E. D. Werner, J. L. Brodsky, A. A. McCracken, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13797 (1996).
21. C. M. Cabral, P. Choudhury, Y. Liu, R. N. Sifers, *J. Biol. Chem.* **275**, 25015 (2000).
22. Y. Liu et al., *J. Biol. Chem.* **274**, 5861 (1999).
23. ———, *J. Biol. Chem.* **272**, 7946 (1997).
24. M. de Virgilio, H. Weninger, N. E. Ivesa, *J. Biol. Chem.* **273**, 9734 (1998).
25. D. Qu, J. H. Teckman, S. Omura, D. H. Perlmuter, *J. Biol. Chem.* **271**, 22791 (1996).
26. D. Stock et al., *Curr. Opin. Biotechnol.* **7**, 376 (1996).
27. A. Mehta et al., *FEBS Lett.* **430**, 17 (1998).
28. P. D. Kwong et al., *J. Virol.* **74**, 1961 (2000).
29. P. B. Fischer et al., *J. Virol.* **70**, 7143 (1996).
30. J. A. Lindquist, O. N. Jensen, M. Mann, G. J. Hammerling, *EMBO J.* **17**, 2186 (1998).
31. E. A. Hughes, P. Cresswell, *Curr. Biol.* **8**, 709 (1998).
32. N. A. Morrice, S. J. Powis, *Curr. Biol.* **8**, 713 (1998).
33. B. Ortmann et al., *Science* **277**, 1306 (1997).
34. B. Sadasivan et al., *Immunology* **5**, 103 (1996).
35. K. L. Rock, A. L. Goldberg, *Annu. Rev. Immunol.* **17**, 739 (1999).
36. J. Beninger, K. L. Rock, A. L. Goldberg, *J. Biol. Chem.* **273**, 18734 (1998).
37. T. Elliott, A. Willis, V. Cerundolo, A. Townsend, *J. Exp. Med.* **181**, 1481 (1995).
38. T. Suzuki, Q. Yan, W. J. Lennarz, *J. Biol. Chem.* **273**, 10083 (1998).
39. T. Elliott, *Adv. Immunol.* **63**, 47 (1997).
40. J. C. Skipper et al., *J. Exp. Med.* **183**, 527 (1996).
41. T. Suzuki et al., *Biochem. Biophys. Res. Commun.* **194**, 1124 (1993).
42. T. Dudler, F. Altmann, J. M. Carballido, K. Blaser, *Eur. J. Immunol.* **25**, 538 (1995).
43. P. A. van der Merwe et al., *Semin. Immunol.* **12**, 5 (2000).
44. A. Grakoui et al., *Science* **285**, 221 (1999).
45. C. R. Monks, H. Kupfer, I. Tamir, A. Barlow, A. Kupfer, *Nature* **385**, 83 (1997).
46. M. L. Dustin et al., *J. Biol. Chem.* **272**, 30889 (1997).
47. P. M. Rudd et al., *J. Mol. Biol.* **293**, 351 (1999).
48. K. C. Garcia et al., *Science* **274**, 209 (1996).
49. J. Wang et al., *EMBO J.* **17**, 10 (1998).
50. Y. H. Ding et al., *Immunology* **8**, 403 (1998).
51. D. N. Garboczi et al., *Nature* **384**, 134 (1996).
52. M. R. Wormald, R. A. Dwek, *Structure* **7**, 155 (1999).
53. J. S. Haurum et al., *J. Exp. Med.* **190**, 145 (1999).

54. G. W. Hart et al., *Annu. Rev. Biochem.* **58**, 841 (1989).
55. D. Hudrisier et al., *J. Biol. Chem.* **274**, 36274 (1999).
56. R. M. Chicz et al., *J. Exp. Med.* **178**, 27 (1993).
57. J. S. Haurum et al., *J. Exp. Med.* **180**, 739 (1994).
58. U. M. Abdel Motal et al., *Eur. J. Immunol.* **26**, 544 (1996).
59. T. Jensen et al., *J. Immunol.* **158**, 3769 (1997).
60. B. Deck, M. Elofsson, J. Kihlberg, E. R. Unanue, *J. Immunol.* **155**, 1074 (1995).
61. A. Glithero et al., *Immunity* **10**, 63 (1999).
62. M. B. Deck, P. Sjolin, E. R. Unanue, J. Kihlberg, *J. Immunol.* **162**, 4740 (1999).
63. J. A. Speir et al., *Immunity* **10**, 51 (1999).
64. J. S. Haurum et al., *Eur. J. Immunol.* **25**, 3270 (1995).
65. A. Corthay et al., *Eur. J. Immunol.* **28**, 2580 (1998).
66. J. A. Speir et al., *Immunity* **14**, 81 (2001).
67. E. Michaelsson et al., *J. Exp. Med.* **180**, 745 (1994).
68. X. J. Zhao, N. K. Cheung, *J. Exp. Med.* **182**, 67 (1995).
69. S. Joyce et al., *Science* **279**, 1541 (1998).
70. D. B. Moody et al., *Nature* **404**, 884 (2000).
71. Z. Zeng et al., *Science* **277**, 339 (1997).
72. J. Deisenhofer, *Biochemistry* **20**, 2361 (1981).
73. P. M. Rudd et al., *Mol. Immunol.* **28**, 1369 (1991).
74. T. S. Mattu et al., *J. Biol. Chem.* **273**, 2260 (1998).
75. J. A. K. Mestecky, M. Kilian, *Methods Enzymol.* **116**, 37 (1985).
76. J. Baenziger, S. Kornfeld, *J. Biol. Chem.* **249**, 7270 (1974).
77. M. C. Field et al., *Biochem. J.* **299**, 261 (1994).
78. R. Shogren, T. A. Gerken, N. Jentoft, *Biochemistry* **28**, 5525 (1989).
79. M. R. Wormald et al., *Biochemistry* **36**, 1370 (1997).
80. P. Gettins, J. Boyd, C. P. Claudemans, M. Potter, R. A. Dwek, *Biochemistry* **20**, 7463 (1981).
81. W. I. Weis, M. E. Taylor, K. Drickamer, *Immunol. Rev.* **163**, 19 (1998).
82. P. Garred et al., *J. Rheumatol.* **27**, 26 (2000).
83. X. Dong, W. J. Storkus, R. D. Salter, *J. Immunol.* **163**, 5427 (1999).
84. P. M. Rudd et al., *J. Biol. Chem.* **272**, 7229 (1997).
85. G. F. Gao et al., *Nature* **387**, 630 (1997).
86. P. M. Rudd et al., *J. Mol. Biol.* **293**, 351 (1999).
87. Y. Ghendler, A. Smolyar, H.-C. Chang, E. L. Reinhertz, *J. Exp. Med.* **187**, 1529 (1998).
88. E. Y. Jones, S. J. Davis, A. F. Williams, K. Harlos, D. I. Stuart, *Nature* **360**, 232 (1992).
89. D. L. Bodian, E. Y. Jones, K. Harlos, D. I. Stuart, S. J. Davis, *Structure* **2**, 755 (1994).
90. P. M. Rudd et al., *Glycobiology* **9**, 443 (1999).
91. A. Merry, R. A. Dwek, P. M. Rudd, unpublished data.
92. W. I. Weis, K. Drickamer, W. A. Hendrickson, *Nature* **360**, 127 (1992).
93. C. M. Fletcher, R. A. Harrison, P. J. Lachmann, D. Neuhaus, *Structure* **2**, 185 (1994).
94. We thank M. Wormald for the molecular modeling, D. Shore for preparing Fig. 2, A. Karadimitris for his contribution to the section on CD1, S. J. Davis for his contribution to the section on CD2 and CD48, and D. Wing for careful reading of the manuscript. I.A.W. is supported by NIH grants CA58896 and AI42266. R.A.D. and P.M.R. thank Hilary C. Lister for her constant inspiration.

VIEWPOINT

Glycosylation of Nucleocytoplasmic Proteins: Signal Transduction and O-GlcNAc

Lance Wells,* Keith Vosseller,* Gerald W. Hart†

The dynamic glycosylation of serine or threonine residues on nuclear and cytosolic proteins by O-linked β -N-acetylglucosamine (O-GlcNAc) is abundant in all multicellular eukaryotes. On several proteins, O-GlcNAc and O-phosphate alternatively occupy the same or adjacent sites, leading to the hypothesis that one function of this saccharide is to transiently block phosphorylation. The diversity of proteins modified by O-GlcNAc implies its importance in many basic cellular and disease processes. Here we systematically examine the current data implicating O-GlcNAc as a regulatory modification important to signal transduction cascades.

Cells respond to their environment through dynamic posttranslational modification of their existing proteins. There are more than 20 posttranslational modifications that occur on eukaryotic proteins (1). Several of these modifications, with phosphorylation being the hallmark, participate in signal transduction. Generally, glycosylation is not thought to participate directly in signaling. Complex N- and O-linked glycosylation occurs on membrane-bound or secreted proteins that are synthesized in the endoplasmic reticulum and Golgi apparatus. The luminal or extracellular localization of these glycans restricts their potential for dynamic responsiveness to signals. In contrast, O-GlcNAc is a simple monosaccharide modification that is abundant on serine or threonine residues of nucleocytoplasmic proteins (2, 3). An O-GlcNAc site consensus motif has not yet been identified. However, many attachment sites are identical to those used by serine/thre-

onine) kinases, and a neural network program has been developed to predict O-GlcNAc sites (4). Unlike phosphorylation, O-GlcNAc modification of tyrosine residues has yet to be observed. Many proteins have been identified that carry this modification, including transcription factors, cytoskeletal proteins, nuclear pore proteins, oncogene products, and tumor suppressors (5–7). O-GlcNAc appears to modify a large number of nucleocytoplasmic proteins (Fig. 1). The attributes of O-GlcNAc, which are distinct from those of complex carbohydrates, predict that it plays an important role in signaling.

O-GlcNAc Meets the Requirements for a Signal Transduction Modification

In order for a protein modification to play an active role in signal transduction, it needs to have certain key features. First, the modification needs to be dynamic. For the proteins that have been examined to date, the O-GlcNAc half-life is much shorter than that of the modified polypeptide chain (8). Second, the removal or attachment of the modification should be inducible by certain stimuli. O-GlcNAc modification of certain proteins is known to change in response to T cell acti-

vation, insulin signaling, glucose metabolism, and cell cycle progression (6). Thus, O-GlcNAc displays features essential for a role in signal transduction.

Consistent with O-GlcNAc being dynamic and inducible, regulated nucleocytoplasmic enzymes for the attachment [O-GlcNAc transferase (OGT)] and for the removal (O-GlcNAcase) of the modification have been purified, characterized, and cloned (9–12). The OGT enzyme is modified by both O-GlcNAc and tyrosine phosphorylation and has 11 protein-protein interaction domains known as tetratricopeptide (TPR) repeats (13). OGT specifically interacts with a variety of other proteins. Because the transferase has a variety of binding partners and is itself Viewpointposttranslationally modified, OGT's substrate specificity, localization, and/or activity are likely to be regulated by signal transduction cascades. Furthermore, the activity of this enzyme is exquisitely responsive to intracellular uridine 5'-diphosphate (UDP)-GlcNAc and UDP concentrations, which are in turn highly sensitive to glucose concentrations and are known to fluctuate in response to a number of stimuli (14). Disruption of OGT activity is lethal in mouse embryonic stem cells, underscoring the importance of this modification (15). O-GlcNAcase, which specifically catalyzes the removal of O-GlcNAc from proteins, is a cytosolic neutral β -N-acetylglucosaminidase, unlike the general acidic lysosomal hexosaminidases. Both OGT and O-GlcNAcase are highly conserved from *Caenorhabditis elegans* to humans (9, 10, 12). Regulation of O-GlcNAc levels by these two enzymes may be analogous to the regulation of phosphorylation by kinases and phosphatases.

Department of Biological Chemistry, Johns Hopkins School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205 USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: gwhart@jhmi.edu