Critical Reviews in Biochemistry and Molecular Biology, 32(1):1–100 (1997)

Glycosylation: Heterogeneity and the 3D Structure of Proteins

Pauline M. Rudd and Raymond A. Dwek

Glycobiology Institute, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, U.K.

Referee: Dr. Tim Block, Dept. of Microbiology and Immunology, Jefferson Cancer Institute, 1020 Locust St., Thomas Jefferson University, Philadelphia, PA 19107

TABLE OF CONTENTS

1040-9238/97/$.50

ABSTRACT: Glycoproteins generally exist as populations of glycosylated variants (glycoforms) of a single polypeptide. Although the same glycosylation machinery is available to all proteins that enter the secretory pathway in a given cell, most glycoproteins emerge with characteristic glycosylation patterns and heterogeneous populations of glycans at each glycosylation site. The factors that control the composition of the glycoform populations and the role that heterogeneity plays in the function of glycoproteins are important questions for glycobiology.

A full understanding of the implications of glycosylation for the structure and function of a protein can only be reached when a glycoprotein is viewed as a single entity. Individual glycoproteins, by virtue of their unique structures, can selectively control their own glycosylation by modulating interactions with the glycosylating enzymes in the cell. Ex-

amples include protein-specific glycosylation within the immunoglobulins and immunoglobulin superfamily and site-specific processing in ribonuclease, Thy-1, IgG, tissue plasminogen activator, and influenza A hemagglutinin.

General roles for the range of sugars on glycoproteins such as the leukocyte antigens include orientating the molecules on the cell surface. A major role for specific sugars is in recognition by lectins, including chaperones involved in protein folding. In addition, the recognition of identical motifs in different glycans allows a heterogeneous population of glycoforms to participate in specific biological interactions.

KEY WORDS: chaperones, glycoforms, glycosylphosphatidylinositol anchors, immunoglobulins, immunoglobulin superfamily, N-glycans, O-glycans, oligosaccharides, oligosaccharide sequencing, variable site occupancy, virus.

I. INTRODUCTION

A. Glycoproteins Consist of Heterogeneous Populations of Glycoforms

The synthesis of the polypeptide chain of a glycoprotein is under genetic control. In contrast, the oligosaccharides are attached to the protein and processed by a series of enzyme reactions without the rigid direction of nucleic acid templates. Consequently, a single polypeptide that is glycosylated normally emerges from the biosynthetic pathway as a mixture of glycosylation variants, known as glycoforms.

For quiescent cells, each glycoprotein has a reproducible and characteristic glycosylation profile, suggesting that the proteins themselves direct the processing of their own glycan chains within the constraints imposed by a given array of enzymes and sugar nucleotides. Interestingly, within the endoplasmic reticulum (ER) all glycoproteins contain the same limited range of oligomannose sugars, and it is later, within the Golgi, that the extensive heterogeneity develops.

This observation poses a number of questions that are explored in this review.

1. What structural features of the individual protein are involved in initiating and developing N-glycosylation in the ER?
2. Are there processes that require the homogeneous glycosylation associated with glycoproteins within the ER?
3. Is the extensive glycan heterogeneity that develops on most glycoproteins in the Golgi controlled mainly by the protein or the cell?

Abbreviations: CDGS: carbohydrate-deficient glycoprotein syndrome; **CE:** capillary electrophoresis; **CHO:** Chinese hamster ovary; **CTL:** cytotoxic T-lymphocytes; **DMJ:** deoxymannojirimycin; **EBV:** Epstein Barr virus; **ER:** endoplasmic reticulum; **GalNAc:** *N*-acetyl galactosamine; **GPI:** glycosylphosphatidylinositol; **HA:** hemagglutinin; **HBV:** hepatitis B virus; **HuE:** human erythrocyte; **HIV:** human immunodeficiency virus; **HLA:** human leukocyte antigen; **Ig:** immunoglobulin; **IgSF:** immunoglobulin superfamily; **K:** kringle; ***n*-BuDNJ:** *n*-butyldeoxynojirimycin; **NMR:** nuclear magnetic resonance; **NP:** nucleoprotein; **MBP:** mannose binding protein; **MHC:** major histocompatability complex; **PI:** phosphatidylinositol; **PNGase:** peptide *N*-glycosidase; **RA:** rheumatoid arthritis; **RNase:** ribonuclease; **SAP:** serum amyloid protein; **TAP:** transporter associated with antigen processing; **TCR:** T cell receptor; **t-PA:** tissue plasminogen activator. **Nomenclature for describing oligosaccharide structures:** A(1,2,3,4) indicates the number of antennae linked to the trimannosyl core; G(0–4) indicates the number of terminal galactose residues in the structure; F: core fucose; B: bisecting N-acetyl glucosamine (GlcNAc); S: sialic acid; G, Gal: galactose; M, Man: mannose.

4. Are there functions for glycosylation that can be performed by any glycan, so that specific oligosaccharide structures are not required?
5. Are there roles for glycoproteins that require a heterogeneous rather than a homogeneous glycoform population?
6. Is heterogeneity at a single site compatible with a recognition function for glycans?
7. Are there roles for glycosylation motifs conserved within heterogeneous arrays of glycoforms, either within the cell or on the cell surface?
8. How can the current technology be developed to give rapid, detailed, and specific insights into oligosaccharide structure?
9. How can structural data from both oligosaccharide and protein analyses be combined (with anchor analysis where appropriate) to gain an overview of a glycoprotein as a single entity?

These are some of the questions that are addressed in the following sections of this review.

1. Section II focuses on the ER, addressing the role of the protein in the initiation of N-glycosylation and the addition of the GPI anchor precursor. It also explores how, in the ER, a restricted number of oligosaccharides common to all glycoproteins containing N-linked glycans can participate in protein folding and assembly.
2. Section III deals with glycosylation events in the Golgi. It explores the roles of the cell and the 3D structure of the protein in the development of heterogeneity and discusses the initiation of O-glycosylation (which may also occur in the ER) and the further processing of O-glycans and GPI anchors.
3. Section IV compares the structure and function of oligosaccharides in the immunoglobulins and immunoglobulin superfamily. Following on from this is a discussion of the extent to which the protein can control its own glycosylation with reference to the site-specific glycosylation of a range of glycoproteins. This leads to a discussion of the role that homogeneous and heterogeneous glycosylation play in specific recognition functions and in more general features such as cell surface packing.
4. Section V addresses one of the main conclusions of this review, which is that it is important to envisage a glycoprotein and its array of sugars as a whole rather than two separate entities. It suggests how newly developing technology may be used to analyze glycan populations rapidly, and how oligosaccharide structural data bases may be used to model oligosaccharides and glycoproteins.

This review has been organized so that each section can be read independently. In addition, we have restricted the discussion to a small number of molecules. Although this approach inevitably leads to some repetition, it serves to emphasize first that individual proteins contain a number of structural features that play a role in determining their glycosylation and second that their oligosaccharides do not necessarily have a unique functional role.

II. GLYCOSYLATION IN THE ENDOPLASMIC RETICULUM

A. Initiation and Development of N-Glycosylation in the Endoplasmic Reticulum (ER)

There are three main classes of covalent glycosidic linkages to protein, and these give rise to N-linked glycans, glycosylphosphatidlyinositol (GPI) anchors and O-linked

glycans (Figure 1). The cellular mechanisms for each of these modifications are available to all proteins that carry appropriate signals in the polypeptide chain. First, the polypeptide must contain a signal for entry into the secretory pathway, so that, as it leaves the ribosome, the NH_2 terminus translocates into the lumen of the endoplasmic reticulum (ER), where it first encounters the N-glycosylation machinery.

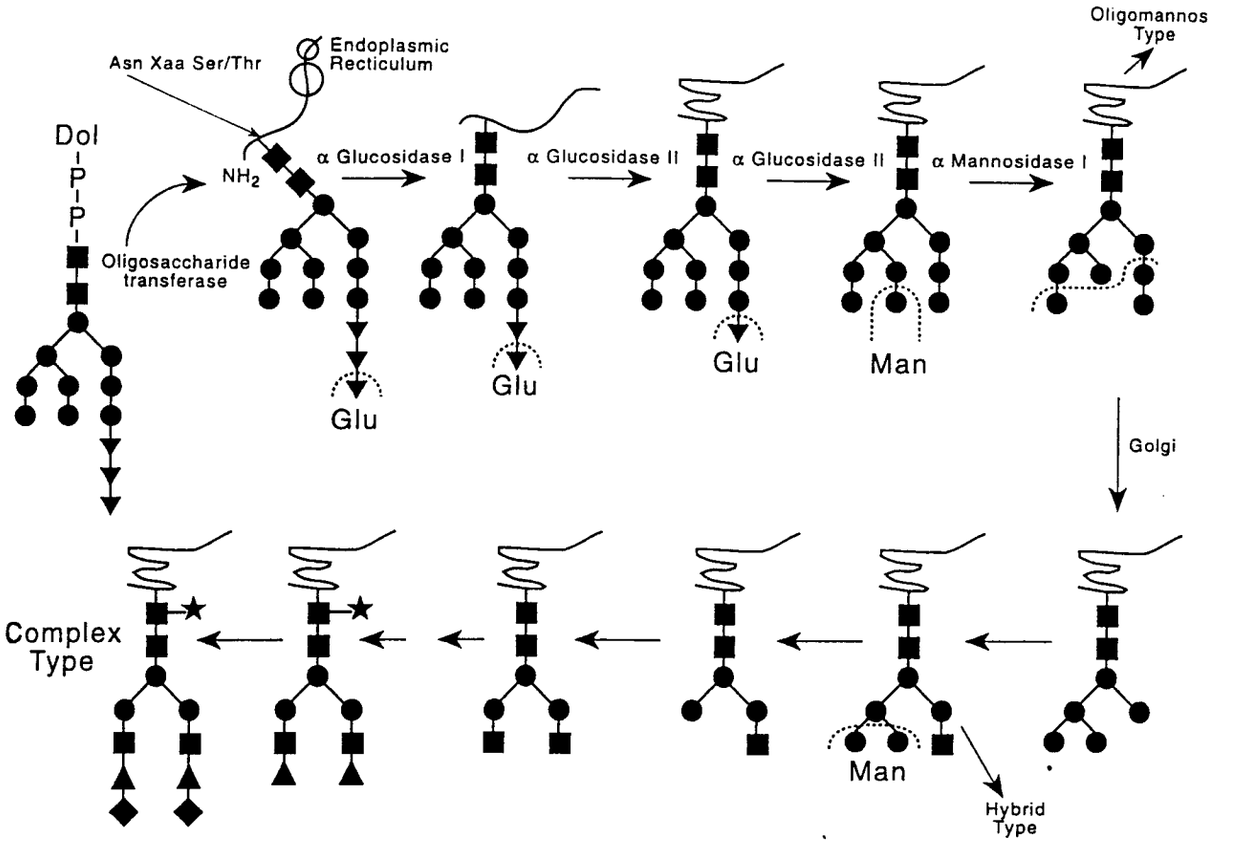
N-glycosylation is a cotranslational modification available to, but not necessarily used by, all proteins that contain the triplet (sequon) AsnXaaSer/Thr (where Xaa is any amino acid except Pro). A block of sugars ($Glc_3Man_9GlcNAc_2$) is transferred to nitrogen in the side chain of some Asn residues (Figure 1A) by the interaction of the glycosylation sequon with an oligosaccharyl transferase and the dolichol diphosphate oligosaccharide precursor, both of which are located in the membrane of the ER.

Similarly, the addition of a GPI anchor to the polypeptide chain, which also occurs in the ER, involves the transfer of a preassembled precursor to a specific site on the protein via ethanolamine phosphate (Figure 1B). However, while N-glycosylation is generally a cotranslational event that takes place as the protein is folding, a GPI anchor is attached posttranslationally to the fully folded protein. GPI anchors are attached selectively to those proteins that contain a GPI signal sequence. This sequence is cleaved and replaced by a preassembled GPI anchor precursor, which may subsequently be modified (Ferguson, 1991). In contrast to N-glycosylation and the addition of GPI anchors, O-glycosylation does not begin with the addition of a precursor, but of a single monosaccharide, usually *N*-acetylgalactosamine (GalNAc) (Figure 1C). This is transferred to the oxygen on appropriate Ser or Thr side chains in the fully folded and assembled protein within a ternary complex formed between the protein, UDP-GalNAc, and GalNAc transferase. The addition of further monosaccharide residues that elongate the sugar is commonly believed to take place in the Golgi complex. However, the location of the subcellular compartment where O-linked glycosylation is initiated is still controversial. It may range from subregions of the ER, a proximal Golgi compartment, an ER-Golgi compartment intermediate (ERGIC), and beyond the ERGIC in the Golgi apparatus (Perez-Vilar et al., 1991). The location of the glycosyl transferases may also be cell-type specific (Roth et al., 1986) and vary depending on the stage of cell differentiation (Perez-Vilar et al., 1991).

1. The Primary Structure of the Protein and Sequon Occupancy

Although all of the proteins that are translated in a particular cell and enter the secretory pathway are exposed to the same glycosylation machinery, potential glycosylation sites are not invariably occupied. It has been estimated that between 10 and 30% of potential N-glycosylation sites are not occupied (Gavel and von Heijne, 1990), Mononen and Karjalainen, 1987). Initiation of glycosylation in the ER is mediated by a membrane-bound oligosaccharyl transferase, the active site of which is located some 30 to 40 Å from the membrane (Nilsson et al., 1993). This enzyme transfers the $Glc_3Man_9GlcNAc_2$ oligosaccharide from the dolichol diphosphate precursor, which is also located in the membrane, to side chains of some asparagine residues in the polypeptide chain.

As the growing peptide chains translocate through the membrane of the ER, asparagine residues that form part of the glycosylation sequon (AsnXaaSer) may be



A

FIGURE 1. **(A)** Schematic diagrams of some of the steps in the biosynthesis of N-linked glycans. In the ER, the dolichol phosphate precursor oligosaccharide is added cotranslationally to the nascent peptide chain, and the three glucose residues are removed sequentially by glucosidase I and II. The Man_9 oligomannose glycoform is trimmed by mannosidase I and transported to the Golgi, where the glycans may be processed to hybrid or complex-type sugars. In the figure, a biantennary sugar is shown at the end of the pathway, although multiantennary glycan and structures with lactosamine extensions may also be developed in the Golgi. The enzyme reactions do not always go to completion; therefore, a single glycoprotein normally leaves the pathway as a collection of glycosylated variants of the same amino acid sequence. ●: mannose, ■: *N*-acetylglucosamine; ▼: glucose; ★: fucose; ▲: galactose; ◆: sialic acid. **(B)** Schematic diagram of some of the steps in the attachment of glycosylphosphatidylinositol (GPI) anchors to proteins. The figure shows some of the steps involved in the addition of the preformed GPI anchor precursor to the side chain of one of six specific amino acid residues, marked X, in the peptide chain. These amino acids must be either Cys, Asn, Asp, Ser, Gly, or Thr. Subsequently, the lipid chains may be remodeled and additional sugar residues may be attached to the glycan core. **(C)** A schematic representation of the O-glycosylation pathway showing the structures of the common cores. O-glycans are attached by the stepwise addition of single monosaccharides, normally beginning with *N*-acetylgalactosamine, to the hydroxyl side chain of Ser or Thr. To date, eight common core structures have been identified (Hounsell et al., 1996), and these may be extended to form the backbone region by the addition of galactose (linked β-1-3/4) and GlcNAc (linked β-1-3/6). Terminal residues include galactose, *N*-acetylgalactosamine, fucose, sialic acid, and sulfate.

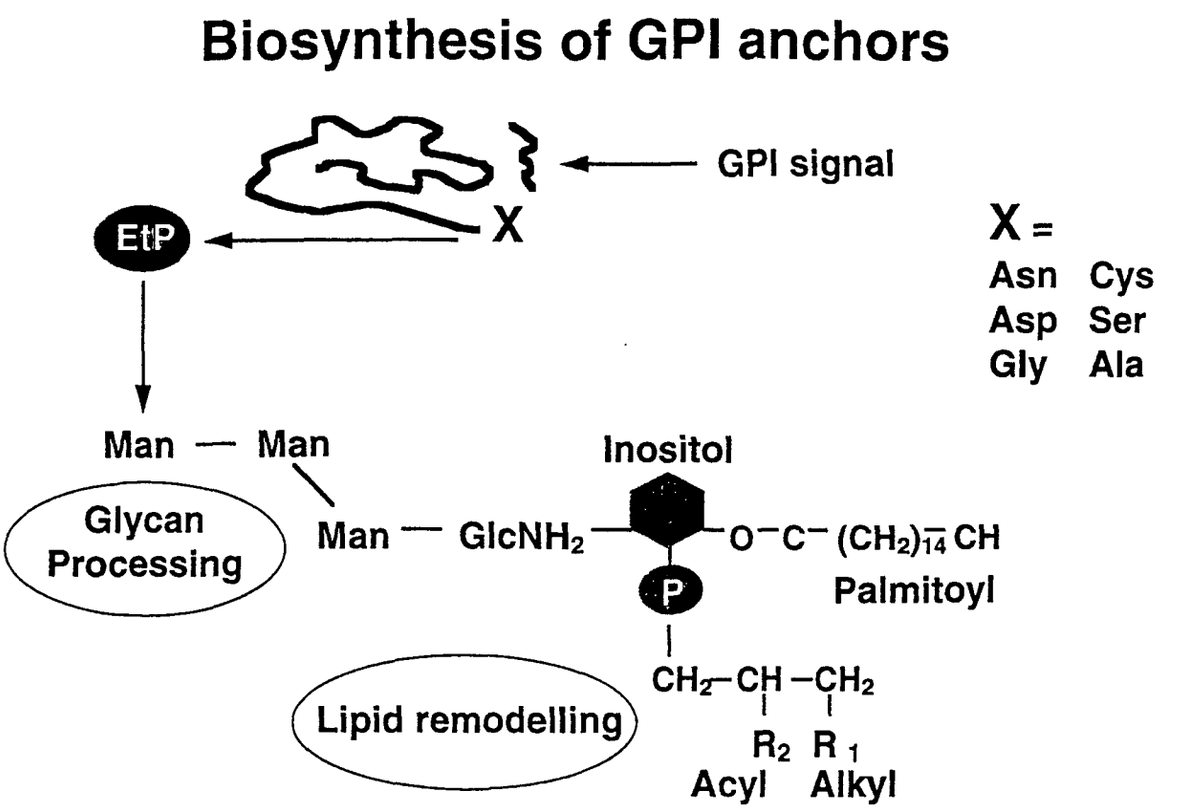


FIGURE 1B

temporarily in a position to form a ternary complex with the enzyme and the precursor. $Glc_3Man_9GlcNAc_2$ may then be transferred to the N-glycosylation site and the oligosaccharyl transferase and the dolichol phosphate released. Interestingly, oligosaccharides on the outer branches of the dolichol phosphate donor appear to induce conformational changes in the active site of the oligosaccharyl transferase, influencing the association constant of the peptide substrate that contains the glycosylation sequon (Breuer and Bause, 1995).

N-glycosylation is a complex process, and its efficiency depends on many factors. The structure of the central amino acid (X) in the glycosylation sequon is one important determinant. In a study of rabies virus glycoprotein, variants containing each of the 20 amino acids in the X-position in the sequon at Asn37LeuSer were generated by site-directed mutagenesis (Shakin-Eshleman et al., 1996). The efficiency of glycosylation ranged from zero to nearly 100%. Pro completely blocked core glycosylation, while sequons containing Trp (5%), Asp (19%), Glu (24%), and Leu (43%) were inefficiently glycosylated. Other amino acids gave values from 70% (Phe) to 97% (Ser). A number of characteristics have been identified that may explain these results.

First, large hydrophobic amino acids may produce an unfavorable local protein conformation for the oligosaccharide transfer. Second, the negative charge of some amino acids may increase the ability of oligosaccharyl transferase to interact with the sequon and the negatively charged dolichol-PP-oligosaccharide precursor. Third, it has been noted that the presence of

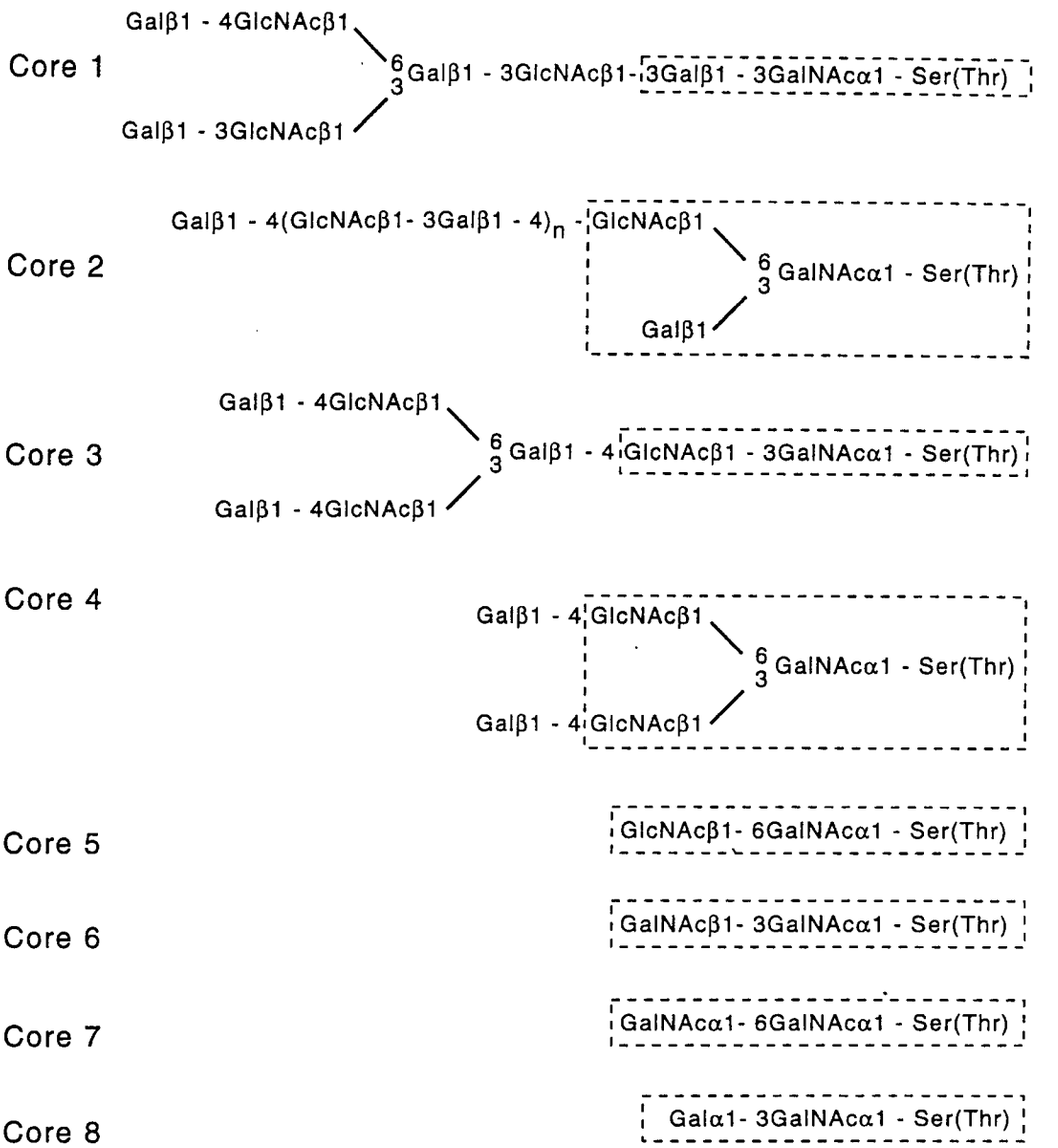


FIGURE 1C

hydroxyl groups, as opposed to amide side chains, in the region close to the sequon are associated with highly efficient glycosylation.

Other factors in the primary sequence of the protein that inhibit glycosylation include Pro immediately following the sequon (Bause, 1983). Replacing Ser with Thr in the sequon results in a general increase in site occupancy (Kasturi et al., 1995). Sequons close to either the amino or carboxy terminal are generally less efficiently glycosylated. For example, the cytokine interleukin-1β, which has a potential N-glycosylation site at Asn7CysThr, is 50% glycosylated when expressed in *Saccharomyces cerivisiae*. Increasing the distance between the Asn7 residue and the signal-processing site to at least nine amino acids results in almost 100% glycosylation of the sequon (Livi et al., 1991).

2. Primary Structure of the Protein and the Transfer of Glycosylphosphatidyl Inositol Membrane Anchors

The primary sequence also determines whether the protein receives a GPI anchor (Ferguson, 1991; McConville and Ferguson, 1993). First, the peptide sequence must contain an amino-terminal signal for entry into the lumen of the endoplasmic reticulum where the anchor precursor is preassembled. Then the carboxy terminus of the protein must contain a GPI signal sequence. This commonly contains a series of 12 to 20 hydrophobic residues at the carboxy terminus of the primary translation product preceded by a polar region of amino acids. The signal sequence is cleaved and replaced by a pre-assembled GPI precursor (Figure 1B) in what appears to be an ATP- and GTP-independent transamidation reaction (Mayor et al., 1991; Gerber et al., 1992). In the fully formed anchor (Figure 2) the carboxy-terminal cleavage amino acid (restricted to Cys, Asp, Asn, Gly, Arg, or Ser) is linked via ethanolamine phosphate to a glycan with a conserved backbone sequence (Manα1-2Manα1-6Manα1-4GlcNH$_2$). The backbone is linked to the 6-position of the *myo*-inositol ring of phosphatidyl inositol (PI). All GPI anchors contain two lipids, normally one acyl and one alkyl, attached to the glycan through the phosphate and glycerol on the inositol ring. In addition, a subset of anchors, including CD59 (Figure 2), are palmitoylated at C3/4 on the inositol ring. Interestingly, molecular modeling suggests that proteins with anchors containing three lipids rather than two are differently oriented with respect to the membrane, both during biosynthesis and eventually on the cell surface. During transit, the lipid chains may be remodeled (Ferguson, 1991), and the backbone of the core glycan may be further processed to give a range of glycoforms, as in the *Trypanosome brucei* variant surface glycoprotein (Ferguson et al., 1988).

3. The Secondary Structure of the Protein and Sequon Occupancy

The predicted secondary structures of a large number of glycoproteins indicates that about 70% of the AsnXaaSer/Thr sequences are located in β-bends, 20% in β-sheets, and 10% in α-helices. No restrictions were found in the location of glycosylated and non-glycosylated sequences within the 3D structure of the protein (Mononen and Karjalainen, 1987). However, whereas the sequon occupancy may not be affected by the final 3D structure of the protein, the local and possibly temporary structure of the nascent polypeptide may play a role in N-glycosylation (Allen et al., 1995; Bullied et al., 1992). The transfer of the precursor occurs at about 30 to 40 Å above the membrane (Nilsson et al., 1993), and this allows space and time for secondary structure to form within the nascent protein before the interaction with the oligosaccharyl transferase. As a result, in cases where the secondary structure inhibits the transfer of the dolichol phosphate precursor, there may be a competition between glycosylation and folding. Disulfide bond formation is one process that may make sites inaccessible to the dolichol phosphate precursor, and it has been noted that sequons containing Cys residues that are involved in disulfide bond formation are not normally glycosylated (Bause et al., 1982; Bullied et al., 1992).

The importance of disulfide bond formation in determining site occupancy is illustrated by two studies, the first of which involved bovine pancreatic RNase. The re-

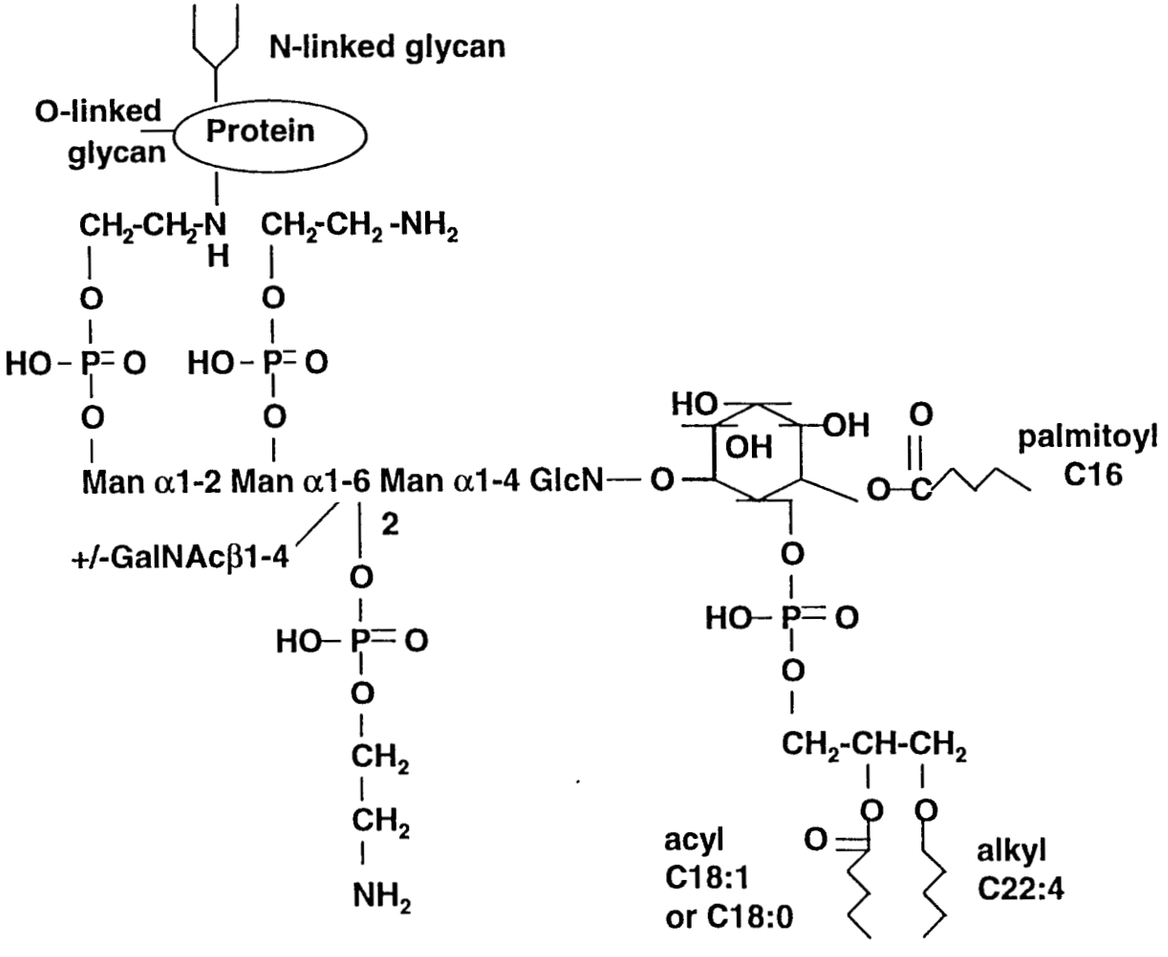


FIGURE 2. Schematic drawing of the CD59 glycan anchor. The structures of the lipids and glycans are based on data derived in Rudd et al. (1997). The inclusion of the ethanolamine groups in this model is solely by analogy with other mammalian GPI anchors such as rat brain Thy-1 antigen (Homans et al., 1988), human erythrocyte acetylcholinesterase (Deeg et al., 1992), and human leukocyte CD52 (Treumann et al., 1995).

gion of the RNase polypeptide close to the glycosylation site at Asn34 contains cysteine residues at sites 26 and 40. It may be that these, or other cysteine residues, form disulfide bonds during translation that produce a temporary conformation that partially inhibits the transfer of the dolichol phosphate precursor. Support for this comes from a study by Williams and Lennarz (1984) in which RNase, which is normally only 12% glycosylated (Carson et al., 1981), was 100% glycosylated when expressed in the presence of a reducing agent that prevented the formation of disulfide bonds.

A similar study involved tissue plasminogen activator (t-PA) (Figure 3A), an enzyme that contains three N-linked glycosylation sites, one of which, Asn184-GlySer in the kringle 2 (K2) domain, is normally 50% glycosylated. The sites in K1 (Asn117SerSer) and in the protease domain (P) (Asn448ArgThr) are both fully glycosylated. (Pohl et al., 1984; Parekh et al., 1989a; Spellman et al., 1989). A fourth potential glycosylation site, Asn228ProSer, is never occupied, presumably because it contains a Pro residue that inhibits glycosylation. In contrast, the amino acids Arg, Ser, and Gly that occupy the central position in the other three sequons are all expected to be efficiently glycosylated (Shakin-Eshleman et al., 1996). Moreover, these three sites are similarly located on loops in the protein structure and appear to be fully accessible, at least on the fully folded protein (Ponting, C. and Downing, K., personal communication). The transfer of the dolichol phosphate precursor to Asn184 takes place during or after folding of the finger (F) and epidermal growth factor (EGF) domains and kringle 1. One possible explanation of the inefficient glycosylation at site 184 (K2) is that the temporary 3D structures taken up by the folding nascent protein may inhibit the efficient transfer of the dolichol phosphate precursor to Asn184. In support of this proposal, a study by Allen et al. (1995) showed that Asn184 was fully glycosylated when expressed in the presence of a reducing agent. At least 20 cysteine residues that take part in disulfide bonding are translated before or soon after the glycosylation site at Asn184 (Figure 3B), including Cys180 and Cys191, which are on either side of the sequon. This suggests that the oligosaccharyl transferase and the dolichol phosphate precursor may be interacting with many different partially folded substrates, each with a characteristic binding constant. In some cases the concentration of the transferase/precursor complex may be too low to allow Asn184, presented in temporary conformations with low binding constants, to be glycosylated efficiently.

4. Cellular Enzyme Levels and Sequon Occupancy

The extent to which potential N-glycosylation sites are occupied also depends on cellular factors, such as the level of the oligosaccharyl transferase, the nucleotide donor, and the dolichol diphosphate precursor. The concept that dolichol diphosphate levels limit the percentage of occupied sites is supported by a study of bovine pancreatic ribonuclease (RNase), which occurs in two forms: RNase A, which is not glycosylated, and RNase B, which contains one N-linked oligomannose-type sugar. The addition of dolichol phosphate to tissue slices increased the percentage of RNase B from 12 to 90% (Carson et al., 1981). Interestingly, decreased levels of phosphomannomutase, the enzyme that is involved in the production of the mannose 1-phosphate required for the synthesis of the dolichol phosphate precursor, have been found in patients with carbohydrate deficient glycoprotein syndrome (CDGS) type I (Van Schaftingen and Jaeken,

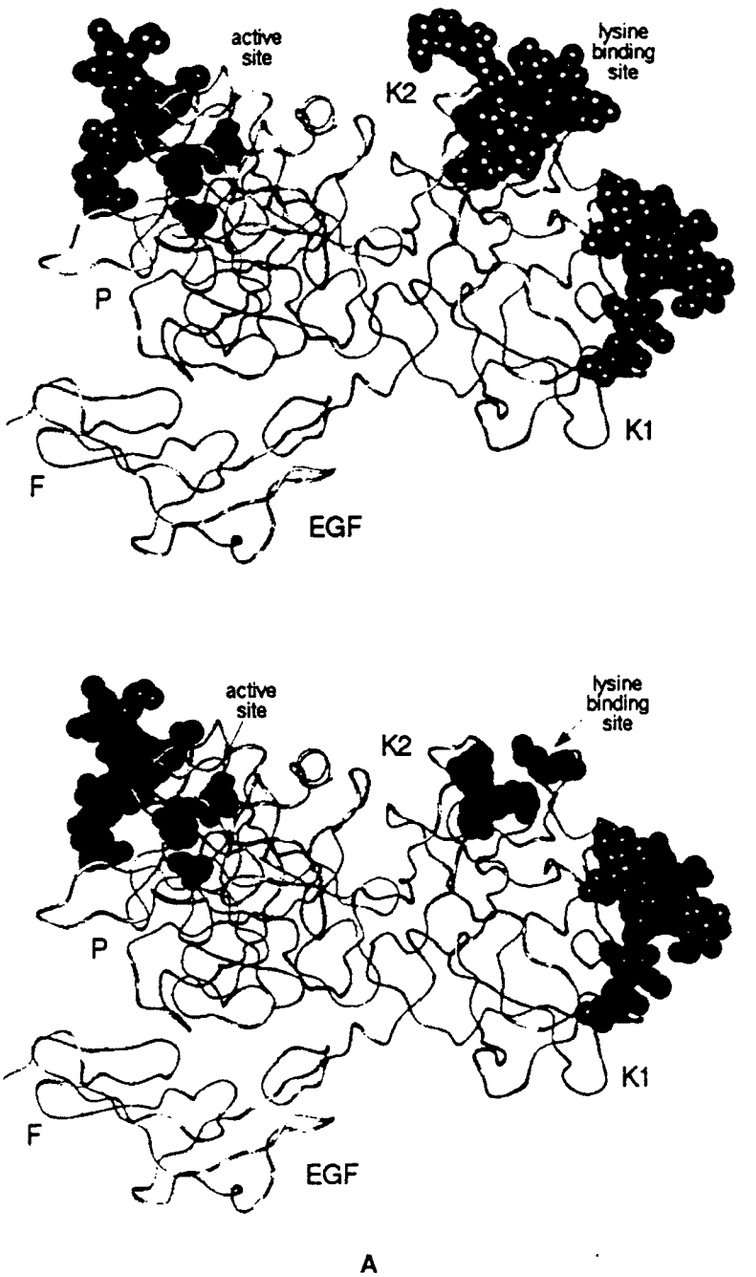


FIGURE 3. Tissue plasminogen activator (t-PA). **(A)** Schematic model of t-PA types 1 (top) and II (bottom). t-PA is composed of five domains: a fibronectin type 1 finger module (F), an epidermal growth factor (EGF)-like module, two kringles (K1 and K2), and a serine protease domain (P). This model was constructed using the coordinates of the finger growth factor pair (Smith et al., 1995) and kringle 2 from human t-PA. Kringle 1 and the serine protease domains were modeled by homology. In type I, Asn117 contains an oligomannose sugar, and complex sugars are shown at Asn184 and 448. The lysine-binding site is shown in kringle 2. In type II t-PA, the sugar at Asn184 is absent, suggesting that glycosylation at this site in type I t-PA may sterically hinder lysine binding. **(B)** Primary structure of human t-PA showing the three N-linked glycosylation sites occupied in type I (Asn117, Asn184, and Asn448) and site of O-fucosylation at Thr61. Disulfide bonds are shown by black lines, the catalytic triad by *, and the cleavage site for the conversion of single to two chain t-PA is marked with an arrow.

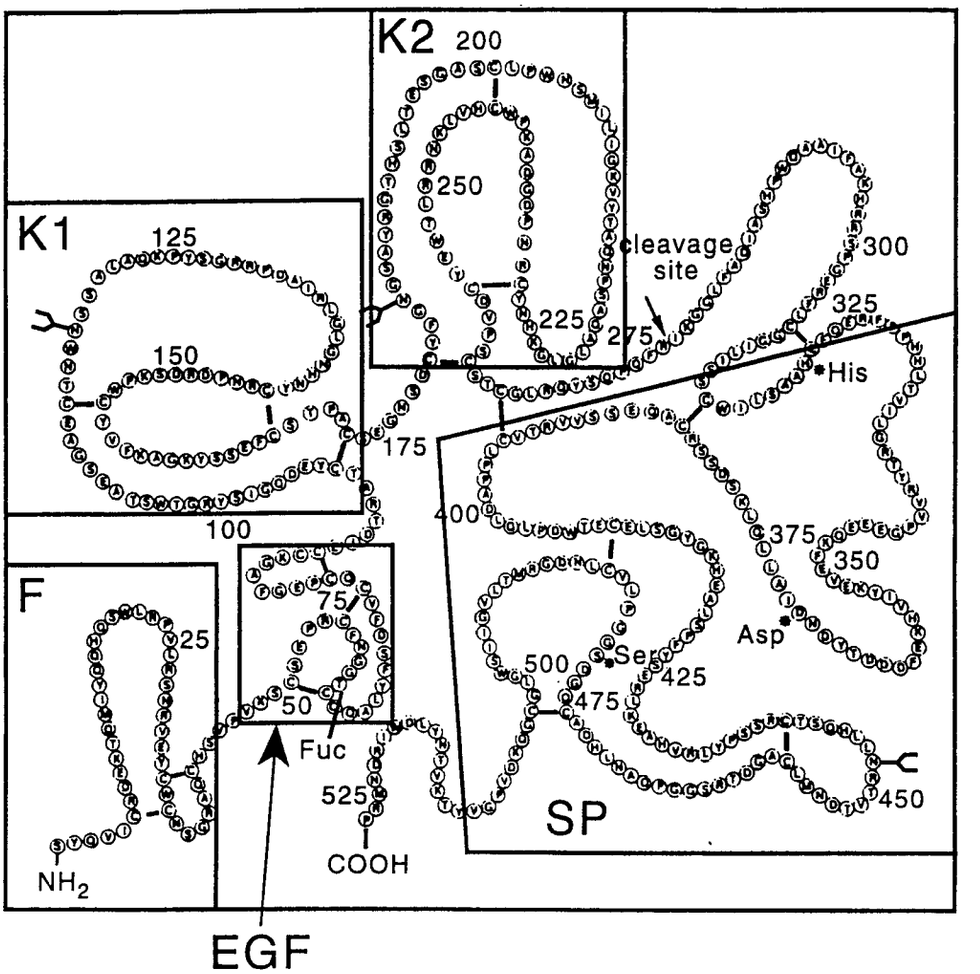


FIGURE 3B

1995). This finding may account for the fact that, in this disease, many N-glycosylation sites on a wide range of proteins are underoccupied (reviewed by Jaeken et al., 1993). For example, normal serum transferrin contains two fully occupied sites, 95% of which contain the disialylated, biantennary complex glycan A2G2SA2. This glycan carries four negative charges due to sialic acid. In contrast, transferrin from CDGS serum also contains molecules with one or zero sites occupied (Yamashita et al., 1993).

These molecules contain the same biantennary glycan (A2G2SA2), but underoccupancy of the glycosylation sites leads to molecules carrying either two or zero negative charges due to sialic acid. Figure 4 (Iourin et al., 1996) shows the capillary electrophoresis (CE) profiles of intact normal and CDGS transferrin. This indicates that, in addition to peak 4, CDGS transferrin contains two glycoforms (0 and 2) that are not present in normal transferrin. These represent the carbohydrate-deficient transferrin containing 0 and 2 sialic acid residues. Another example in which cellular enzyme levels control site occupancy comes from a comparison of human pancreatic and urine RNase. These proteins are derived from

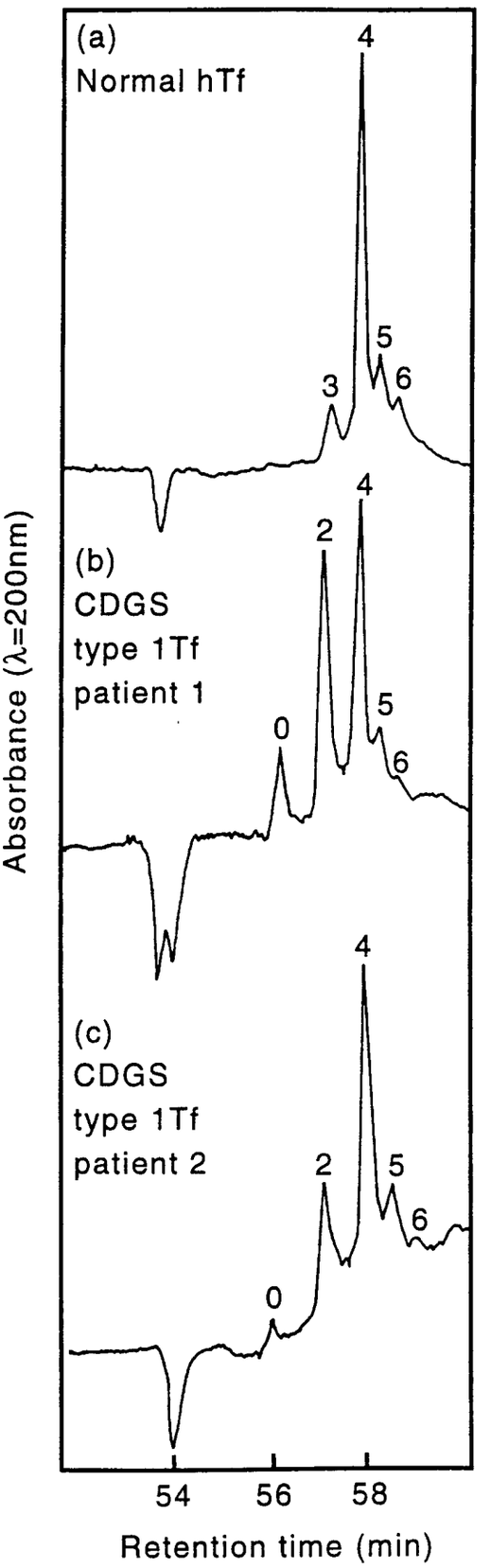


FIGURE 4. Capillary electrophoresis charge separation profiles of intact serum transferrin from (a) an individual normal control and (b,c) two patients with type I carbohydrate-deficient glycoprotein syndrome (CDGS). Normal human serum transferrin consists of glycoforms in which both N-linked glycosylation sites are occupied mainly by biantennary sialylated glycans that together carry four negatively charged sialic acid residues (a; peak 4). CDGS transferrin has two additional peaks, numbered 2 and 0. Peak 2 contains a population of glycoforms with only one occupied site, which contains a fully sialylated biantennary glycan and therefore carries two negative charges. Peak 0 contains transferrin in which both glycosylation sites are unoccupied and which therefore carries no charge that can be ascribed to sialic acid.

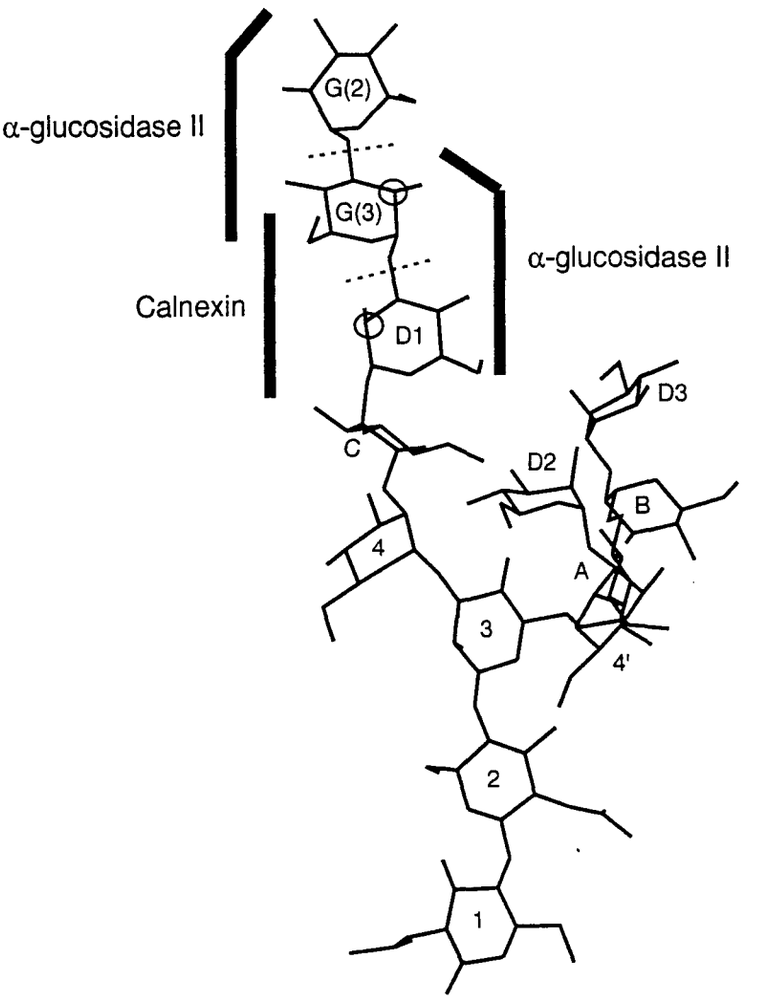
the same gene product and may be assumed to have the same 3D structure (Ribo et al., 1994, Breukelman et al., 1993, Beintema et al., 1984). The occupancy of the three glycosylation sites is different in the enzymes isolated from the two sources. In urine, each of the three sites (Asn34MetThr, Asn76SerSer, and Asn88GlySer) is fully occupied, whereas in human pancreatic RNase, Asn34 and Asn76 have carbohydrate attached to 100 and 50% of the molecules, respectively, and only a small population of molecules are glycosylated at Asn88 (Ribo et al., 1994). These data indicate that the transfer of the $Glc_3Man_9GlcNAc_2$ oligosaccharide precursor to RNase is less efficient in human pancreas than in the cells from which urine RNase is derived.

5. Glycan Processing in the ER

The initial steps that modify the $Glc_3Man_9GlcNAc_2$ precursor depend on the successive removal of the three glucose residues attached to the terminal mannose D1 on the α-1,3 arm of the oligosaccharide, by α-glucosidases I and II (Figures 5A and B) (Hubbard and Ivatt, 1991). This process is common to all N-linked glycoproteins, and is completed only when the protein is fully folded. Examination of the 3D solution structure of the $Glc_{3-1}Man_9GlcNAc_2$ precursors has suggested possible functions for these glucose residues (Petrescu et al., 1997). The rapid removal of the freely accessible terminal α-1,2-linked Glc G(1), by glucosidase I (within 30 s) facilitates the release of the oligosaccharyl tranferase following transfer of the dolichol phosphate precursor to the protein. The second Glcα-1,3-Glc linked residue, G(2), is removed by glucosidase II in a protein-independent manner within 30 minutes. This residue spaces G(1) from the protein and orients it. The final α-1,3-linked Glc, G(3), is removed more slowly by glucosidase II, allowing the G(3) residues attached to proteins that use the calnexin or calreticulin pathways to interact with these chaperones.

Either during or after folding and deglucosylation, from one to four of the α-1,2 linked mannose residues are removed by the ER α-mannosidase I. Interestingly, proteins expressed by Chinese hamster ovary (CHO) cells cultured in the presence of the ER glucosidase inhibitor nBuDNJ contained $Glc_3Man_7GlcNAc_2$ (Petrescu et al., 1997). This suggests that in the biosynthesis of N-linked sugars, some mannose residues may be cleaved before the glucose residues are removed, and the protein is fully folded. This is consistent with the finding that calnexin recognizes all glycoforms containing Glc(1), although the affinity is greater for $Glc_1Man_{7-9}GlcNAc_2$ structures (Ware et al., 1995).

In the early stages of processing, the protein may, in some cases, shield the oligosaccharide from the ER α-mannosidase 1. For example, in bovine pancreatic RNase B, 6% of the structures are not processed beyond Man_9, 17% remain as Man8 glycoforms, 11% terminate at Man7, 20% of the glycans reach the Man6 stage, and only 48% are processed fully by the mannosidase enzymes to Man5 (Rudd et al., 1992). The possibility that the protein may protect the sugars from the glycosylating enzymes is consistent with the finding that the rate of digestion of the α-1,2 linked oligomannose sugars attached to RNase B with *A. Saitoi* α-1,2 mannosidase was 150 times slower than the rate of digestion of the free sugars. In both the free sugars and the intact glycoprotein the Man6 sugar was more resistant to the enzyme than the Man7, 8, or 9 oligosaccharides, suggesting that these data reflect differ-



A

FIGURE 5. **(A)** *Schematic structure of* Glc_2Man_9 *showing the proposed epitope for glucosidase* II on Glc_2Man_9 and calnexin and glucosidase II on Glc_1Man_9. The circles indicate the points of epimerization between glucose and mannose. **(B)** Schematic representation of some of the steps in the N-glycosylation pathway. **(1)** A representative protein subunit has been drawn with three $Glc_3Man_9GlcNAc_2$ sugars attached to N-glycosylation sites (AsnXaaSer/Thr) (A, B, and C) located in different regions of the polypeptide. **(2)** As the subunit folds, the two terminal glucose residues in both sugars are cleaved by α-glucosidase I and II. One region of the protein (containing site A) is shown bound to BiP, excluding the possibility that this region will use the calnexin pathway. **(3)** The second region of the protein (containing site B) is shown folding via a calnexin- dependent pathway, and contains the chaperone bound to the terminal glucose residue (G1) of the $Glc_1Man_9GlcNAc_2$ glycoform. The third region (containing site C) is shown with the terminal glucose residue of the $Glc_1Man_9GlcNAc_2$ glycoform bound to calreticulin. When folding is complete, α-glucosidase II removes the final glucose residue from $Glc_1Man_9GlcNAc_2$. **(4)** The fully folded subunit is shown with $Man_5GlcNAc_2$, $Man_8GlcNAc_2$, and $Man_6GlcNAc_2$ at sites A, B, and C, respectively. These glycans are the products of a series of interactions of the oligomannose sugars with α-mannosidase I, an enzyme that removes from one to four of the α1-2 linked mannose residues from the oligosaccharide precursor. **(5)** The fully folded protein (assembled into subunits) may then be transported to the Golgi by ERGIC 53, where the processing of oligomannose to hybrid or complex sugars takes place on the fully assembled molecule.

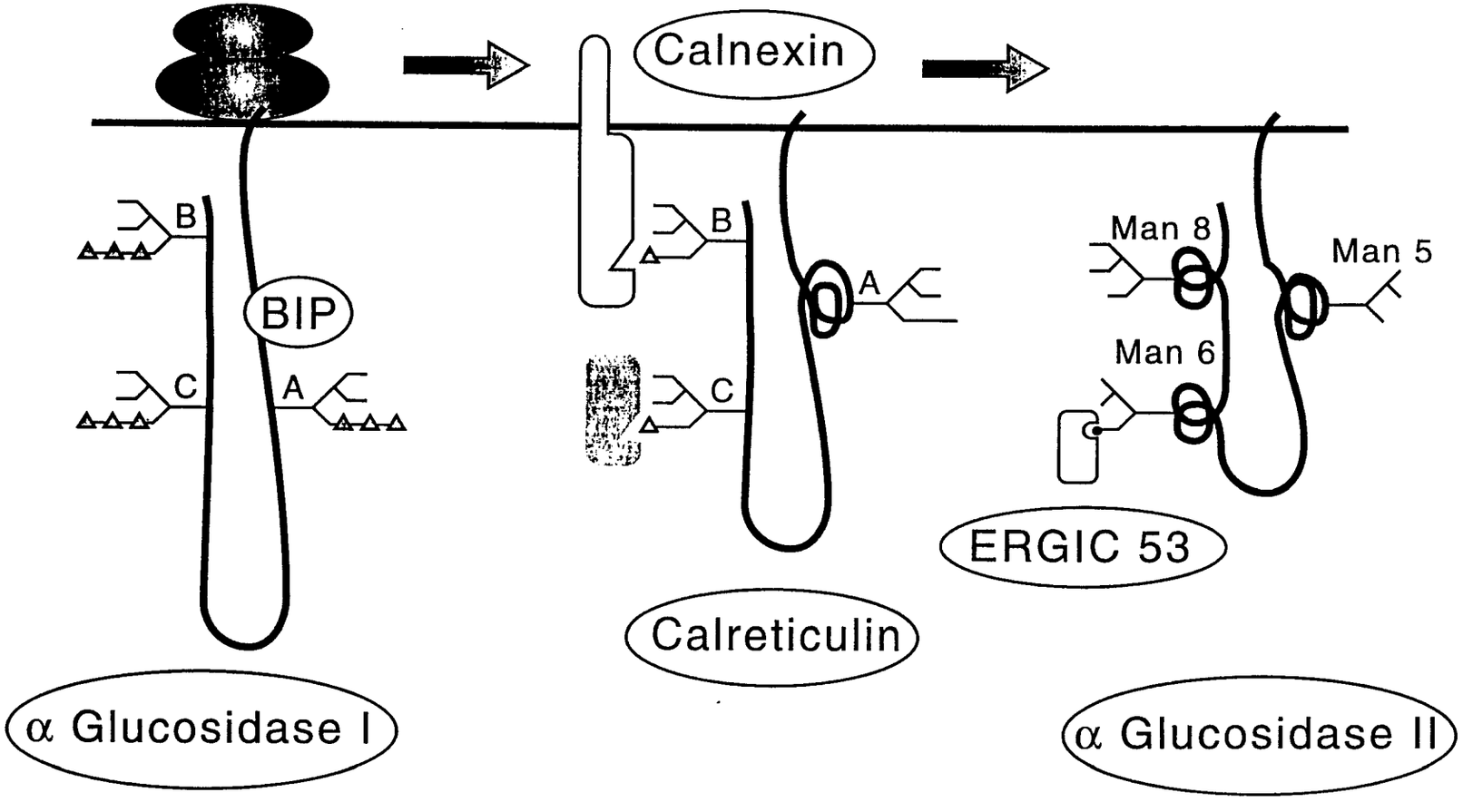


FIGURE 5B

ences in the accessibility of the sugars over and above the difference in diffusion rates of the two species (Rudd et al., 1992).

It is interesting to note that most proteins that consist of assemblies of several different protein chains, or multimers of the same chain, are assembled during their passage through the ER and the ERGIC. These include the immunoglobulins, influenza A hemagglutinin, and viral coat proteins. For such proteins, some glycan processing in the ER and all processing in the Golgi takes place within a multimeric structure.

B. General and Specific Roles for Glycans in the ER

The addition of N-linked sugars to proteins invariably requires the transfer of a common precursor, $GlcNAc_2Man_9Glc_3$, to nitrogen in the side chain of an asparagine residue. The fact that this glucosylated oligomannose structure is immediately deconstructed in the ER to yield oligomannose structures, and subsequently reprocessed in the ER and Golgi to form complex or hybrid-type structures is an interesting question for cell biology. Almost all the glycosylation-processing events that take place in the ER are general mechanisms that involve trimming, and many are independent of the structure of the individual protein. This suggests that, during protein folding within the ER, there are common roles for sugars that can be accomplished by a relatively small range of structures.

All sugars can impart a similar range of physical properties to the proteins to which they are attached. In the ER, these may include increasing solubility, shielding the protein surface with a network of glycans and organized water that may prevent aggregation, and orienting anchored glycoproteins in the membrane of the ER. In addition, any sugar may be able to modulate sterically the interactions of the protein (e.g., with chaperones) if it is located close enough to the binding site (Rudd et al., 1994a; Mori et al., 1995). There may also be roles for specific oligosaccharide motifs that can impart particular functions to the proteins that contain them. For example, all proteins using calnexin and calreticulin-folding pathways require the Glcα1-3Man motif, and all glycoproteins that are targeted to the lysozome contain a mannose-6-phosphate residue (Kornfeld, 1992).

1. Glycosylation and Protein Folding in the ER

The transfer of the precursor to the Asn side chain takes place during translocation of the polypeptide into the ER before protein folding is complete. Studies of the conformational dynamics of flexible oligopeptides indicate that glycosylation can alter the conformational profile of a polypeptide and allow it to take up conformations not accessible to the unglycosylated peptide (Imperiali and Rickert, 1995). This implies that, in some cases, glycosylation may direct a protein through particular folding pathways, producing structural nucleation elements important for the efficiency of the folding process. In the initial stages of protein synthesis, all sugars may, by reason of their size and hydrophilicity, prevent nonspecific binding and aggregation of newly formed proteins. Their hydrophilic properties may also prevent hydrophobic sections of the protein, destined to be on the surface from folding inward. When translation is completed, the hydrophilic sugars may increase the solubility of proteins within the highly concentrated lumen of the ER. Particular oligosaccharides are not required for

any of these functions, as they are properties conferred on proteins by all sugars.

2. Glycosylation in Calnexin/ Calreticulin-Dependent Protein Folding

Calnexin (Ou et al., 1993) and calreticulin (Wada et al., 1991) are molecular chaperones which bind transiently to the monoglucosylated forms of some glycoproteins. Both are located in the ER, but while calnexin is a membrane-bound glycoprotein, calreticulin is soluble and retained in the ER through the interaction of a KDEL sequence with the KDEL receptor (Kelly, 1990). Interestingly, calreticulin and endomannosidase have been copurified from a Golgi fraction, suggesting a possible alternative scheme for deglucosylation through the action of an endomannosidase rather than glucosidase II (Spiro et al., 1996). The mechanism of action of calnexin and calreticulin is under debate. Calnexin, which has been the subject of more research than calreticulin, is involved in a complex series of interactions that involve α-glucosidase II, glucosyl transferase, UDP-glucose, and the unfolded glycoprotein (Bergeron et al., 1994; Otteken et al., 1996). The unfolded or misfolded glycoprotein contains the monoglucosylated sugar $G(3)Man_9GlcNAc_2$. At intervals during the folding process, G(3) is removed by α-glucosidase II; if the refolding is incomplete, glucosyl transferase reglucosylates the oligomannose sugar at the D1 position on the α-1,3 arm using UDP-glucose as the donor nucleotide. The G(3) residue again binds calnexin, which retains the molecule in the ER, while the process is repeated until the correct folding pattern is achieved. The terminal glucose residue G(3) is finally removed when the protein is correctly folded (Figure 5B).

Following the transfer of $Glc_3Man_9GlcNAc_2$ to the protein, glucosidase I removes G(1). Glucosidase II then rapidly removes G(2). The resulting $Glc_1Man_9GlcNAc_2$ structure is a substrate for both glucosidase II and calnexin. Nuclear magnetic resonance (NMR) structural analysis of the G(2)-G(3) and G(3)-D1 linkages shows that the two disaccharide motifs are superimposable except for the epimerization at the C2 carbon atoms (circled in Figure 5A) (Petrescu et al., 1997). Figure 5 shows how this feature can provide a common epitope for glucosidase II, which cleaves both these linkages. In contrast, calnexin can distinguish between the G(2)-G(3) and G(3)-D1 linkage and binds only the latter. This suggests that the C2 position of residue D1 is involved in the binding of calnexin to the sugar.

The full details of the folding mechanism of proteins that use the calnexin-dependent pathway are not yet fully understood, but it is clear that initially calnexin acts as a lectin that recognizes the $Glc\alpha\text{-}1\text{-}_3Man_9GlcNAc_2$ structure (Ware et al., 1995). Further, a study by Zapun et al. (in press) demonstrated that calnexin binds native, partially folded, and unfolded RNase B, indicating that, in some cases, calnexin may act exclusively as a lectin and does not require a protein-binding site. An alternative mechanism is suggested by a study that showed that the assembly of class I human leukocyte antigen (HLA) heavy chains with β2-microglobulin was enhanced fivefold by calnexin (Vassilakos et al., 1996). The data suggest that, in the first instance, calnexin binds through its lectin site to the $Glc\alpha\text{-}1\text{-}_3Man_9GlcNAc_2$ sugar on the nascent glycoprotein. Calnexin then locates a secondary binding site on the protein, releasing the sugar from the lectin site and allowing access to glucosidase II. The protein-bound calnexin then functions as a conventional chaperone, preventing aggregation or deg-

radation and stabilizing folding or assembly intermediates. As long as the protein remains unfolded, G1 is not removed and remains available to interact again with the lectin site.

Calreticulin, which contains several regions with 42 to 78% sequence identity to calnexin (Wada et al., 1991), has been detected in a wide range of cell types, where it binds *in vitro* to a large number of proteins isolated from the ER. Although calreticulin binds optimally to Glcα-1-3$Man_9GlcNAc_2$, substantial interaction was retained after trimming of the mannose residues to Glcα-1-3$Man_5GlcNAc_2$ (Spiro et al., 1996).

Several models for a combined role for these chaperones have been proposed. For example, in a study of the HIV envelope glycoprotein GP160, Otteken et al. (1996) found evidence for a ternary complex of calnexin, calreticulin, and GP160. Sequential immunoprecipitation data suggested that, while most of the GP160 associated with calreticulin was also bound to calnexin, a portion of the GP160 was only associated with calnexin. This may suggest that calnexin is required for calreticulin binding to GP160. The study suggested several models; for example, calnexin may hold the newly translated protein in the ER while calreticulin associates with and dissociates from the protein. Alternatively, membrane-bound calnexin may bind to the protein through glycosylation sites close to the ER, whereas the soluble calreticulin may bind to glycans at sites further from the ER membrane. In the $Glc_1Man_1GlcNAc_2$ structure the G(3) is some 30 Å from the point of N-glycan attachment (Petrescu et al., in press). This suggests that recognition by calnexin is independent of the protein, and this would be consistent with the finding that a wide range of glycoproteins use this pathway.

The oligosaccharides on a glycoprotein do not invariably bind calnexin or calreticulin. At present, there are no rules by which it can be predicted which glycans will bind the chaperones, and each protein must be evaluated on a case by case basis. In some cases, it may be necessary to examine the role of the glycans in protein folding on a site by site basis, for while some glycosylated regions of a protein may use the calnexin pathway, other glycosylated regions may not (Figure 5B). An example is the heavily glycosylated HIV surface glycoprotein GP160, which is discussed below (see Section II.B.3.a).

When glycosylation sites were deleted from a number of proteins that depend on the calnexin/calreticulin folding pathways, incorrectly folded proteins were retained in the ER. The proteins that have been studied include human transferrin receptor, human glucuronidase, human influenza A virus hemagglutinin, and human-truncated FcεRI (discussed in Helenius, 1994). Interestingly, unglycosylated FcεRI is secreted normally from *Escherichia coli* (Letourneur et al., 1995), suggesting that in bacteria, mechanisms such as the GroEL and GroES chaperonins (Ellis, 1994) may assist protein folding. Importantly, complete enzymatic deglycosylation of the glycoprotein did not generally affect the 3D structure of these proteins, suggesting that while glycosylation is essential for folding to take place in the ER, once the proper folding has been achieved the sugars are not required to maintain it.

3. Glycosylation and Protein Folding and Assembly

Glycosylation plays an important role in protein folding and assembly. Some examples are discussed in the sections that follow. First, in the HIV surface glycoprotein GP120, different domains fold inde-

pendently to produce the correctly folded protein (Section II.B.3.a). Interestingly, the enzyme tyrosinase may have its folding pattern altered if the glycan processing is inhibited. This suggests an involvement of the calnexin pathway. Although the enzyme is secreted normally, its function is impaired (Section II.B.3.b). Similarly, at least some, if not all, of the regions of both the influenza A virus (Section II.B.4) and the hepatitis B virus envelope proteins (Section II.B.5) also seem to depend on the calnexin pathway to achieve correct folding and secretion. Inhibiting the glycosylation of both HBV and HA coat proteins produced misfolded protein that was unable to assemble into viral particles and was retained in the ER. Additionally, carbohydrates also have a role in the stabilization of the trimer formed by the influenza A virus hemagglutinin.

a. Domains in the HIV Coat Protein GP120 May Fold Independently

Although, in principle, the calnexin/calreticulin-dependent folding pathways are available to all proteins carrying N-glycans, glycoproteins do not necessarily bind these lectins. Moreover, in a single glycoprotein with more than one glycosylation site, the sugars at some individual sites may not bind to the chaperones. The interaction of the protein with other (nonlectin) chaperone molecules may make some specific sites inaccessible, whereas there may be other regions of the same protein that rely exclusively on this lectin pathway. As a result, different regions in one protein may fold independently using either calnexin/calreticulin-dependent pathways or pathways involving other chaperones. This has been shown for the HIV surface glycoprotein GP120 in a study in which the glycoprotein was expressed in CHO cells in the presence and absence of the ER glucosidase inhibitor n-BuDNJ. In the presence of nBuDNJ, none of the glucose residues attached to $Glc_3Man_9GlcNAc_2$ is removed, and therefore none of the glycans can interact with calnexin. Probing recombinant GP120 expressed in the presence of nBuDNJ with a panel of antibodies indicated that the conformation of the V1/V2 loops were altered, whereas those of the C1-5 and the V3 regions were unaffected (Figure 6) (Fischer et al., 1996a, b). This indicates that only the V1/V2 region requires the removal of the glucose residues for correct folding, and that other regions of the protein can fold normally even though this event has not taken place. This suggests that, while the V1/V2 regions may fold through a calnexin and/or calreticulin-dependent pathway, other regions fold independently by a different mechanism, for example, using the protein-binding chaperone BiP (Figure 5B) (Ellis, 1994; Flynn et al., 1991). In virus-infected cells that have been treated with nBuDNJ the Gp120 that is present on the viral surface is also misfolded, preventing it from undergoing conformational changes after CD4 binding. In the case of the hepatitis B virus (Section II.B.5), the geometry required for the assembly of the viral coat protein appears to be such that the virion fails to assemble when glycosylation processing is inhibited. In the case of tyrosinase (Section II.B.3.b) inhibition of the glycan processing of the sugars attached to tyrosinase leads to a viable protein but an inactive enzyme.

b. Inhibition of the Glycan Processing of the Sugars Attached to Tyrosinase Leads to a Failure to Bind Copper

Tyrosinase is a key enzyme in melanin biosynthesis in mammalian cells. It is synthesized in the melanocytes, from which the

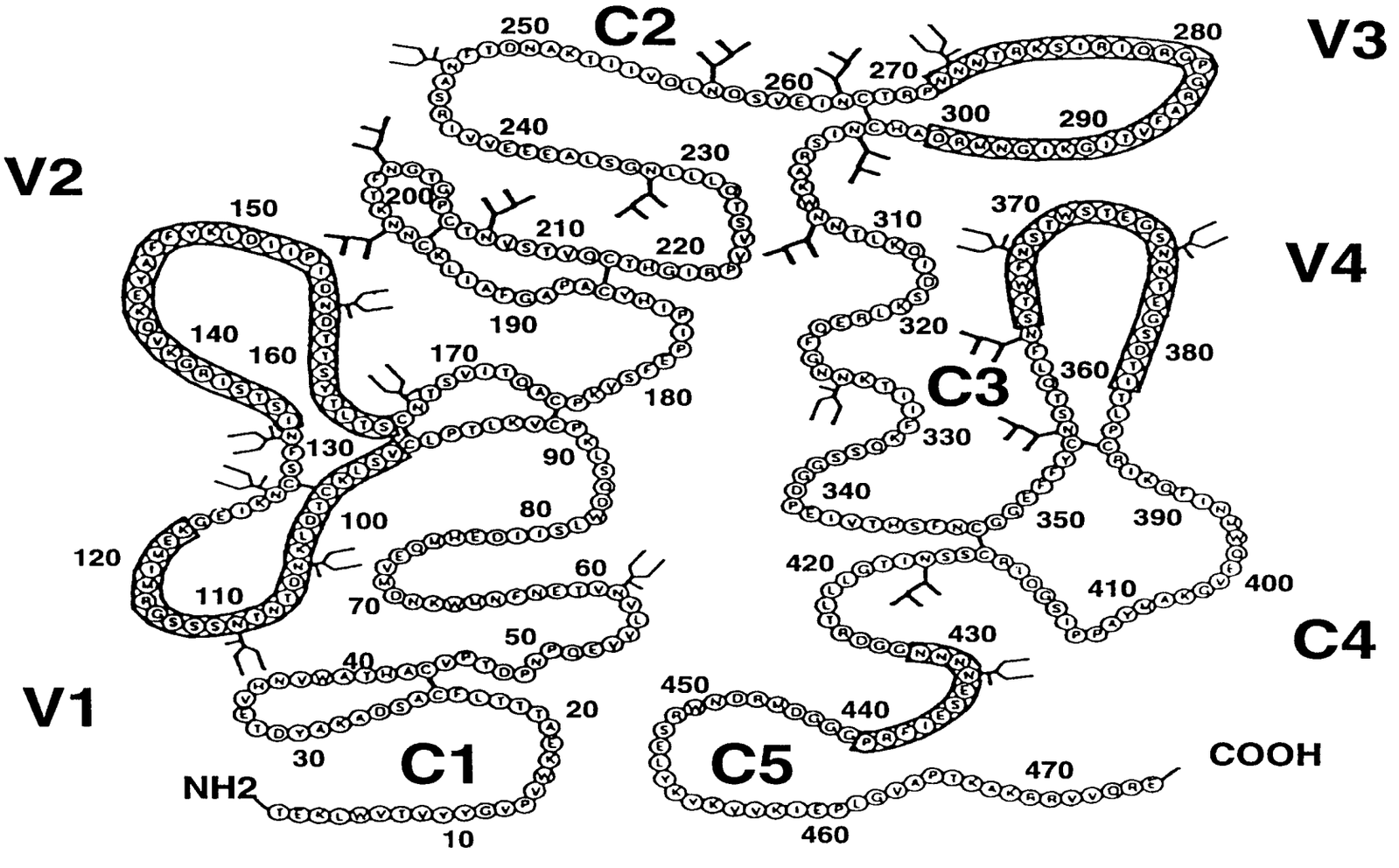


FIGURE 6. Schematic diagram of the HIV-1 envelope glycoprotein gp120. Glycosylation sites containing oligomannose and/or hybrid oligosaccharides and those containing complex-type oligosaccharides are indicated. The disulfide-bonded domains are labeled with Roman numerals, and the hypervariable regions are enclosed in boxes and labeled V1-V5. (Adapted from Leonard et al., 1990.)

mature protein is transported to specialized intracellular organelles known as melanosomes (Figure 7) (Laskin et al., 1986). Tyrosinase is a 60-kDa protein with six potential glycosylation sites, of which three are located within or close to the active site domains (Muller et al., 1988). Tyrosinase contains two copper-binding sites, in each of which three histidine residues co-ordinate a copper atom. Both copper atoms coordinate a molecule of O_2 (Jolley et al., 1974). The two copper atoms are essential for the activity of the enzyme.

In B16 mouse melanoma cells cultured in the presence of the ER glucosidase inhibitor nBuDNJ, the enzyme was synthesized and transported to the melanosomes (Petrescu et al., in press). However, analysis of the glycans indicated that three oligosaccharide structures ($Glc_3Man_{7-9}GlcNAc_2$) predominated, demonstrating that the processing of the glycans was incomplete. This resulted in the loss of copper binding and total inhibition of enzyme activity. This was not a result of steric hindrance by the abnormal sugars because activity could not be restored by removing the glucose residues. A similar effect was achieved when the cells were cultured with tunicamycin, which prevents the assembly of the dolichol phosphate precursor and leads to the synthesis of non-glycosylated tyrosinase. However, replacement of the glycans with oligomannose glycans, achieved by culturing with deoxymannojirimycin (DMJ), an inhibitor of Golgi α1-2mannosidase, had no effect on enzyme activity, indicating that the precise nature of the glycans on the mature glycoprotein is not important. Together, these data suggest that the inability to bind copper may be the result of misfolding around the catalytic site. Correct folding of the active site domains may depend on the interaction of the $Glc_1Man_{7-9}GlcNAc_2$ oligosaccharides with calnexin. In the presence of the ER glucosidase inhibitor n-BuDNJ, this intermediate is not formed.

4. Glycosylation and Subunit Assembly in Influenza A Virus Hemagglutinin

Influenza viral proteins, like other viral proteins, control the processing of their own sugars in a site specific manner using the glycosylation machinery available in the host cells (Figure 8) (Mir-Shekari et al., in press). Hemagglutinin (HA) is a type 3 membrane glycoprotein expressed on the surface of the influenza A virus as a homotrimer (Wilson et al., 1989). Carbohydrate covers 17 to 20% of the protein surface, and in the intermediate stages of folding and oligomeric assembly, the monomers are associated with calnexin. This indicates that some or all of the protein folds through association with this chaperone (Tatu et al., 1995). The absolute requirement for carbohydrate in the folding process was demonstrated by Sing et al. (1990), who showed that inhibiting the glycosylation of soluble, monomeric, anchor-free forms of the influenza HA precursor produced misfolded protein that was retained in the ER.

The assembly of the homotrimers occurs within the ER and the intermediate compartment without calnexin involvement (Tatu et al., 1995). However, at this stage, the carbohydrates are required to stabilize the trimer. The glycans shield a large area of the HA surface, mainly covering hydrophilic patches, with Ser and Thr residues most frequently interacting with the sugars (Wilson et al., 1989). Specific noncovalent carbohydrate-protein interactions are involved in the subunit–subunit interface and inter- and intrachain contacts that stabilize both the monomer units and the trimeric structure. In addition, by interacting with Ser and Thr residues, the N-linked sugars may prevent the addition of O-linked glycans to potential sites.

The glycans also protect the protein by covering two Lys, three Arg, and five aro-

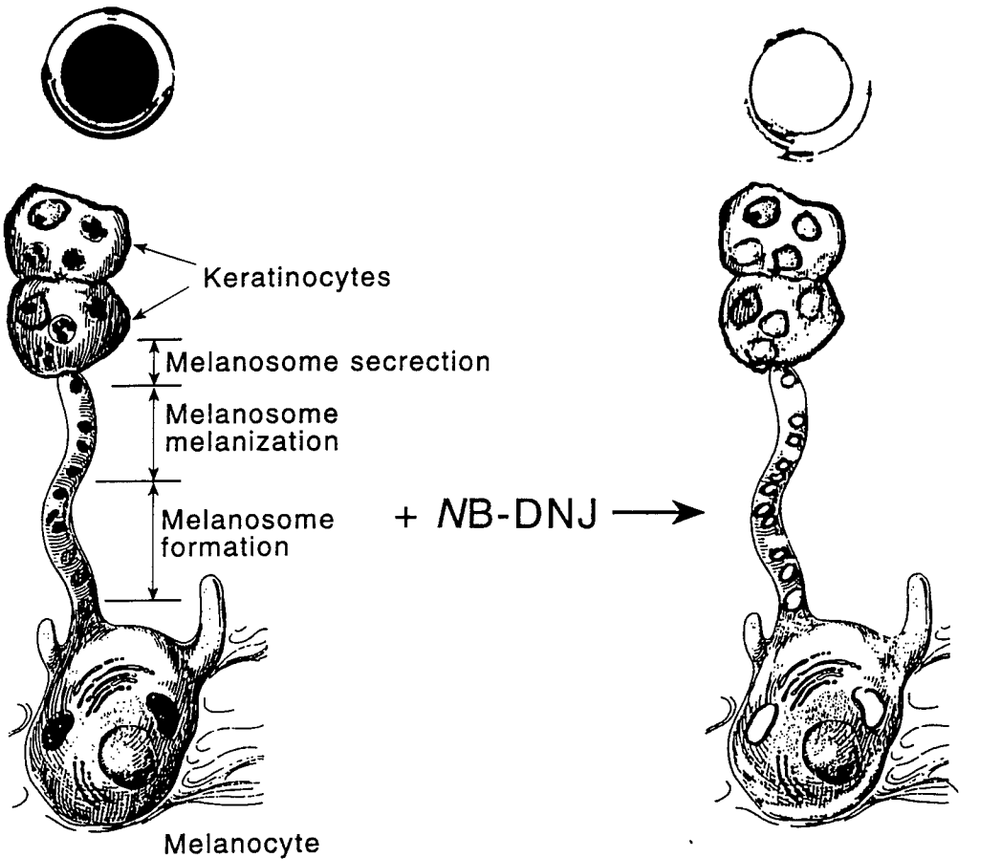


FIGURE 7. The function of tyrosinase. Tyrosinase is a key enzyme in melanin biosynthesis in mammalian cells. It is synthesized in the melanocytes from which the mature protein is transported to specialized intracellular organelles, known as melanosomes. Tyrosinase contains two copper atoms that are essential for the activity of the enzyme. In the presence of the ER glucosidase inhibitor, nBuDNJ, the enzyme was synthesized and transported to the melanosomes. However, the enzyme, which did not contain copper, was totally inactive, as illustrated in the right-hand figure.

matic residues that are sites for trypsin and chymotrypsin cleavage (Wilson et al., 1989). In addition, the sugars may modulate host recognition by shielding immunogenic epitopes from the host's immune system. The sugars themselves are poor antigens, having been processed by the biosynthetic pathway of the host. Importantly, although the sugars cover a large area of the protein surface, they do not obstruct either the receptor-binding pocket, which binds sialic acid on host cells or the cleavage of HA1/2. Both of these are essential features that maintain the infectivity of the virus.

5. Glycosylation and Multisubunit Assembly in the Hepatitis B Virus

The hepatatis B virus (HBV) specifies three envelope glycoproteins (L, M, and S) that together form the envelope of the virus. Figure 9A shows the amino acid locations, using HBV genotype adw as a basis. These proteins are related and result from the alternate translation initiation of a single open reading frame. L consists of three domains known as preS1, preS2, and S. M contains preS2 and S, whereas S contains only the

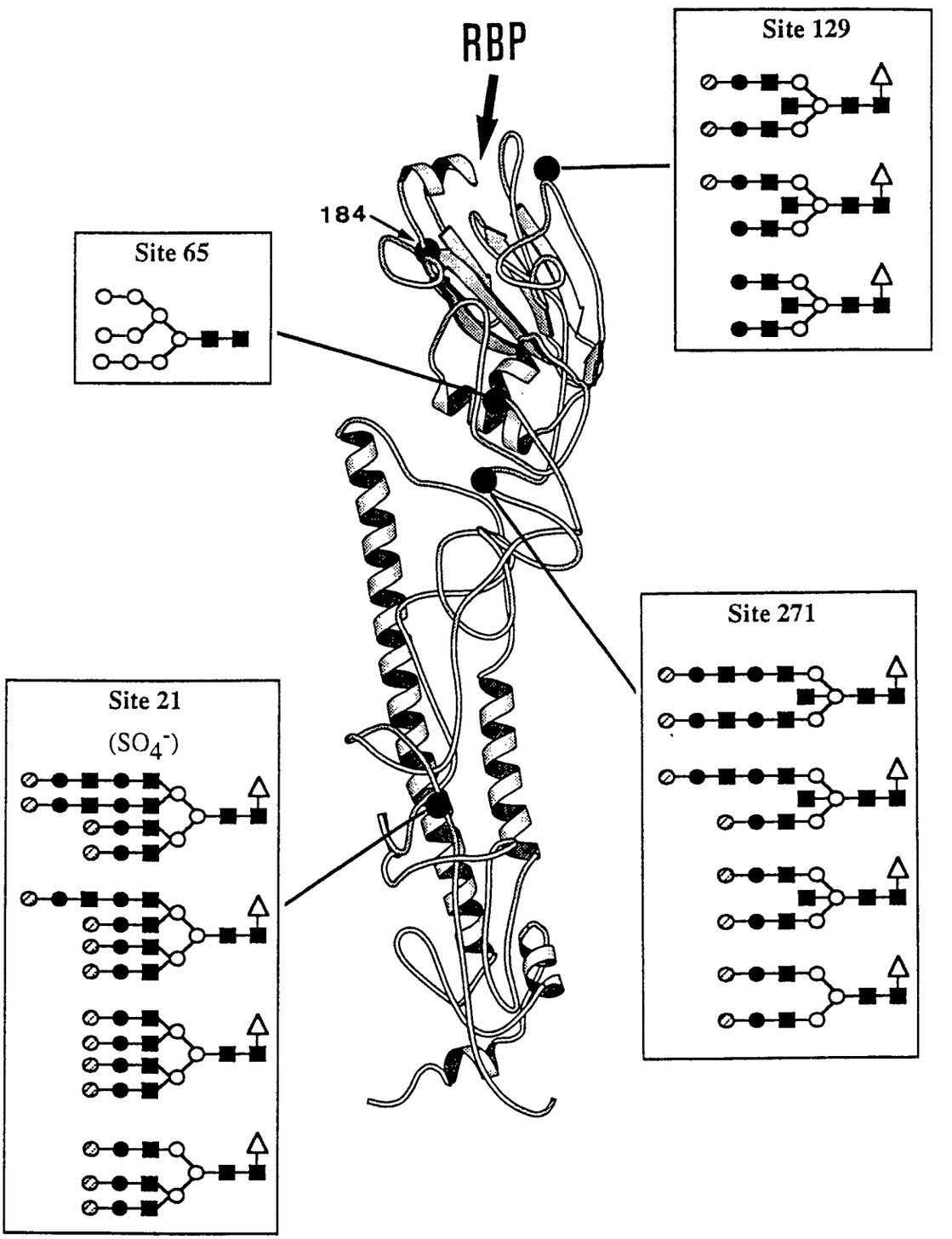
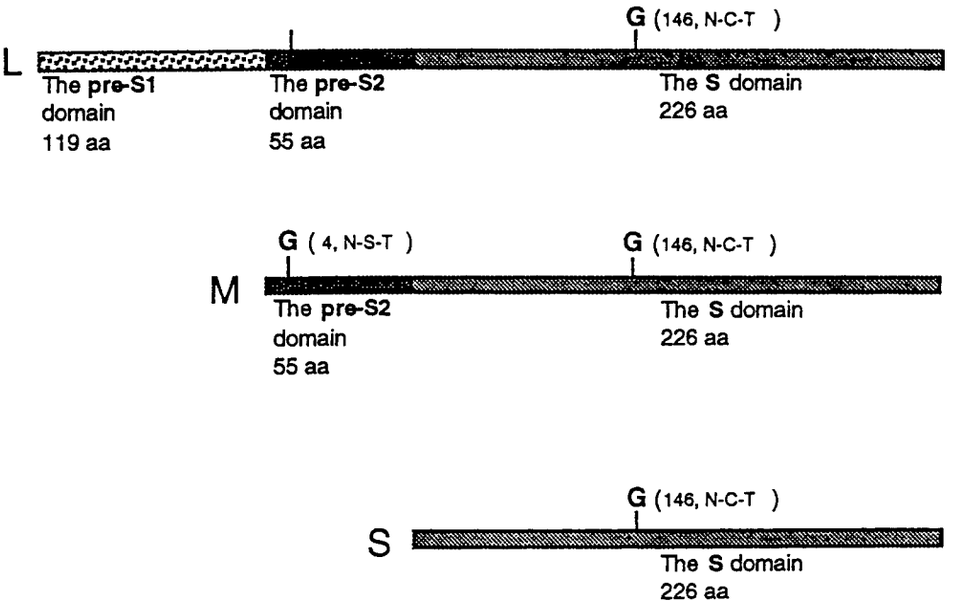
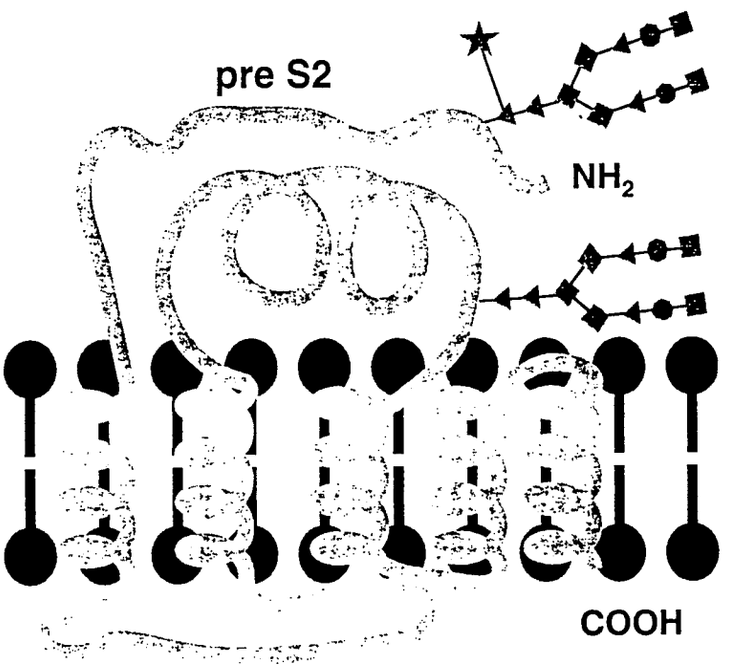


FIGURE 8. Diagram showing the major carbohydrate structures at each of the N-linked glycosylation sites in hemagglutinin expressed in MDBK cells. The receptor-binding pocket (RBP) is indicated with an arrow; the positions of glycosylation sites 21, 65, 129, and 271 are indicated by shaded circles. For the oligosaccharide structures, triangles represent fucose residues; closed squares, GlcNAc residues; open circles, mannose residues; closed circles, β-GalNAc residues; and hatched circles, α-Gal residues.



A



B

FIGURE 9. **(A)** Schematic drawing of the hepatitis B virus envelope proteins. All three proteins have a common N-linked glycosylation site near their carboxy terminus (G). The M protein contains a unique site at the amino terminal end. The L protein, although it contains the pre-S2 glycosylation sequon, only utilizes the shared S glycan site at Asn146. **(B)** Structural diagram of the M-protein with attached N-linked glycans. The extra pre-S2 domain of the M protein is indicated (Heerman and Gerlich, 1992).

single domain. All three proteins have a complex N-linked oligosaccharide at Asn146CysThr of the S domain. In addition, the M protein contains an additional glycosylation site (Asn4SerThr) at its amino terminus in the preS2 domain (Figure 9B). The L, M, and S proteins are incorporated into the envelope of noninfectious subviral particles, which do not contain DNA, and also into the envelope of the intact virus.

When HBV was expressed in HepG2 cells in the presence of the ER glucosidase inhibitor n-BuDNJ, only subviral particles were secreted, and these contained only the S and L proteins (Mehta et al., 1997). M protein and viral DNA accumulated in the cell. The data suggest that the assembly and secretion of the virus and the assembly of M protein into the subviral particles can be prevented by interfering with the glycosylation of the M protein.

In the ER, the S and L proteins are assembled normally into subviral particles even though the glycans they contain are hyperglucosylated. In the Golgi, the glycans on these subviral particles are processed to complex sugars, presumably through the endomannosidase pathway (Rabouille and Spiro, 1992), and the particles are secreted normally. This suggests that the common glycosylation site on S and L may not be involved in glycan-dependent folding.

In contrast, the unique site on the M protein may assist the folding of the preS2 domain in the M protein, either directly, by affecting the conformation of the region, or indirectly, by mediating an interaction with calnexin. Both of these events may be inhibited in cells cultured in the presence of n-BuDNJ, which prevents the removal of the two terminal glucose residues from the $Glc_3Man_9GlcNAc_2$ glycan precursor. Therefore, misfolded M protein, which cannot be assembled into either subviral particles or virions, accumulates in the cell. Interestingly, when the cells were cultured in DMJ, both the virus and M protein were normally glycosylated, assembled, and secreted. DMJ does not prevent the removal of the glucose residues in the ER, nor interfere with calnexin binding. It inhibits the Golgi mannosidase, leaving the $Man_{7\text{-}9}GlcNAc_2$ structure on the protein, and prevents further processing to complex glycans. This finding is consistent with the proposal that the specific complex glycans attached to the M protein are not required for the assembly of M into subviral particles or virions. However, correctly folded M is necessary. Correct folding seems to require the presence of a $Glc_1Man_9GlcNAc_2$ glycan in the ER, suggesting that the folding of the preS2 region of the M protein follows a calnexin-dependent pathway. Further, the possibility remains that in M (in contrast to L and S), the folding of the S region that contains the common glycosylation site, may also be calnexin dependent.

6. Glycosylation and Protein Degradation in the ER

Proteins that are incorrectly folded or that fail to assemble into the correct oligomeric complexes are not normally released from the ER (Hurtley and Helenius, 1989). For example, the T cell receptor (TCR) is composed of at least seven subunits, and neither partial complexes not individual subunits are expressed on the protein surface. In a study of this receptor complex, Lippincott-Schwartz et al. (1988) identified an independent proteolytic system in a pre-Golgi compartment where such misfolded or unassembled proteins are degraded. For example, in the absence of light chains, IgG heavy chains are also degraded in the ER (Dulis et al., 1982) as are a range of other proteins, including IgM monomers or oligomers, which are not assembled into pentamers, and incorrectly folded asialoglycoprotein receptors (for a review see Klausner and Sitia, 1990).

A role for ER proteases in the generation of peptides for presentation to major histocompatibility (MHC) class I-restricted cytotoxic T lymphocytes (CTL) has also been demonstrated. Cells lacking TAP(transporter associated with antigen processing) are unable to present cytosolically derived antigen to CTL, and this defect can be overcome by targeting peptides known to be class I epitopes to the ER using a leader sequence. In such circumstances, proteases in the ER are also capable of generating an epitope from a large fragment of influenza nucleoprotein (NP) but not from the full-length protein (Elliott et al., 1995). This full-length NP contains two sequons for N-linked glycosylation (both of which are used) and is degraded only slowly (50% remaining after 8 h in pulse-chase studies) after an initial lag period. Prevention of glycosylation by tunicamycin treatment markedly increases the rate of degradation of the protein and restores the ability of TAP cells to present epitopes derived from NP to CTL. A mutant NP with the glycosylation sites removed by site-directed mutagenesis behaves in a similar fashion (Wood and Elliott, unpublished results). Glycoproteins may thus be protected from the action of ER proteases by virtue of their N-linked glycans, during an initial "trial-period" of folding, mediated perhaps via calnexin/calreticulin. Subsequently, these molecules may function to transport newly synthesised glycoproteins away from or toward areas of the ER that are rich in proteases.

The conclusions from this exploration of the role of glycosylation within the ER suggest that there are some roles for glycosylation that may be performed by any one of a range of sugars attached to a single glycosylation site. For example, several different sugars may serve to protect residues on correctly folded proteins from cleavage by proteases that are part of the ER degradative pathway (Klausner and Sitia, 1990). Although this has not been demonstrated specifically in the ER, it has been shown *in vitro* for RNase B that a range of oligomannose structures, as well as the unsubstituted chitobiose core, gave some protection from digestion by pronase and trypsin compared with the unglycosylated protein (Rudd et al., 1994a).

Alternatively, there are some recognition processes for which specific glycan structures are required. For example, in calnexin-dependent protein folding, the protein domain must contain a Glc_1ManD1 oligosaccharide epitope within a $Glc_1Man_{7-9}GlcNAc_2$ structure. However, as in the HIV surface glycoprotein GP120 (Section II.B.3.a), not all domains that present this sugar use the calnexin pathway, suggesting that other chaperones may compete more effectively for the protein, possibly hindering the calnexin binding. Alternatively, in some cases it has been suggested that the protein may contain a secondary binding site for calnexin that promotes the interaction, as in HLA class I heavy chains (see Section II.B.2). The importance of the interrelationships between proteins and their sugars and the need to view a glycoprotein and its array of sugars as a single entity if the full implications of glycosylation are to be understood is highlighted throughout this review. This concept is particularly relevant to the enzyme interactions that develop glycan heterogeneity within the Golgi.

III. GLYCOSYLATION IN THE GOLGI

A. The Development of Glycan Heterogeneity in the Golgi is Protein Specific

The individual protein may interact with the glycosylation machinery in the Golgi to develop particular motifs required for spe-

cialized recognition functions. One example is the specific recognition marker sialyl Lewis X, which is recognized by L selectin and is involved in the adhesion of circulating leukocytes to endothelium during inflammation (Bevilacqua, 1993). In other cases, specific oligosaccharides may stabilize or protect the protein by interacting with specific amino acid residues on its surface, for example, the complex biantennary N-glycans in IgGFc (Leatherbarrow, 1984; Padlan, 1991; Padlan, 1994) and the influenza A HA (Wiley et al., 1989) or the sialylated GalGalNAc O-glycans in IgA (Mattu T. S., unpublished data). The precise structure of the glycans may also be a factor that controls the half-life and destination of a protein. For example, restricting sialylation or increasing the proportion of triantennary glycans in a glycoform population may increase the turnover rate of the protein because asialo, triantennary glycans have a higher affinity for the asialoglycoprotein receptor than sialylated biantennary oligosaccharides (Section IV.E.2) (Lee et al., 1983; Pepys et al., 1994).

There may also be general roles for the complex-type glycans that are attached to many human cell surface proteins. For example, complex sugars, which frequently terminate in galactose or sialic acid, may serve to distinguish human cells from yeast or bacterial cell walls that carry multiply presented oligomannnose sugars. Recognition of yeast and bacteria by the mannose-binding protein is an important feature of the innate immune defense system (Section IV.E.2).

1. The Efficiency of Key Enzyme Reactions Enzymes and the 3D Structure of the Protein

In a single cell, all proteins trafficking through the ER and Golgi encounter a common glycosylation pathway that is defined by the levels and locations of the glycosylating enzyme arrays, and several key enzyme reactions are involved in the biosynthesis of N-linked oligosaccharides. The interactions of these enzymes with the developing glycoprotein determine which of the potential glycosylation sites will be occupied, the class of the oligosaccharides that will be located at each site, and the terminal residues on the sugars. The efficiency of these key enzyme reactions, and those that refine the basic structures, determine the final structures of the glycans (for a review see Schachter, 1986).

A number of features of the 3D structure of the individual protein may influence the specific glycosylation patterns at each site by limiting sterically the access of enzymes and sugar nucleotides to the developing glycan chain. These include (1) the relative positions of the subunits and the domains within a subunit (quaternary structure), (2) the structure of the domain to which the sugar is attached (tertiary structure), (3) the local 3D structure of the domain to which the glycans are attached, for example, loops, helices, or pleated sheets and the position of disulfide bonds (secondary structure), (4) the structure and orientation of the amino acid side chains within and around the glycosylation site (primary structure), and (5) the position of the glycosylation sequon within the primary sequence of the peptide (primary structure).

Within the Golgi, in a given cell, the sequential additions of monosaccharide residues to $Man_5GlcNAc_2$ are events that are controlled mainly by the protein structure (Parekh et al., 1987). We recently demonstrated this for a series of leukocyte antigens that had characteristic glycosylation patterns even though they were all expressed in the same CHO cell line. Many parameters combine to ensure that each develop-

ing glycan chain on every protein is uniquely presented within a defined structural environment. Thus, within a particular cell, every protein can achieve specific and reproducible glycosylation at each glycosylation site. While the relative importance of the various structural features of the protein in the processing of its glycans will be decided on a case by case basis, some general principles are emerging.

2. The Development of Hybrid or Complex Glycoforms

In the cis Golgi, $Man_6GlcNAc_2$ glycans may be trimmed to $Man_5GlcNAc_2$ (Figure 1A). In the context of an appropriate protein structure, $Man_5GlcNAc_2$ is a substrate for *N*-acetyl glucosaminyl transferase I that substitutes *N*-acetyl glucosamine (GlcNAc) on the α-1,3 arm of the glycan. This reaction is a key step in the processing of oligomannose structures to both hybrid and complex glycans (for review see Natuska and Lowe, 1994; Schachter, 1986). Whether the glycans will be processed from hybrid to complex sugars depends on whether the 3D structure of the protein can allow the two remaining terminal mannose residues on the α-1,6 arm of the glycan to be removed by α-mannosidase II. The product of this reaction is a substrate for β-*N*-acetyl glucosaminyl (GlcNAc) transferase II, which adds GlcNAc to the α-1,6 arm of the glycan, forming a biantennary glycan and committing the glycoprotein to the complex sugar-processing pathway. If GlcNAc transferase III introduces a GlcNAc residue in a bisecting position between the two arms, further branching is blocked (Harpaz and Schachter, 1980).

Alternatively, it has been suggested that some proteins may contain positive signals responsible for the branching of their glycans to tri- and tetraantennary complex structures through the GlcNAc transferase enzymes IV and V (Figure 10) (Schachter, 1986). This suggests that the protein may regulate the extent of branching, not only by steric hindrance, but also by stimulating particular interactions.

The complex processing pathway may terminate before or after the addition of galactose to one or more arms of the glycan or continue to extend the complex glycan in an arm specific manner (Fukuda 1991) by the addition of polylactosamine units (Galβ-1,4GalNAc). In a tetraantennary structure the extensions occur preferentially on the GlcNAcβ-1,6Manα-1,6GlcNAc2 arm (Figure 11) (Fukuda and Hindsgaul, 1994). The amino acid sequences of proteins carrying polylactosamines have been compared, but no consensus sequence has emerged that could provide a signal. However, the possibility that there may be a 3D signal sequence in the protein cannot be ruled out. At present, the data suggest that glycoforms carrying lactosamine extensions arise mainly as a result of particular cellular pathways. For example, many proteins expressed on the surface of human erythrocytes, such as band 3 protein (Fukuda et al., 1984) and CD59 (Rudd et al., 1997) (Figure 12), contain lactosamine-type glycans. Interestingly, when expressed in CHO cells, the leukocyte antigens, CD59, CD48, and CD2 all contained the same range of polylactosamine glycans in similar proportions (Figure 13).

3. The Competition between Extension and Termination

Processing may terminate at any stage in the pathway. However, the addition of sialic acid is invariably a chain termination step. In contrast to other monosaccharide

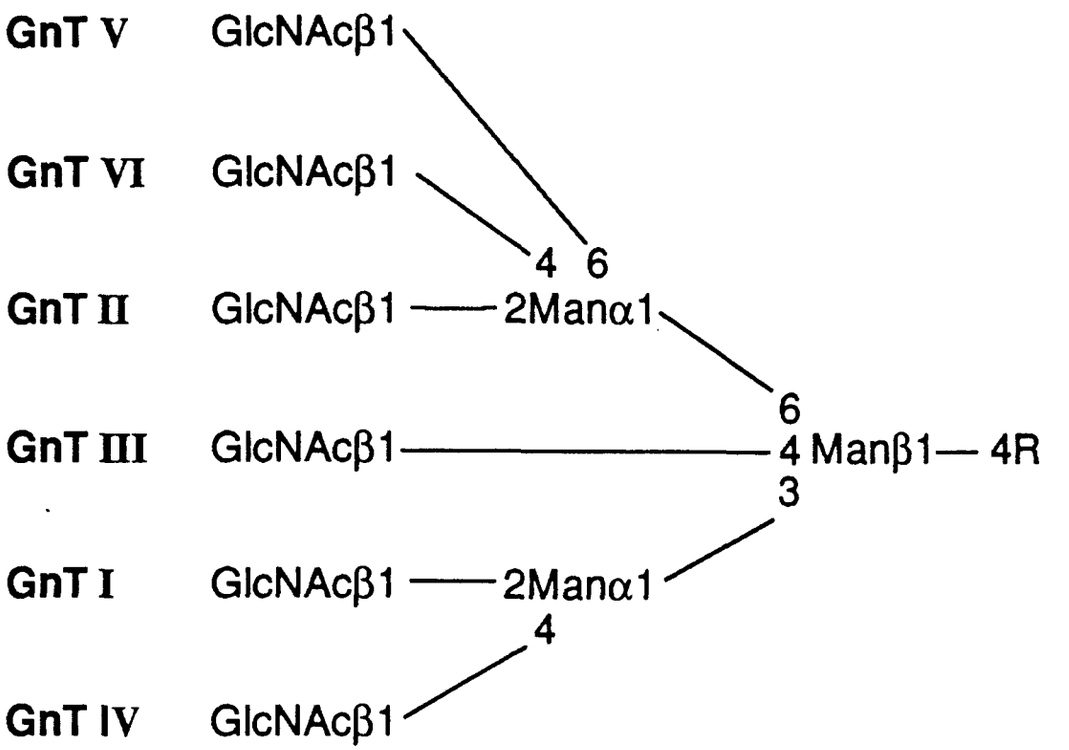


FIGURE 10. The 'branching' GlcNAc transferases (GnTs). Five antennae can be initiated on the Manα1-3Manβ1-4GlcNAcβ1-4GlcNAcβ-Asn core of N-glycans by the actions of GnT I, II, IV, V, and VI. A bisecting GlcNAc can be added by GnT III. GnT VI has not been detected in mammals.

residues, sialic acid is normally found only at the nonreducing termini of N-glycan chains, where it is linked α–2-3,-6, or, rarely, -8 to terminal galactose. Although some glycosyl transferases have been located to specific regions of the Golgi stacks (Kornfeld and Kornfeld, 1985), other glycosyltransferases are diffused throughout. One of these is β-galactoside α-2,6-sialyltransferase. This enzyme competes with β-*N*-acetyl glucosaminyl transferase for terminal galactose residues, thus initiating a competition between chain elongation (which results from the addition of GlcNAc to terminal galactose) and chain termination (which results from the addition of sialic acid to terminal galactose) throughout the *cis*- and the *trans*-Golgi.

Interestingly, *in vitro* studies with purified sialyltransferases have demonstrated that Galα-2,6-sialyl transferase also competes with Galα-2,3-sialyl transferase (Weinstein et al., 1982). The addition of sialic acid to complex glycans has been explored in studies of several glycoproteins. The relationship between chain length, charge, and sialic acid linkage has been investigated in glycoproteins from a mouse lymphoma cell line (Merkle and Cummings, 1987), thyroid glycoprotein (Edge and Spiro, 1985), and fetal lactosaminoglycan (Fukuda et al., 1984). In general, these studies show that sialylation is more likely to occur in the early stages of processing, and the incidence falls off as the chains become more elongated. Moreover, the sialic acid present on shorter oligosaccharides (less than three linear units of the repeating disaccharide (Galβ-1,4GlcNAc) is more often α-2,6, while the α-2,3 linkage is more common in the rela-

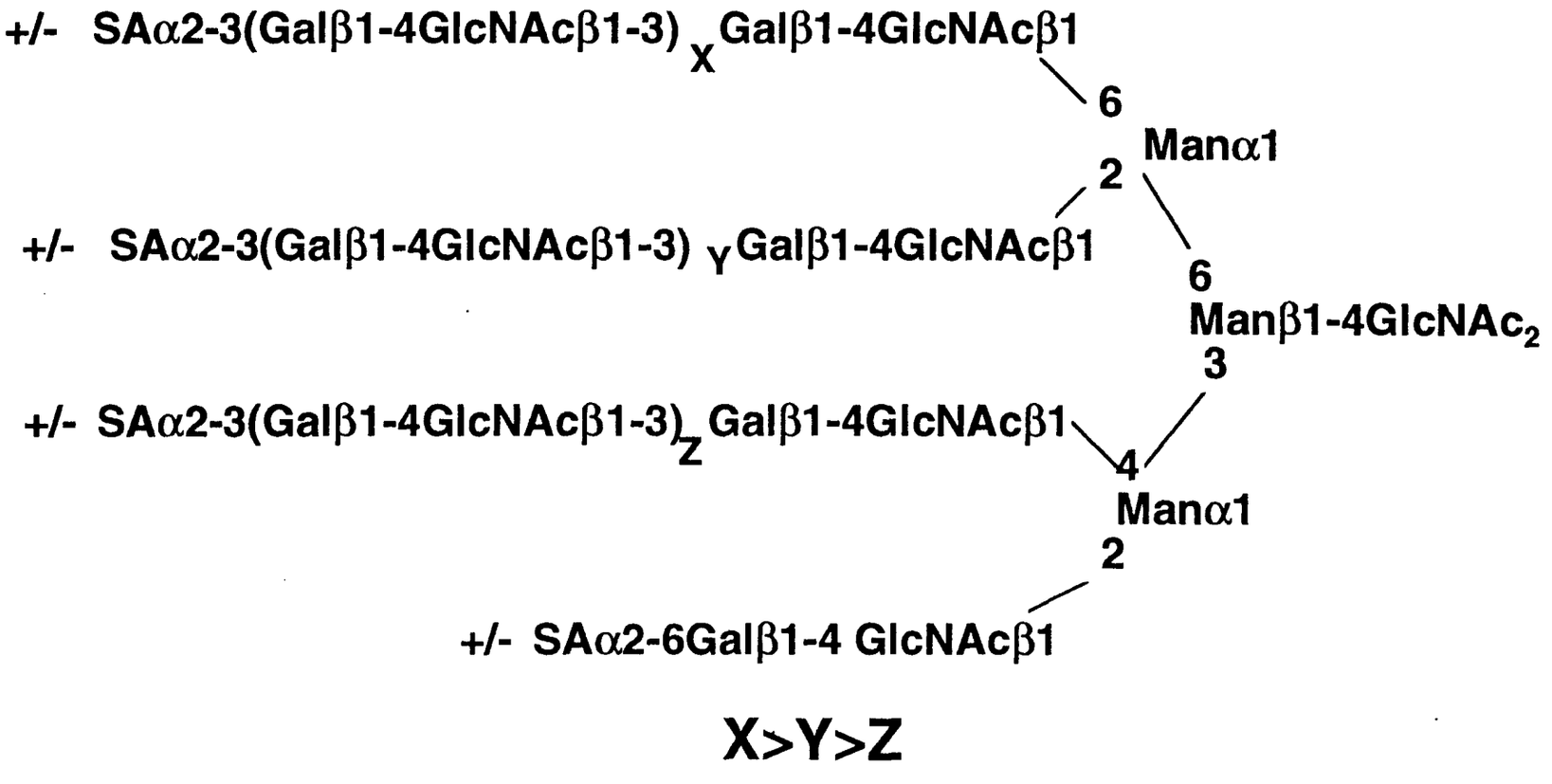


FIGURE 11. Complex tetrantennary structure with poly-*N*-acetyllactosamine extensions on three branches. The lactosamine extensions (x, y, and z) are preferentially added to the α-1,6 arm of the Manα-1,6 branch of the primary mannose (x), which is > y > z. Sialic acid is more commonly linked in the α-2,3 position to the Manα1,6 arm of the primary mannose, while the α-1,2 branch of the α-1,3 arm is more often associated with α-2,6-linked sialic acid (Fukuda, 1994).

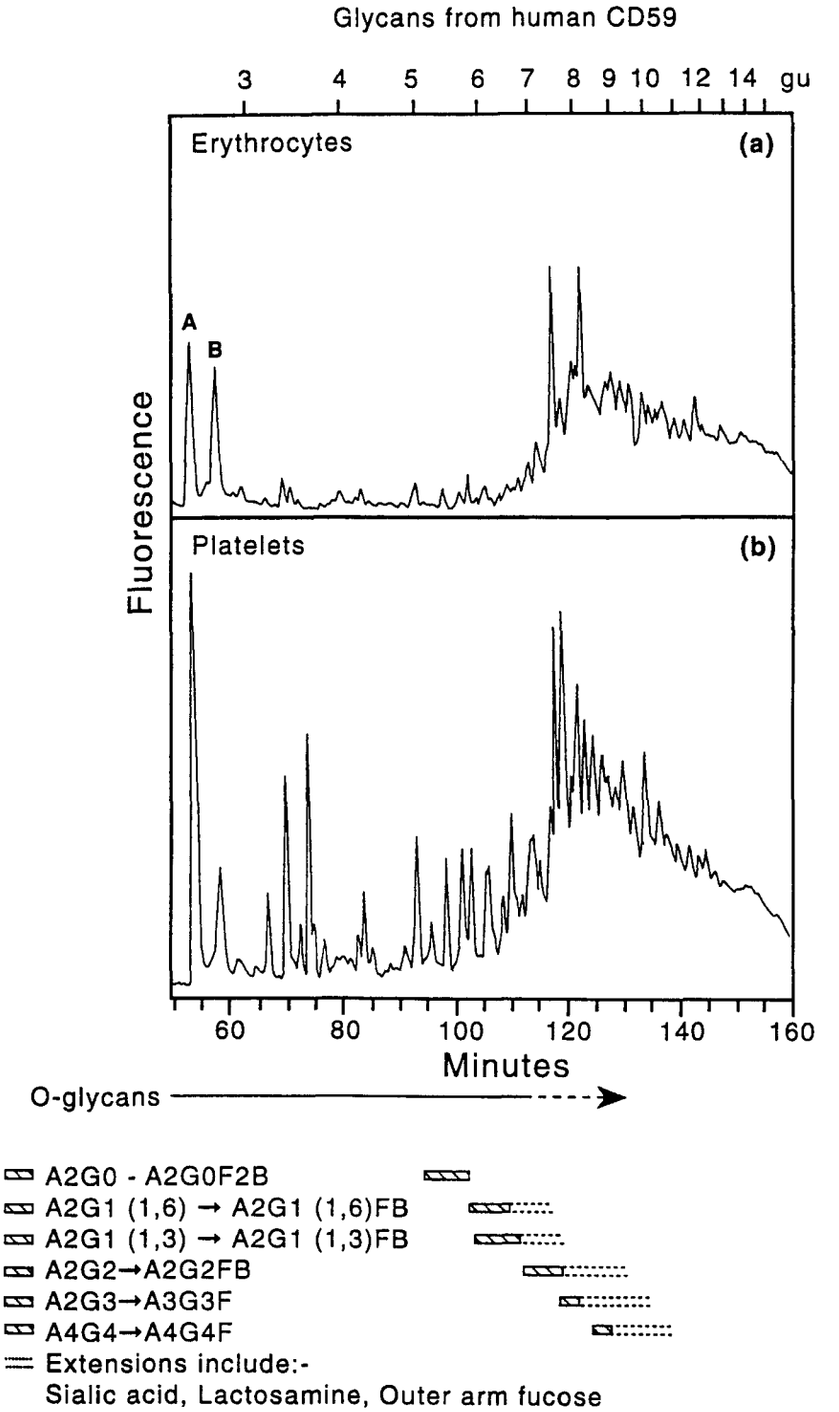


FIGURE 12. Normal phase HPLC profiles of fluorescently labeled glycans released from CD59 isolated from (a) human erythrocytes (b) human platelets. The elution positions of the dextran ladder used as a standard and of the different classes of oligosaccharides are shown in the scales above and below the figure, respectively. Structures and compositions were initially assigned on the basis of the gu values of the peaks, the elution positions of known standards, and the incremental values for the addition of monosaccharide residues (derived in Guile et al., 1996). **A** and **B** are sugars eluting in the region of O-glycans.

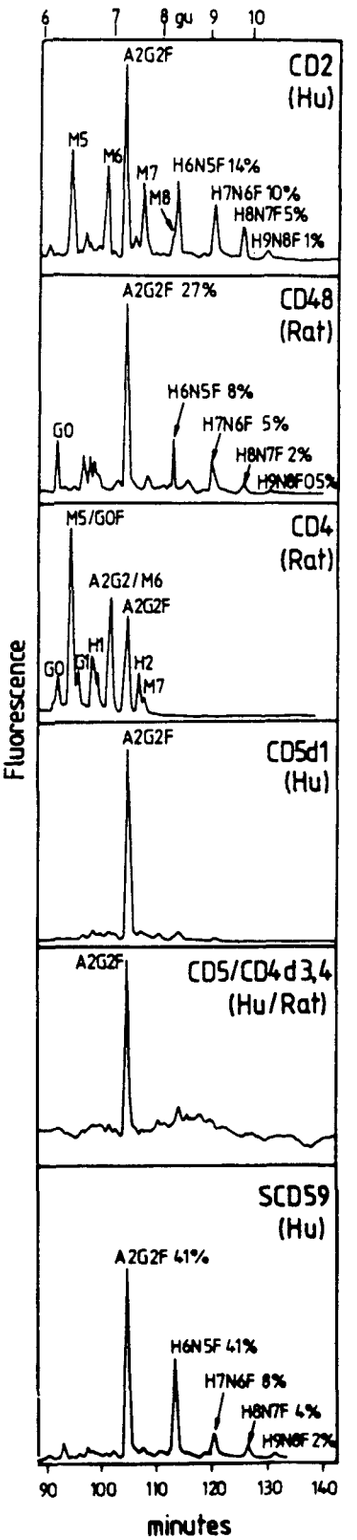


FIGURE 13. Comparison of the normal-phase HPLC profiles of fluorescently labeled oligosaccharides released from soluble leukocyte antigens.

tively less sialylated longer polylactosamine extensions (Merkle and Cummings, 1987). This suggests that the length of the glycan chain and the proximity of the protein matrix are factors that regulate the addition of α-2,3- or α-2,6-linked sialic acid. It has been suggested that sialylation may be controlled by a recognition sequence for sialyl transferase in the glycoprotein. In support of this, sequence analyses of cloned transferases have identified a conserved peptide motif. This 'sialyl motif' contains eight invariant amino acids (C, V, L, G, R, D, V, G) embedded within otherwise dissimilar sequences (Kitagawa et al., 1993). However, recently, this motif has been shown to participate in binding the donor substrate, CMP-NeuNAc, rather than the glycoprotein substrate (Datta and Paulson, 1995).

4. The Development of Glycan Heterogeneity (Glycoforms) within the Golgi Stacks

Although all N-linked glycoproteins enter the oligosaccharide-processing pathway as a single $Glc_3Man_9GlcNAc_2$ glycoform, all but a very few emerge from the *trans*-Golgi as a heterogeneous mixture of glycosylated variants. The development of the extensive heterogeneity associated with most glycoproteins poses major questions for glycobiology.

For example, do cells promote heterogeneity by ensuring that proteins compete for limited amounts of glycosylating enzymes, or does the protein control heterogeneity by modulating the efficiency of the enzymes, either directly by means of its 3D structure or indirectly through transit time, thereby ensuring that some processing steps fail to reach completion? If several enzymes compete simultaneously for a developing glycan, does the primary, secondary, tertiary, or quaternary structure of an indi-

vidual protein selectively influence the outcome, or does the cell alone control glycan processing?

5. Proteins with Single Glycoforms

If the structure of the developing glycoprotein were the only factor controlling glycosylation, proteins would emerge as single glycoforms and there would be no heterogeneity. In fact, examples of homogeneous glycoproteins are rare. One example is soybean agglutinin from plants (Ashford et al., 1991), which contains only the $Man_9GlcNAc_2$ sugar. Another is the soluble CD5 domain 1 expressed in CHO cells (see Section IV.B.1). CD5 domain 1 contains one glycosylation site. In CHO cells, this site contains a biantennary, core-fucosylated complex glycan (A2G2F) that varies only in the extent of sialylation. An example of a mammalian protein that is almost exclusively glycosylated with one sugar (A2G2) is normal human serum transferrin. Another serum glycoprotein that has only one glycoform is serum amyloid protein (SAP) (Figure 14). This contains the biantennary complex sugar (A2G2) in the neutral, mono-, or disialylated forms (Pepys et al., 1994). In this case, it is not clear whether the single glycoform represents a unique biosynthetic pathway or whether other less processed

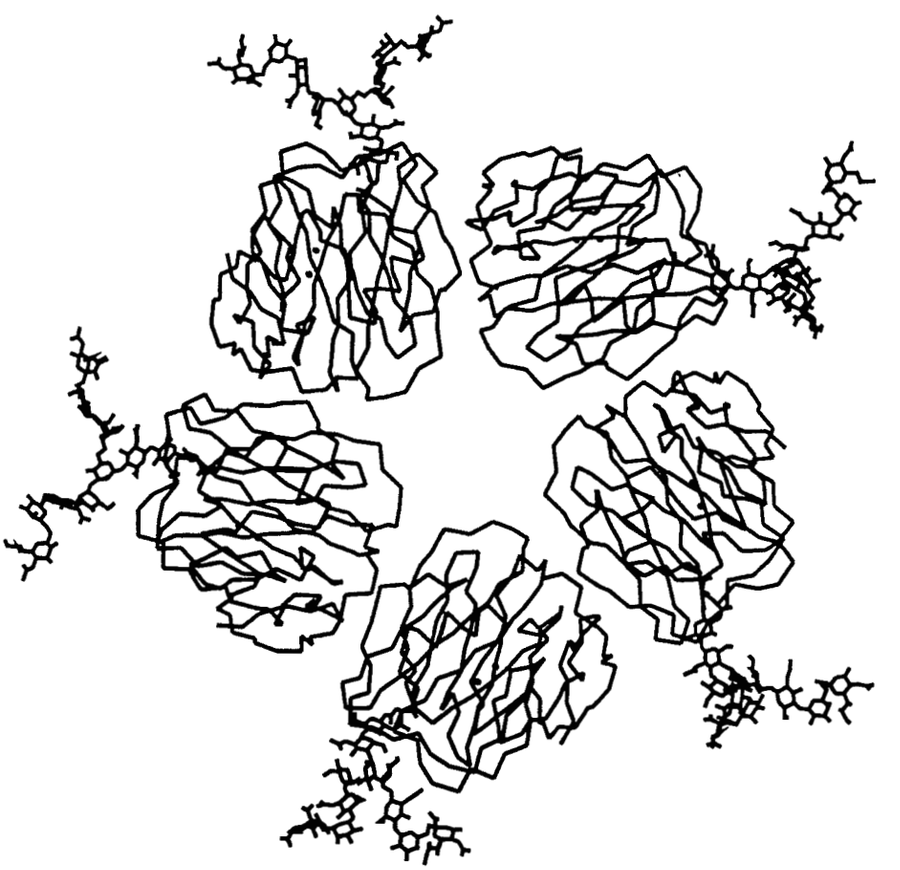


FIGURE 14. Molecular model of human serum amyloid protein pentamer. The model is based on the X-ray structure of the protein (Emsley et al., 1994).The sugar attached at Asn32 in each protomer is the disialylated biantennary complex glycan (Pepys et al., 1994).

glycoforms are removed from the serum, for example, via the asialoglycoprotein receptor. The N-glycosylation site of SAP (Asn32) is located on a β-strand lying underneath the single α-helix on the face of the protomer opposite to the calcium and ligand binding site, and the sugars extend outward from each protomer (Emsley et al., 1994). The glycans are attached to the protomer face that is involved in the pentameric interaction, and the oligomannose sugars present on the protein in the ER may contribute to decamer assembly. Interestingly, the clearance rate of normally sialylated SAP was unaffected by a blockade of the asialoglycoprotein receptor with galactose (Pepys et al., 1994). This indicates that normal SAP turnover may involve a protein rather than a carbohydrate recognition step and is consistent with the finding that the affinity of the biantennary glycan for the receptor is low (Section IV.E.2).

B. The Development of Glycan Heterogeneity in the Golgi is Cell-Type Specific

The observation that fully processed glycoproteins normally exist as mixtures of glycoforms suggests that the control that the 3D structure of the protein has over its own glycosylation is secondary to that exercised by the cell. This allows the cell the possiblity of responding to changes in its external and internal environment by altering the structure of the glycoform populations of the glycoproteins that it secretes. For example, in some cancer cells, the cell surface glycosylation is altered (Hiraizumi et al., 1992; Itzkowitz et al., 1990; Leathem and Brooks, 1987) and in inflammation the glycans attached to α1-acid glycoprotein become more fucosylated and sialyl lewis X structures are expressed on the cell surface (De Graaf et al., 1993). More work is needed to understand the functional implications of such changes.

1. Glycan Processing Depends on Features of the Glycosylation Machinery in the Individual Cell

Although the 3D structure of the protein can exert some control over its glycosylation in its initial stages, the oligosaccharide profiles of glycoproteins are normally characteristic of the cell in which the protein is expressed. For example, a protein such as t-PA, which has been expressed in many cell lines, is associated with several different glycosylation patterns, each of which is characteristic of the glycosylation of t-PA in a particular cell (Parekh et al., 1989; Rudd, P. M., 1995). Another example of cell-specific glycosylation is human CD59, which has a different pattern of glycosylation depending on whether it is expressed in human erythrocytes (HuE) or platelets (Figure 12). Again, human secretory RNases derived from kidney (Mizuta et al., 1990), pancreas (Ribo et al., 1994), and semen (De Prisco et al., 1984) are differently glycosylated from each other and do not contribute to urinary RNase (Hitoi et al., 1987), the source of which has not yet been established.

The processing of N-linked glycans has been shown to depend on cellular factors such as the enzyme repertoire, the competition between different enzymes for one oligosaccharide substrate, and the transit time. Additional factors that may influence glycan processing include the levels of the sugar nucleotide donors and competition between different glycosylation sites on the protein for the same pool of enzymes.

2. Heterogeneity and the Enzyme Repertoire in the Cell

At any time, many different glycoproteins may be trafficking through the glycosylation pathway, simultaneously competing for the active sites of the enzymes which, in general, are located in specific regions of the Golgi. The oligosaccharides attached to some glycoproteins are not processed by some enzymes and, as a result, microheterogeneity develops within the glycan population. Glycoproteins containing a range of glycans representing different stages of processing are therefore transported into the next subcellular compartment, where they encounter a different set of enzymes. Because a glycoprotein that fails to be processed by one enzyme cannot become the substrate for the next, modification of the glycan chains will terminate whenever the developing glycoform is unable to interact further with the specific glycosylation machinery still available in the cell. The glycoforms that emerge from the *trans*-Golgi are therefore the end products of a series of incomplete enzyme reactions.

When a glycosylating enzyme is present in sufficiently high concentrations, the accessible sugar chains on a protein will be processed. For example, in human serum transferrin, 98% of the glycans at the two glycosylation sites are processed to fully sialylated, complex biantennary structures. Alternatively, if the enzyme concentrations are low, the protein structure at each site will determine the relative reaction rates and therefore which sugars will be more likely to be processed. For example, in normal human serum IgG, the glycosylation sites in the Fab region contain approximately 70% of fully galactosylated sugars while the Fc contains only 37%, reflecting the difference in the accessibility of the two sites (Wormald et al., in press). During pregnancy, when the levels of β-galactosyl transferase rise, the percentage of galactosylated structures increases. In rheumatoid arthritis, when the activity of β-galactosyl transferase falls, there is a decrease in galactosylated sugars. These data indicate that the glycosylation status of IgG is controlled initially by the cell and secondly by the 3D structure of the protein.

Clearly, cell-specific glycosylation also reflects the enzymes that are present in the cell. In some cases, the genes for certain transferases are missing; for example, human cells do not contain a gene for α-galactosyl transferase, which is expressed by murine or trypanosome cells. In other cases, a gene may be 'silent', as in CHO cells that do not normally express 2,6 sialytransferase enzyme.

3. Glycan Processing and Transit Time

Interestingly, increased transit times through the glycosylation pathway, achieved by lowering the temperature of cell cultures to 15°C, have been shown to increase the polylactosamine content of the glycans attached to LAMP proteins-1 and -2 (Wang et al., 1991). One factor that may determine the percentage of polylactosamine structures may be the attachment of a GPI anchor. This decreases the transit time of the glycoprotein, but may bring the sugars into closer contact with the glycosylating enzymes that are also located in the membranes of the ER and Golgi. In the GPI-anchored protein HuE CD59, approximately 40% of the glycans were of the polylactosamine type, of which 30% contained more than 15 monosaccharide residues (Figure 12). In contrast, human soluble CD59 expressed in CHO cells did not contain significant amounts of these large polylactosamine structures (Rudd et al., in preparation). However, opposed to the suggestion that the presence of a GPI

anchor increases the percentage of polylactosamine structures is the data from the analysis of the GPI anchored glycoprotein Thy-1 from rat, mouse and human neural tissue, in which no polylactosamine structures were detected (Williams et al., 1993).

4. Multidomain Proteins and Cell-Specific Glycosylation

In multidomain proteins the quaternary structure may influence glycosylation even when the polypeptide structure and the glycosylation machinery remain constant. For example, a study has been made of the MAC-1 integrin that contains the a^M subunit and LFA1 integrin that contains the a^L subunit. These glycoproteins have a common β-subunit and the αβ-dimers form before processing, making it possible to compare the glycosylation of the β-subunit sites and to determine the influence of different α-subunits on processing (Dahms and Hart, 1986). Structural analysis of oligosaccharides from each of the corresponding glycopeptide fraction of the β-chains of MAC-1 or LFA-1 indicated that there were significant differences in the glycosylation of at least four of the corresponding sites on the β-chains, although they were simultaneously expressed in the same cells. The role of tertiary and quaternary structure in controlling glycan heterogeneity is explored further in Section IV.A.12.

C. O-Linked Glycosylation: Initiation and Development of Heterogeneity

1. The Primary Sequence of the Protein and Initiation of O-Glycosylation

All Ser(Thr) O-linked glycosylation is initiated by the incorporation of a single *N*-acetylgalactosamine (GalNAc) residue from UDP-GalNAc into the polypeptide chain, and elongation requires the sequential addition of monosaccharide residues. At least eight core structures have been identified (Figure 1C) (Hounsell et al., 1996). In general, no consensus sequence is required for O-glycosylation; however, it has been noted that acidic residues at position-1 relative to the glycosylation site particularly inhibit glycosylation in a reporter sequence derived from human von Willebrand factor (Nehre, 1996). Proline residues are often close to an O-glycosylated Ser(Thr) residue, and regions of peptides that contain a high proportion of serine, threonine, and proline residues are frequently O-glycosylated.

Inspection of NMR solution structures or X-ray crystallography data may allow possible glycosylation sites to be predicted, but the success may depend on the accuracy of the structure. For example, O-linked sugars have been identified on human erythrocyte CD59, but not yet located. There are eight potential O-glycosylation sites located at Thr10, 15, 29, 51, 52, and 60 and at Ser 20 and 21 (Figure 15). Two NMR solution structures are available for CD59 (Fletcher et al., 1994; Kieffer et al., 1994). Rotation and inspection of both structures indicates that Ser20, which forms part of the N-glycosylation sequon, may not be readily accessible once the N-linked sugar is attached at Asn18. However, the side chain of the adjacent residue, Ser21, presents a feasible site for O-glycosylation. Thr15 is also close to Asn18, but is accessible, as are the side chains of Thr52 and Thr60. Thr29 is inaccessible. In the structure determined by Fletcher et al. (1994), the side chain of Thr51 is also accessible. Interestingly, in the structure determined by Kieffer et al. (1994), a small difference in the conformation of the helix to which Thr51 is attached causes the side chain to be differently oriented so that, in this case, it is inaccessible. Thr10 is partially hindered in both models. As a result of these observations, it may be concluded

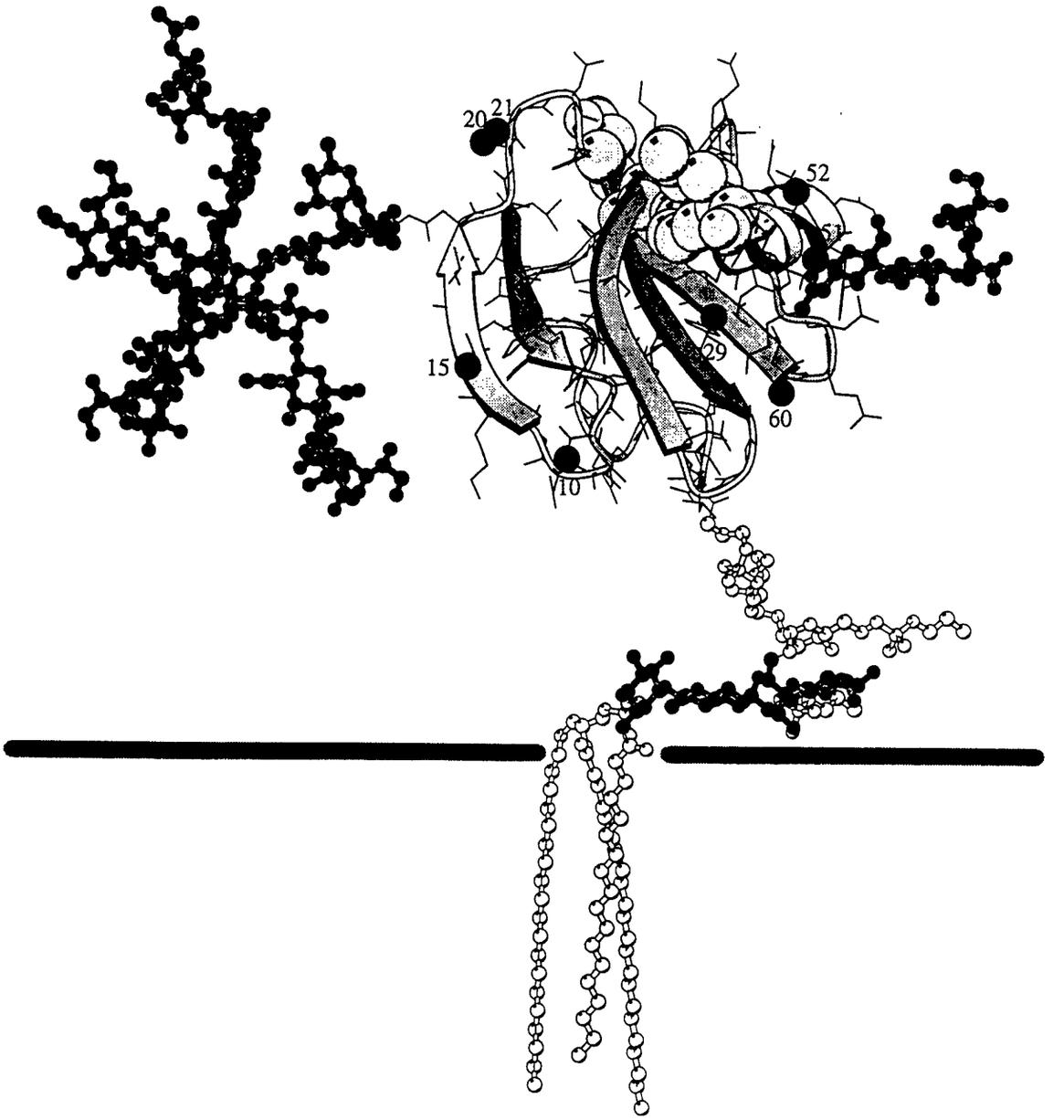


FIGURE 15. Molecular model of CD59 based on the protein coordinates from the solution structure (Fletcher et al., 1996). The binding site residues (modeled with Van der Waals surfaces in gray cpk notation) are located at Trp40, Asp24, Arg53, and Glu56 (Bodian et al., submitted). The glycan anchor is 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, tetraantennary complex N-glycan is shown attached to Asn18. The O-glycan, NeuNAc2,3Galβ1-3GalNAc, is attached to Thr51 to indicate one of the possible linkage positions. Thr and Ser hydroxyl side chains in the protein are numbered. The sugar chains and the glycan contained within the GPI anchor are depicted in black 'ball and stick' notation.

that potential sites at Ser20, Thr10, and Thr29 are unlikely to be O-glycosylated.

While most O-linked sugars contain GalNAc linked to Ser or Thr, t-PA contains an O-linked fucose residue α- linked to Thr61 (Figure 3B) in the growth factor domain (Harris et al., 1991). Thr61 is part of a consensus sequence for O-fucosylation (GlyGlySer/Thr) that has been noted in epidermal growth factor regions in other molecules associated with the blood coagulation cascade such as factors VII, IX, XII (Nishimura et al., 1992; Harris and Spellman, 1993).

Other amino acids carrying O-linked sugars include hydroxylysine linked to galactose (in collagens), hydroxyproline linked to arabinose (plant glycoproteins), Ser linked to GlcNAc (nuclear proteins), Tyr linked to glucose (glycogenin), Tyr linked to galactose (eubacteria), Ser linked to α-galactose (plants and celluloytic bacteria), Ser linked to β-glucose (bovine blood-clotting factor) (Sharon and Lis, in press).

2. The 3D Structure of the Protein O-Glycosylation and Glycosaminoglycan Assembly

O-linked glycans are added to the protein posttranslationally, initially by the addition of a single GalNAc residue in an α-linkage to the hydroxyl group of a Ser or Thr side chain in the fully folded protein. The 3D structure of the protein plays a major role in determining which sites will be occupied and how the sugars will be processed. Clearly, serine and threonine residues that are not accessible to GalNAc transferase will not be glycosylated. O-glycosylation is often associated with regions of peptide that contain many proline residues and as a result assume a helical structure that may make the Ser and Thr side chains more accessible. For example, mucins contain extended stretches of peptide sequence with repeating serine, threonine, and proline residues and are heavily O-glycosylated (Shogren, 1989; Gerken, 1989). Gelatinase B is the largest member of a family of structurally related matrix metalloproteinases. Human and mouse gelatinase B, which is also heavily O-glycosylated, contains seven protein domains, three potential N-glycosylation sites, many Ser and Thr residues, and a section of primary sequence rich in Pro, Ser, and Thr within a collagen type V domain (Figure 16) (Masure et al., in press).

Another example of an O-glycosylated proline-rich domain is the hinge region of IgA1 that is between the Fab and Fc regions (Figures 17A and B). This section of the peptide (Fab-P-V_{222}-P-S-T-P-P-T-P-S-P-S-T-P-P-T-P-S-P-S_{240}Cys-Fc) contains five glycosylation sites at Ser224, 230, 232, 238, and 240. The sites predominantly contain sialylated GalGalNAc sugars (Field et al., 1994). O-glycans may have both general and specific functions. In some cases, repulsion between the charged sugars in regions containing O-linked sugars may serve to extend the peptide chain (Gerken et al., 1989). The hinge region of IgA1 is resistant to many common proteases (Mestecky and Kilian, 1985). The energy minimized structure of the hinge region of IgA1, shown in Figure 17B, suggests that the conformation of the sugars may allow them to protect the hinge region from cleavage. Interestingly, a limited number of highly specific proteases that cleave IgA1 at its extended hinge region are secreted by a number of pathogenic bacteria, such as those that cause meningitis (*Neisseria meningitidis*), respiratory tract infections (*Haemophilus influenzae and Streptococcus pneumonia*), gonorrhea (*Neisseria gonorrhea*), conjunctivitis (*Haemophilus aegyptius*), dental plaque formation (*Streptococcus sanguis*) (Plaut, 1983; Killian and Reinholdt, 1986; Mulks and Shoberg,

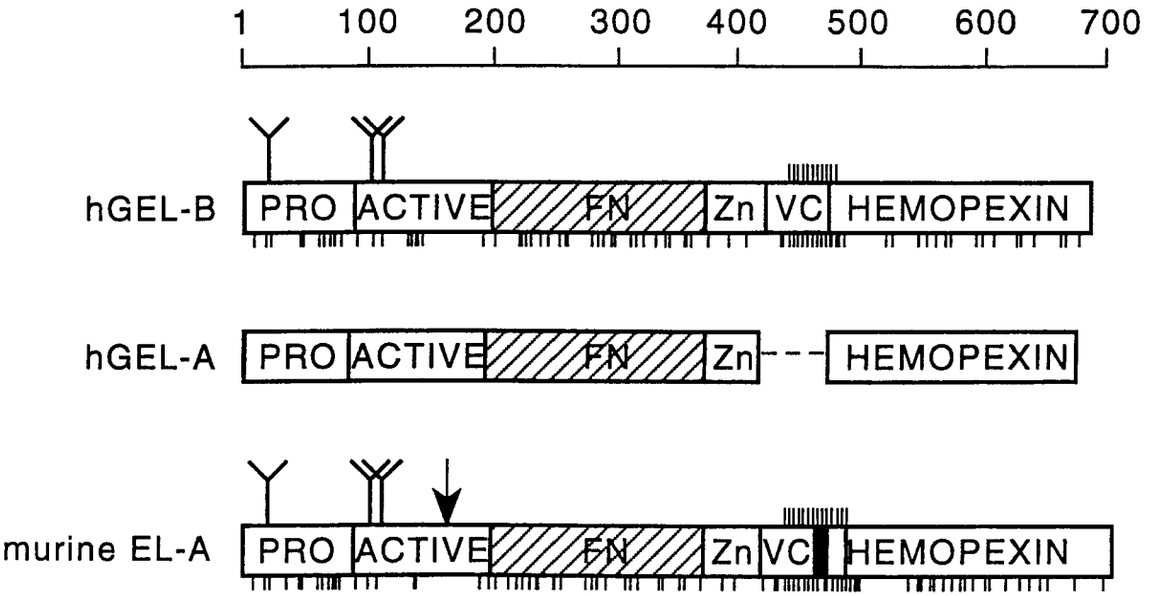


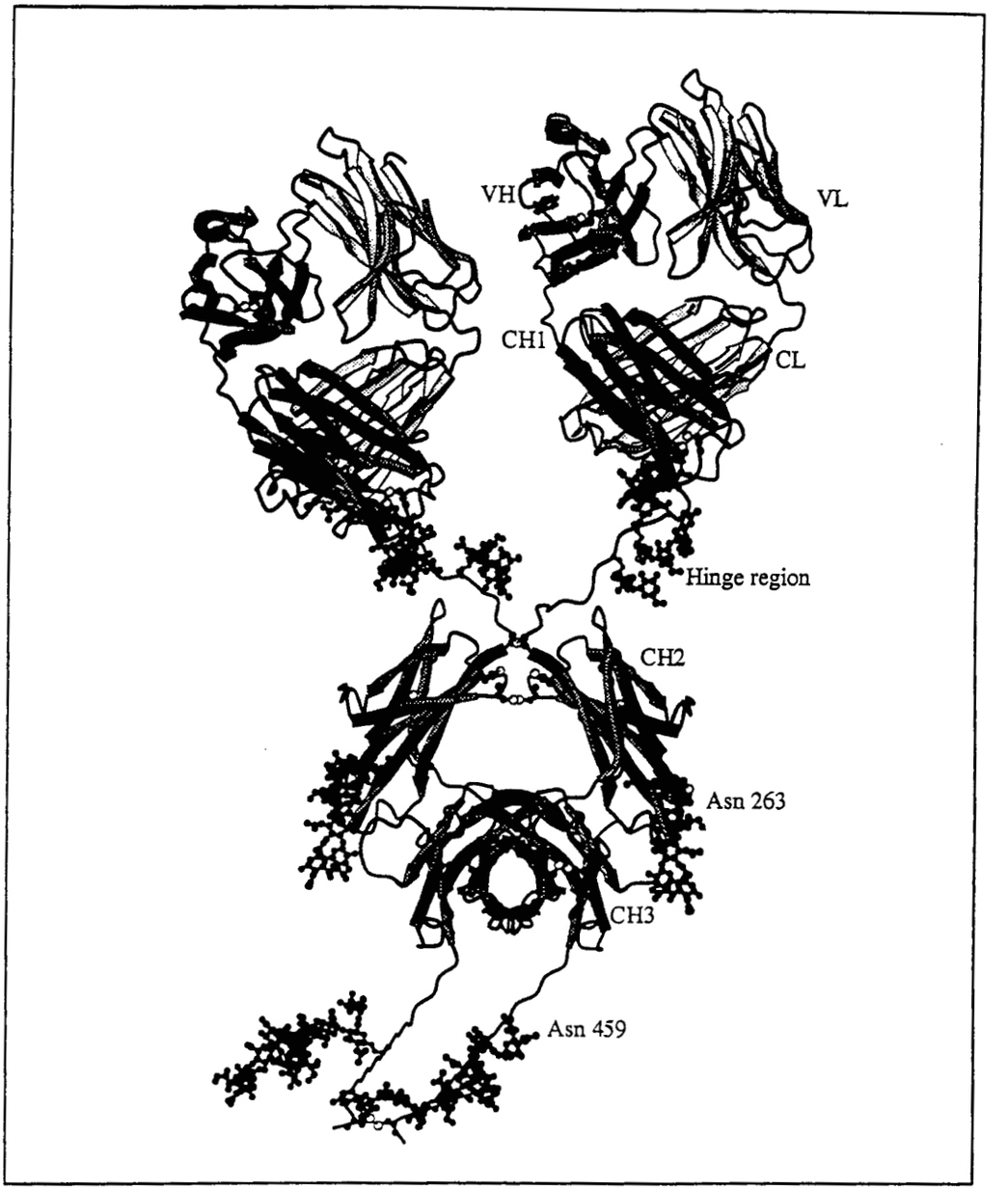
FIGURE 16. Diagram of domains in the proenzyme human gelatinase B (hGEL-B). The scale at the top shows the number of amino acids in the protein relative to the domains in hGEL-B. The enzyme consists of seven domains (Masure et al., 1993), some of which are homologous with previously characterized domains in other proteins (e.g., FN, fibronectin; Zn, zinc-binding domain; VC, collagen type V domain and hemopexin. N-linked glycosylation sites are denoted by the symbol Y; serine and threonine residues, which are potential sites for O-glycosylation, are denoted by the symbol I. Human gelatinase B contains a section of primary sequence (VC) rich in proline, serine, and threonine, which is not conserved in human gelatinase A (Masure et al., 1997).

1994). IgD also contains an O-glycosylated hinge region with glycosylation sites at Ser224, 230, 232, 238, and 240. *In vitro*, both IgA1 and IgD bind to a lectin site on T lymphocytes through the O-glycans in the hinge region (Rudd et al., 1994b).

The core protein sequence is important in the assembly of glycosaminoglycans, which begins with the addition of xylose to Ser (Esko and Zhang, 1996). The attachment sites are exposed on the protein surface and frequently located on β-turns. The immediate region around the sites is generally rich in Asp, Glu, Gly, Ala, Ser. Thr, Phe, Val, and Leu and clustered sites also exist in which Ser-Gly/Ala repeats occur within close proximity or immediately adjacent to each other. While synthesis of heparan sulfate chains requires a positive signal from the aglycone, synthesis of chondroitin chains occurs by default.

3. The Enzyme Repertoire in the Cell and the Synthesis of O-Glycans

The structure of O-glycans also reflects the levels and types of enzymes in the cell. For example, the Tn antigen (GalNAcSer/Thr)$_3$ (Nakada et al., 1993) is present on a subset of glycoforms of glycophorin, a pro-



A

FIGURE 17. **(A)** Molecular model of an IgA1 glycoform. This model was constructed from the mouse IgA Fab crystal structure, which was first altered to comply with both the primary amino acid sequence and disulfide bond pattern of human IgA1 Fab and then energy minimized and attached to the IgA1 Fc model (Mattu, T. S., Pleass, R. J., Woof, J. M., Rudd, P. M., and Dwek, R. A., submitted). The model highlights the orientation of the Fab arms that, as a result of the different pattern of disulfide bonding, occupy space that is inaccessible to the Fab region of IgG. **(B)** Molecular model of a single glycoform of the O-glycosylated hinge region of IgA1. The glycans were added as intact, energy-minimized units, and energy conflicts were reduced by altering side chain and glycosidic torsion angles. The four attached glycans are those identified by Baenziger and Kornfeld (1974): GalNAc, NeuNAc(3)Galβ1-3[NeuNAc(3)]GalNAc and two molecules of NeuNAc(3)Galβ1-3GalNAc. Thr 236 has been left unglycosylated because this site is unoccupied in the majority of human serum IgA1 molecules (Mattu, T. S., unpublished data).

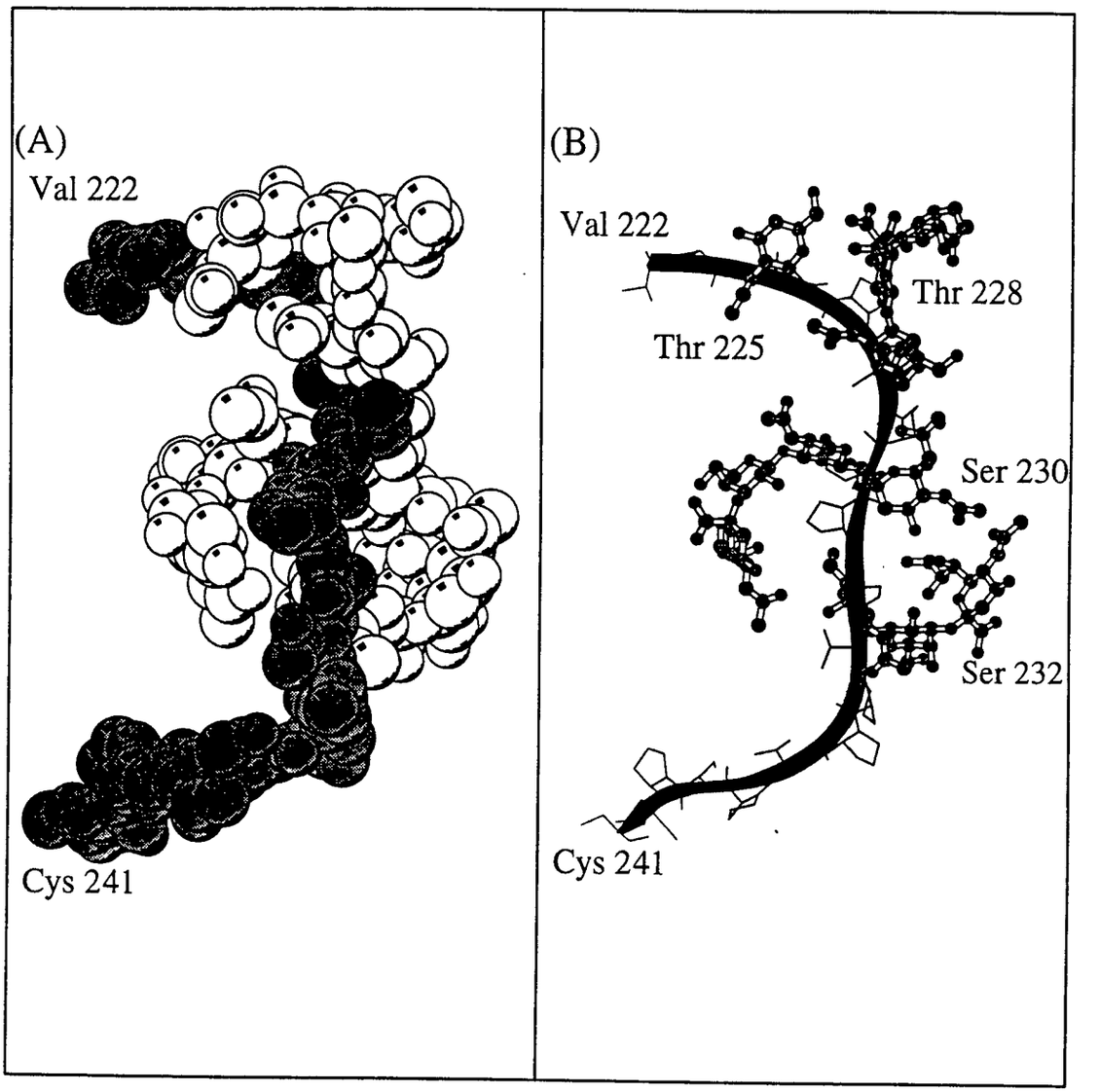


FIGURE 17B

tein present on the surface of erythrocytes. This subset originates from progenitor cells that lack β-galactosyl transferase (Cartron et al., 1978), thereby preventing further processing beyond the initial GalNAc monosaccharide.

The differences between the antigens in the ABO blood group system are controlled primarily by the levels and types of enzymes in the cell. The A, B, and H antigens are glycan motifs that differ only in the sugar residues at the nonreducing ends of the glycans. The H antigen occurs in type O individuals; it is also the precursor oligosaccharide of A and B antigens. Type A individuals express a transferase that specifically adds an *N*-acetylgalactosamine residue to the terminal position of the H antigen; in type B individuals the residue added to the terminal position is galactose.

D. Heterogeneity in GPI Anchor Glycans

1. Heterogeneity in GPI Anchor Glycans Develops in the Golgi and is Cell and Protein Specific

All GPI anchors analyzed to date contain a common glycan core (Manα1-2Manα1-6Manα1-4GlcNH). This may be further processed in a cell-type and protein-specific manner. For example, the rat brain Thy-1 anchor contains additional Manα1-2 and GalNAcβ1-4 residues (Figure 18), whereas the tetrasaccharide backbone of the anchor of the variant surface glycoprotein from *Typanosome brucei* is substituted with branched side chains of α-gal residues (Figure 18). Compositional analysis of the *T. brucei* anchor glycans also revealed the presence of a mannose-6-phosphate residue (Ferguson et al., 1988). In Thy-1, it is not known when the addition of mannose residues to the tetrasaccharide core takes place. It may be that this is a modification to the anchor precursor, or that it is the result of the action of a currently unidentified mannosyl transferase located in the ER or the Golgi.

To date, the only structural information available for two anchored proteins expressed in the same cell line is from studies of acetylcholinesterase and CD59, both from human erythrocytes (Deeg et al., 1992, Rudd et al., submitted). The majority of the GPI anchors of CD59 (Figure 18) were identical, in terms of PI and glycan structure, to that described for acetylcholinesterase. The identical PI moieties in particular, which contain alkyl chains and C22:4 fatty acid components that are not common in mammalian PI pools (Kerwin et al., 1994), suggest that, in reticulocytes, the two proteins receive the same GPI precursor in exchange for their different COOH– terminal GPI signal peptides. The small proportion of CD59 anchors subsequently modified by the addition of β-GalNAc suggests that at least some human reticulocytes possess the β-GalNAc transferase required for this relatively common GPI modification. The difference in the level of GPI modification (9% for CD59 and 0% for AChE) expressed in the same cells presumably reflects factors imposed by the glycoprotein to which the anchor is attached. More comparative studies are needed to explore the measure of control that the individual protein exerts over the glycosylation of its own anchor.

IV. GLYCAN HETEROGENEITY AND PROTEIN STRUCTURE

A. Glycosylation and Immunoglobulins

1. Cell Type-Specific Glycosylation of IgG

An important example of cell-specific glycosylation involves IgG. The glycosylation of IgG is both species specific (Rademacher et al., 1985) and cell specific (Lund et al., 1993), and while the structures of the oligosaccharides (Figure 19A) remain unchanged, the relative proportions of the glycoform populations are altered in pregnancy and some diseases, notably rheumatoid arthritis (RA) (Parekh et al., 1995) (Figures 19 B,C).

Interestingly, in a study of the glycosylation of serum IgG from patients with a range of autoimmune diseases, changes in the relative proportions of the oligosaccharide structures were not confined to a single glycoform structure. Each of six rheumatic diseases showed a statistically different pat-

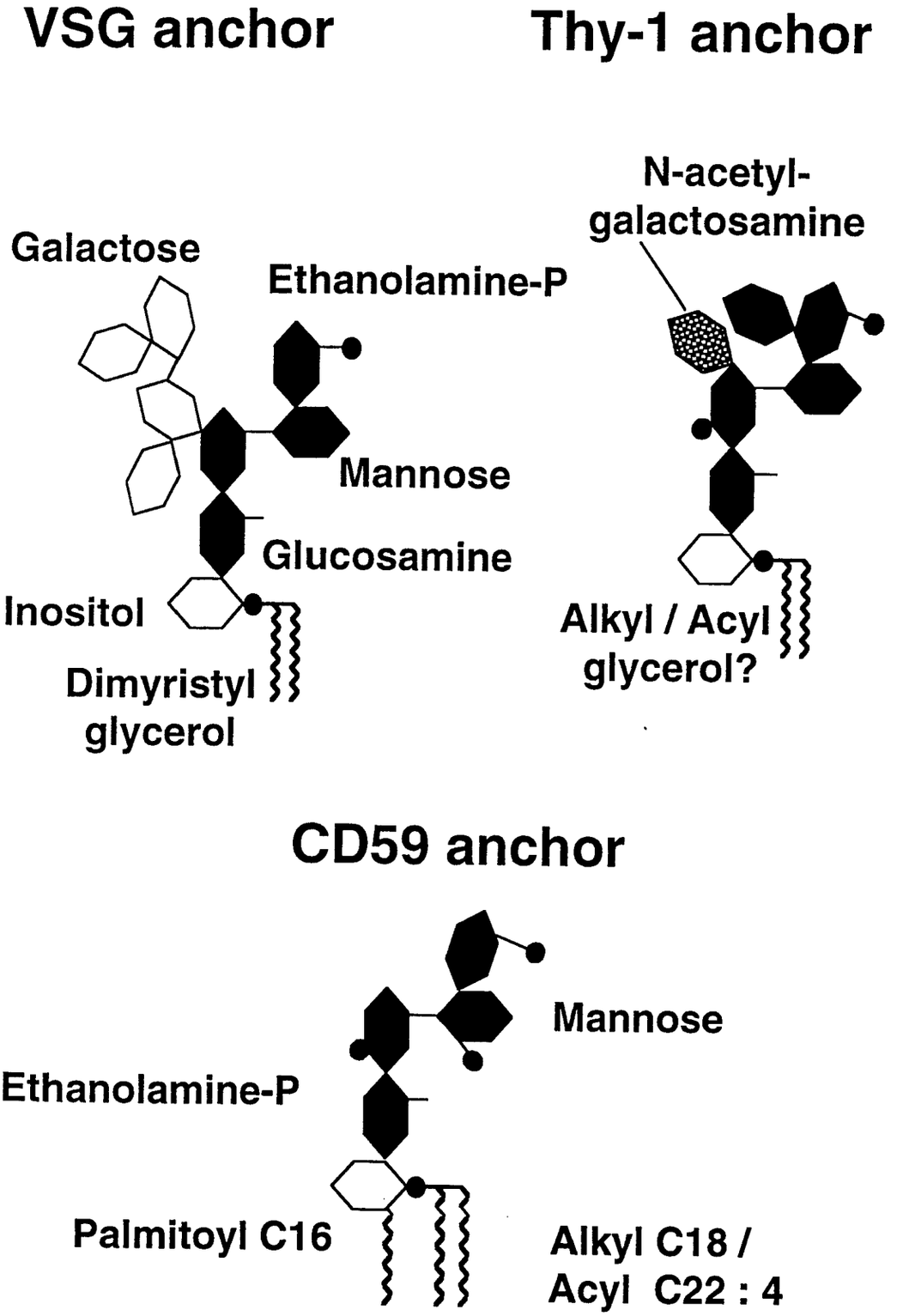
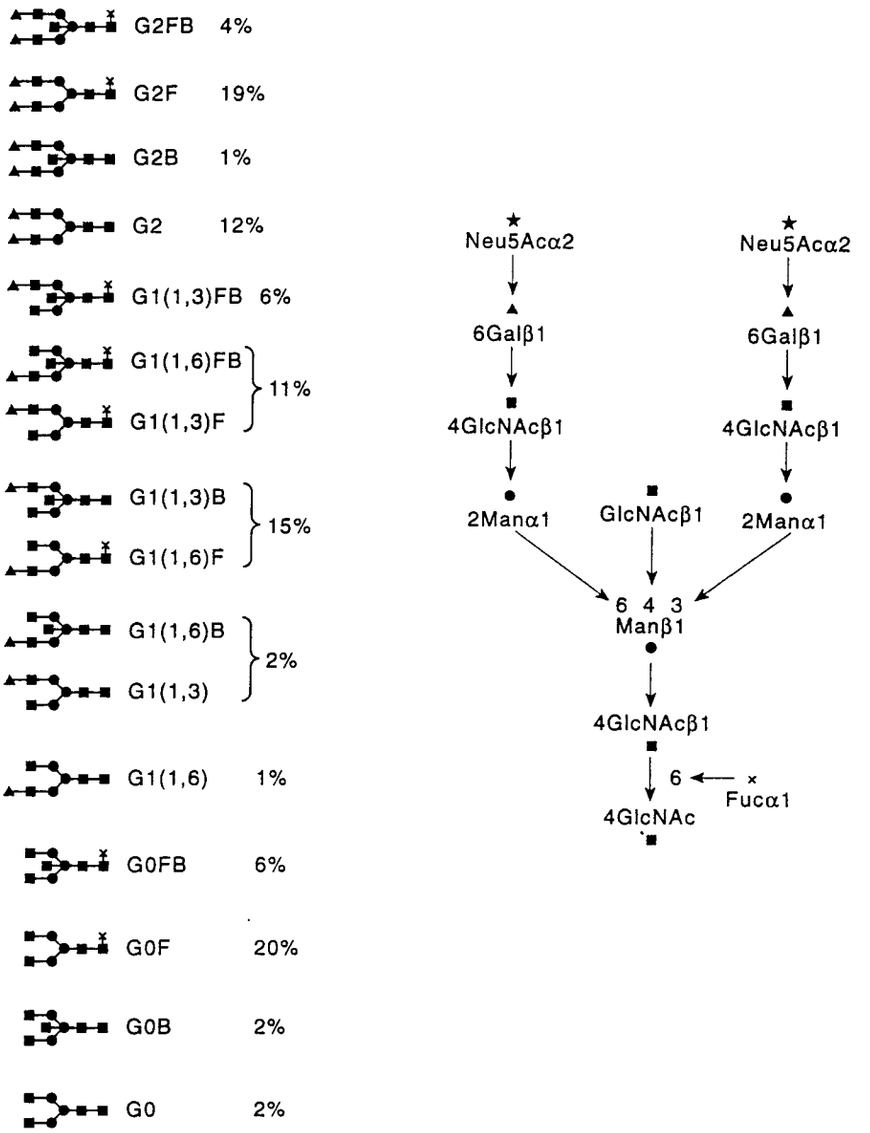


FIGURE 18. Comparison of the structures of the VSG, Thy-1 and CD59 GPI anchors.



A

FIGURE 19. **(A)** Primary sequences of the neutral N-linked oligosaccharides associated with IgG and the percentages of the different glycoforms present in the desialylated glycan pool. The most fully processed oligosaccharide present on IgG in normal human serum is the disialylated, biantennary complex sugar, shown on the right of the figure. The sugars on the left are those present in the oligosaccharide pool after desialylation with *Arthrobacter ureafaciens* sialidase. In the top four structures, both arms terminate in galactose (G2-type sugars). In the next eight structures, either the top (α-1,3) or the bottom (α-1,6) arm terminates in galactose (G1-type structures). In the bottom four structures, both arms lack galactose and terminate in *N*-acetylglucosamine (G0-type structures). **(B,C)** Normal phase HPLC separation of total IgG glycans. The glycans were released by hydrazinolysis from (**A**) pooled normal serum IgG and (**B**) rheumatoid serum IgG from a single patient. The sugars were fluorescently labeled and resolved by normal-phase HPLC. Peaks were assigned glucose unit values by comparison with the elution positions of a standard dextran ladder (shown at the top of the figure). The peaks were assigned structures from their gu values, exoglycosidase digestions, and by coinjection of standard glycans (Guile et al., 1996). The percentage of biantennary sugars in which both arms lack galactose and terminate in GlcNAc (G0 type), was 51.8% in RA IgG compared with 15.2% in normal serum IgG.

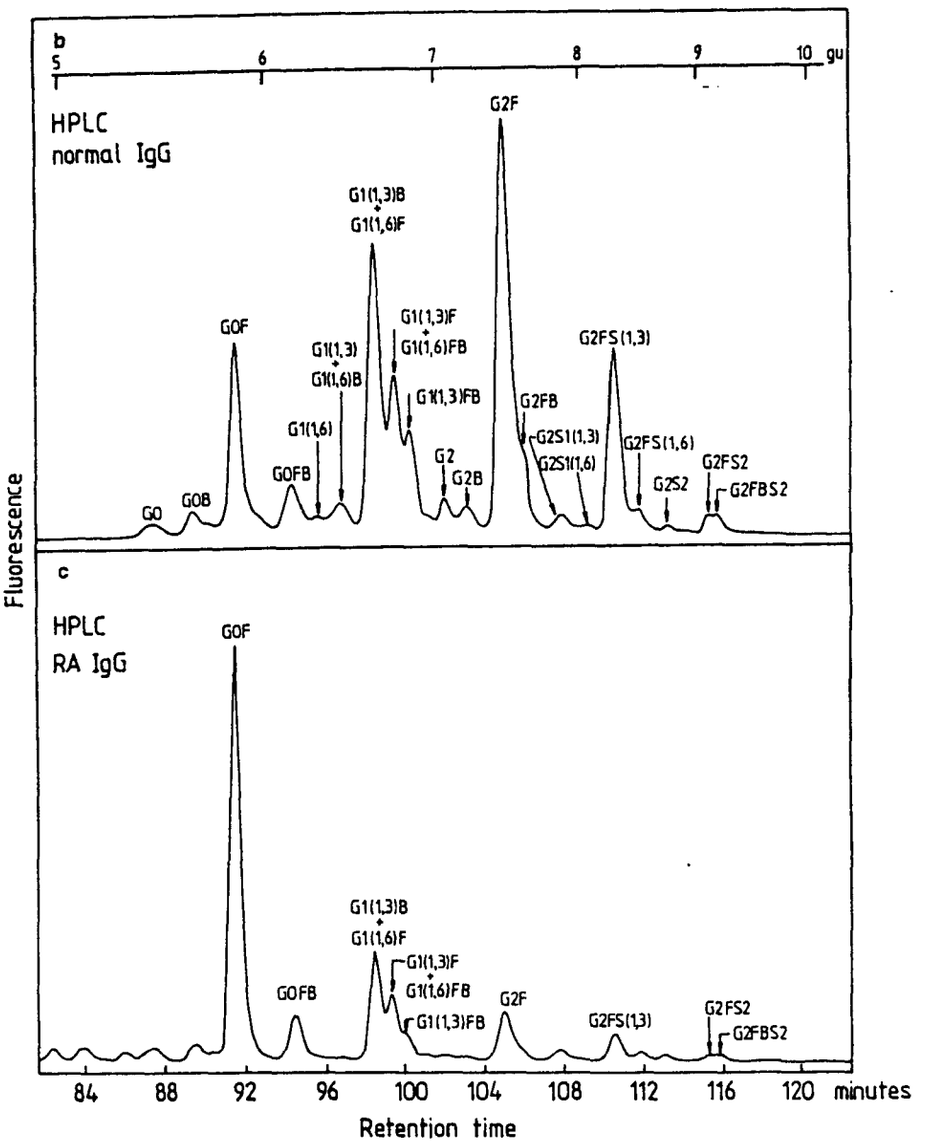


FIGURE 19B and C

tern of glycosylation (Figure 20) (Watson et al., in preparation). This suggests that in each disease different sets of IgG-producing cells are involved and that these contain different levels or activities of the glycosylating enzymes. While serum IgG is polyclonal, in these diseases the increases in the levels of particular sets of glycoforms may reflect the expansion of one or more clones of IgG-producing cells.

2. Glycan Attachment to Immunoglobulins

The immunoglobulins (Igs) IgG, IgA, IgM, IgD, and IgE are a family of structurally related proteins involved in antibody recognition. Igs consist of assemblies of domains, of which the basic structural unit is the immunoglobulin fold (Figure 21). The immunoglobulin fold is also the basic struc-

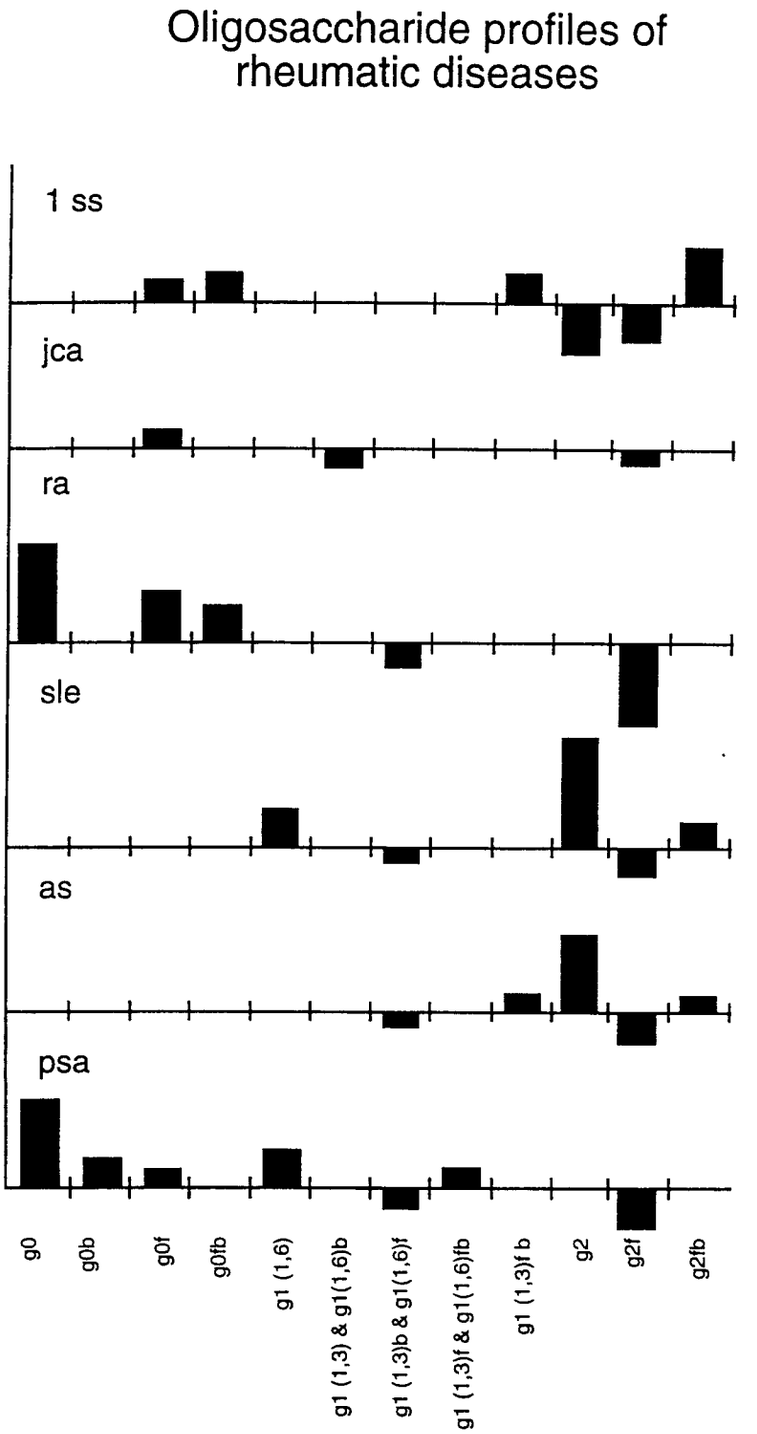


FIGURE 20. Disease-specific glycosylation patterns of serum IgG oligosaccharides from patients with a range of rheumatic diseases. The midline in the histograms represents the levels of each sugar in normal IgG. The histograms indicate the relative increase or decrease (above or below the line, respectively) of each sugar in each disease state. The patterns are characteristic and statistically significant (n = 5 to 10) (Watson, M., Rudd, P. M., Bland, M., Dwek, R. A., and Axford, J. S., in preparation). 1ss, primary Sjogren's syndrome; jca, juvenile chronic arthritis; ra, rheumatoid arthritis; sle, systemic lupus erythrematosus; as, ankylosing spondylitis; psa, psoriatic arthropathy.

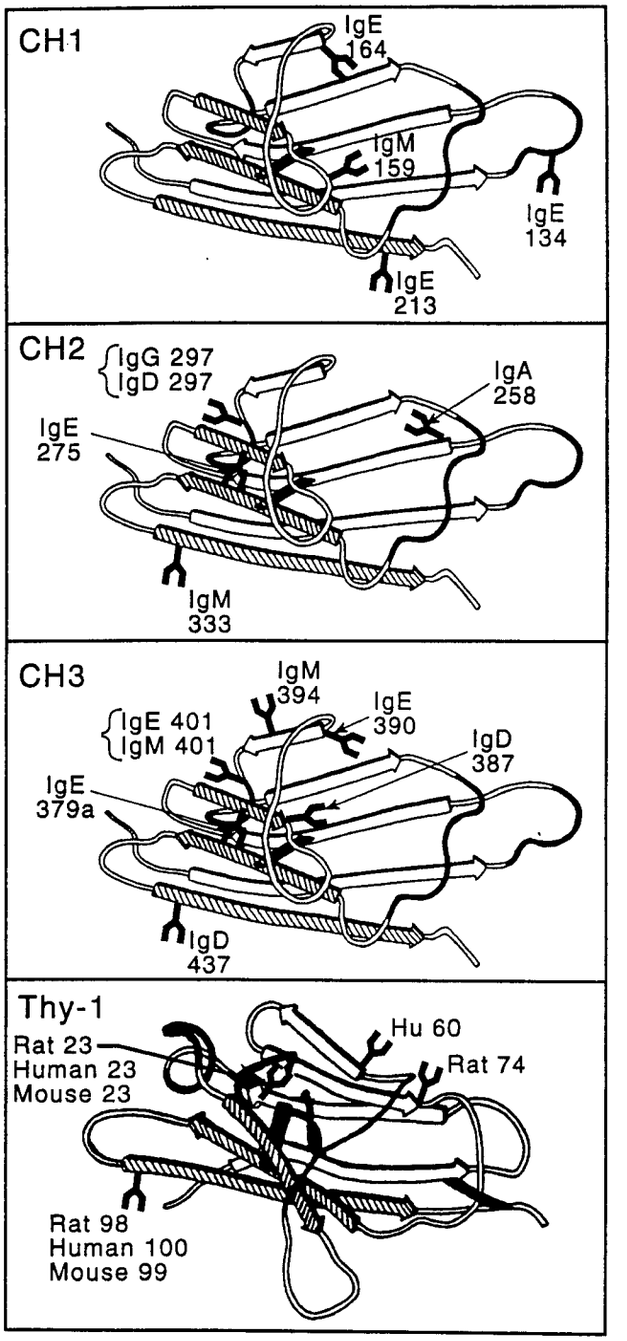


FIGURE 21. Location of the N-glycosylation sites of IgA, IgD, IgE, IgM, and IgG on the constant domains (CH1, CH2, and CH3) of the immunoglobulin fold. The top three panels contain molecular models of the C2 immunoglobulin domain on which the N-glycosylation sites for each class of immunoglobulins has been modeled. From the top, the panels show (1) the location of the N-glycosylation sites in the CH1 domain that are present in IgM and IgE, (2) the location of the N-glycosylation sites in the CH2 domain in IgA1, IgD, IgE, IgM, and IgG, and (3) the location of the N-glycosylation sites in the CH3 domain present in IgD, IgE, and IgM. The bottom panel contains a molecular model of an immunoglobulin V domain on which the human and rat glycosylation sites present in Thy-1 have been located.

TABLE 1
The Location of N-Linked Glycosylation Sites on the CH1, CH2, and CH3 Domains of Immunoglobulins

		IgG (human serum)	IgA (human serum/chimera)	IgM (murine)	IgD human myeloma	IgE
CH1	a[a]					Asn134AlaThr
	e					Asn164GlyThr
	d			Asn158aAsxSer biantennary		
	i					Asn213 LysThr
CH2	c					Asn265 IleThr
	b		Asn258 LeuThr bi- and triantennary (human serum)			
	g	Asn297 SerThr biantennary			Asn297GlySer oligomannose	
	h			Asn333AlaSer triantennary		
CH3	c					Asn379a LeuSer
	c/d				Asn387ThrSer biantennary	
	e					Asn390HisSer (not glycosylated)
	f			Asn394IleSer triantennary		
	g			Asn402AlaThr triantennary	Asn394 (401) GlyThr	
	h				Asn437AlaSer biantennary	

TABLE 1 (continued)
The Location of N-Linked Glycosylation Sites on the CH1, CH2, and CH3 Domains of Immunoglobulins

	IgG (human serum)	IgA (human serum/chimera)	IgM (murine)	IgD human myeloma	IgE
Tailpiece j		Asn459 ValSer bi- and triantennary (chimera)	Asn563ValSer oligomannose		

[a] a, loop; b, middle strand of β-sheet, toward N terminus; c, middle of β-strand; c/d, middle of loop; d, N-terminus of loop; e, end of loop entering N terminal of β-strand; f, middle of β-strand; g, middle of loop; h, middle of β-sheet; i, C-terminal end of β-strand.

Note: Where information is available, the class(es) of the major oligosaccharides present at each site are given.

tural component of the large group of proteins that make up the immunoglobulin superfamily (IgSF). These proteins, which also function in the immune system, are mainly involved in cell-cell recognition (Williams, 1987; Williams and Barclay, 1988). The proteins in both the Ig and IgSF are built of domains of approximately 100 amino acid residues. The tertiary structure is formed by two β-pleated sheets packed face to face and stabilized by one or more intrachain disulfide bonds. The domains are grouped into three sets that are structurally distinct: variable (V), constant (C)1, and C2. Immunoglobulins (Igs) contain V- and C1-type domains, while the Ig-related domains of non-Ig molecules are described as belonging to V or C sets, depending on whether they have a pattern of strands approximating a V or C domain. The glycosylation sites on these structurally similar domains are located in different positions in different molecules (Table 1, Figure 21).

Immunoglobulins are constructed in the ER from assemblies of heavy and light chains (Bole et al., 1986), which are themselves composed of V and C domains. The complex sugars on the proteins are therefore processed on intact Ig molecules. In some cases, intact Ig molecules assemble further, and the processing takes place within a multimer. For example, IgA and IgM form dimers and pentamers, respectively. In general, therefore, the glycosylation of immunoglobulins reflects the quaternary and tertiary structure of the assembled molecules. In contrast, the members of the IgSF (Section IV.B) are monomers composed of several Ig-like domains linked together in a linear chain, between which there are relatively few interactions.

Overall, it does not appear to be an intrinsic functional requirement of Ig domains to contain a particular sugar at a particular site, although the sugar sites are highly conserved in all Igs of the same class. However, as in the case of IgG, which has been studied extensively, the sugars play a critical role in the structure and function of the protein. Sections IV.A.3 to IV.A.12 explore the role of the protein structure in determining the glycosylation of Igs and the role of the sugars which maintain the structure and some of the functions of the molecules.

3. The Structure of Immunoglobulin G

IgG (mol wt 150 to 160 kDa) consists of two heavy (H) and two light (L) chains linked together by disulfide bridges and divided into homologous regions of sequence (V_H, C_H1, C_H2, C_H3, V_L, and C_L) (Figure 22E) known as the immunoglobulin fold (Amzel and Poljak, 1979). V_H and V_L contain hypervariable regions of sequence that together form the antigen-binding site. In humans, there are four subclasses, distinguishable from each other by the arrangement of disulfide bridges in the hinge region of the molecule. IgG can be cleaved by papain into Fab (the antigen-binding fragment) and Fc (the region responsible for the effector functions of the molecule). In IgG, the two CH2 domains each contain a conserved glycosylation site at Asn297SerThr to which complex biantennary oligosaccharides are covalently linked. On average, 2.5 ± 0.3 N-linked sugars are attached to each IgG molecule (Chang, S. C., 1993), of which two are located in the Fc. The additional oligosaccharides have been located to the hypervariable regions in the Fab. IgG is associated with at least 16 different complex neutral glycans (Figure 19A), many of which may also be sialylated. In normal human serum IgG, the V domains in the Fab region contain mainly core-fucosylated, sialylated biantennary glycans of which approximately 30% contain bisecting *N*-acetyl glucosamine residues. The Fc re-

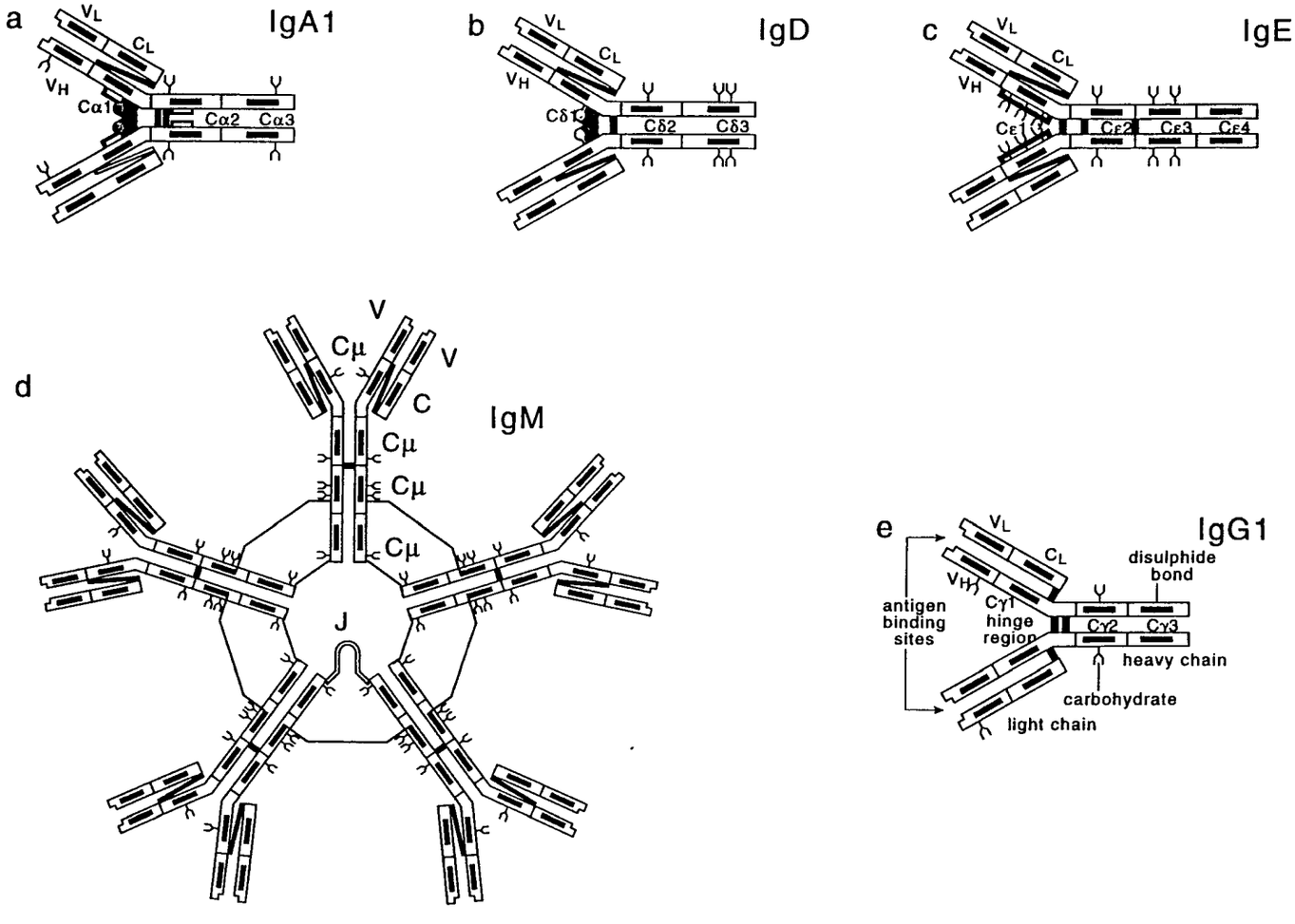


FIGURE 22. Schematic diagrams of examples of each of the five classes of immunoglobulins (IgA1, IgD, IgE, IgM, and IgG1) showing the N- and O-glycosylation sites and the disulfide bonds. (Adapted from Roitt, I. M., Brostoff, J., and Male, D. 1985. *Immunology,* chap. 5. New York: Gower.)

gion contains fewer sialylated and bisected glycans (Figure 23). The glycoforms of IgG can be grouped into three sets, depending on whether they contain 0, 1, or 2 galactose residues in the outer arms of their complex biantennary glycans. These sets of oligosaccharides are commonly known as G0, G1, and G2, respectively. The G1 and G2 type structures may be sialylated. Within each of these sets are four species that result from the presence or absence of core fucose and 'bisecting' GlcNAc (Figure 19A). The structures and relative proportions of the 16 major glycans present in the desialylated pool of IgG sugars are shown in Figure 19A.

IgG is assembled completely in the ER. Two H and two L chains associate to form IgG prior to transfer to the Golgi apparatus (Bole et al., 1986). Processing to complex glycans occurs on the assembled molecule, within the restricted space that the sugars occupy between the two CH2 domains, and in the Fab region of the molecule.

4. Protein-Oligosaccharide Interactions in IgG

Interactions of the IgG Fc protein with the developing sugar stabilize the quaternary structure of the protein, play a role in glycan processing and affect the accessibility of the sugars. The crystal structure of human IgGFc (Deisenhofer, 1981) predicts a total of 38 noncovalent interactions between the hexasaccharide unit ($GlcNAc_2Man_3GlcNAc$) and the Fc face of the CH_2 domain and an additional 13 interactions for GlcNAc and 27 for Gal on the α1,6 arm and seven for the core fucose attached to the primary GlcNAc (Padlan 1991).

The hinge disulfide bond (Cys229) in IgG Fc (Figure 22E) imposes a constraint that provides a rationale for the large interstitial space between the two CH2 domains, and protein-oligosaccharide interactions play a role in maintaining the relative geometry of the CH2 domains (Rudd et al., 1991). Consistent with this is the finding that both aglycosylated and degalactosylated IgG bind less efficiently to the Fcγ receptor (Nose and Wigzell, 1983; Leatherbarrow et al., 1985; Tsuchiya et al., 1989). This is presumably because the orientation of the hinge residues involved has altered (Duncan et al., 1988). In a recent study (Lund et al., 1996), protein engineering was used to replace amino acid residues within the extensive oligosaccharide interaction site that contacts the core hexasaccharide (GlcNAc2Man3GlcNAc). The data suggest that noncovalent interactions of multiple amino acid residues of IgG with oligosaccharide residues are necessary for optimal recognition of IgG by C1q and FcγR1.

In common with IgG, IgD contains a disulfide bond at Cys229 (Figure 22B). By analogy, this may be predicted to impose a constraint on the hinge region leading to a large interstitial space between the two CH2 domains, as in IgG. In addition, the sugars in the CH2 regions of IgD are located at the same position as those in IgG (on a loop at Asn297) and, again by analogy with IgG, may be predicted to lie within the CH2 domains. The interchain disulfide bonds at Cys229 and the sugars on IgD and IgG at Asn297 may therefore perform similar roles in both immunoglobulins, but this has yet to be established.

In contrast, in IgA (Figure 22A), it has been proposed that the relative conformation of the CH2 domains is maintained by the two interchain disulfide bonds at Cys238 and Cys239 (Mattu, 1996). In this case, there is no interstitial space and the sugars are fully accessible. In IgE (Figure 22C), there are interchain disulfide bonds at Cys229 and Cys260. Although no structural details are known, by analogy with IgA, the sugars at Asn 275 on IgE might be expected to

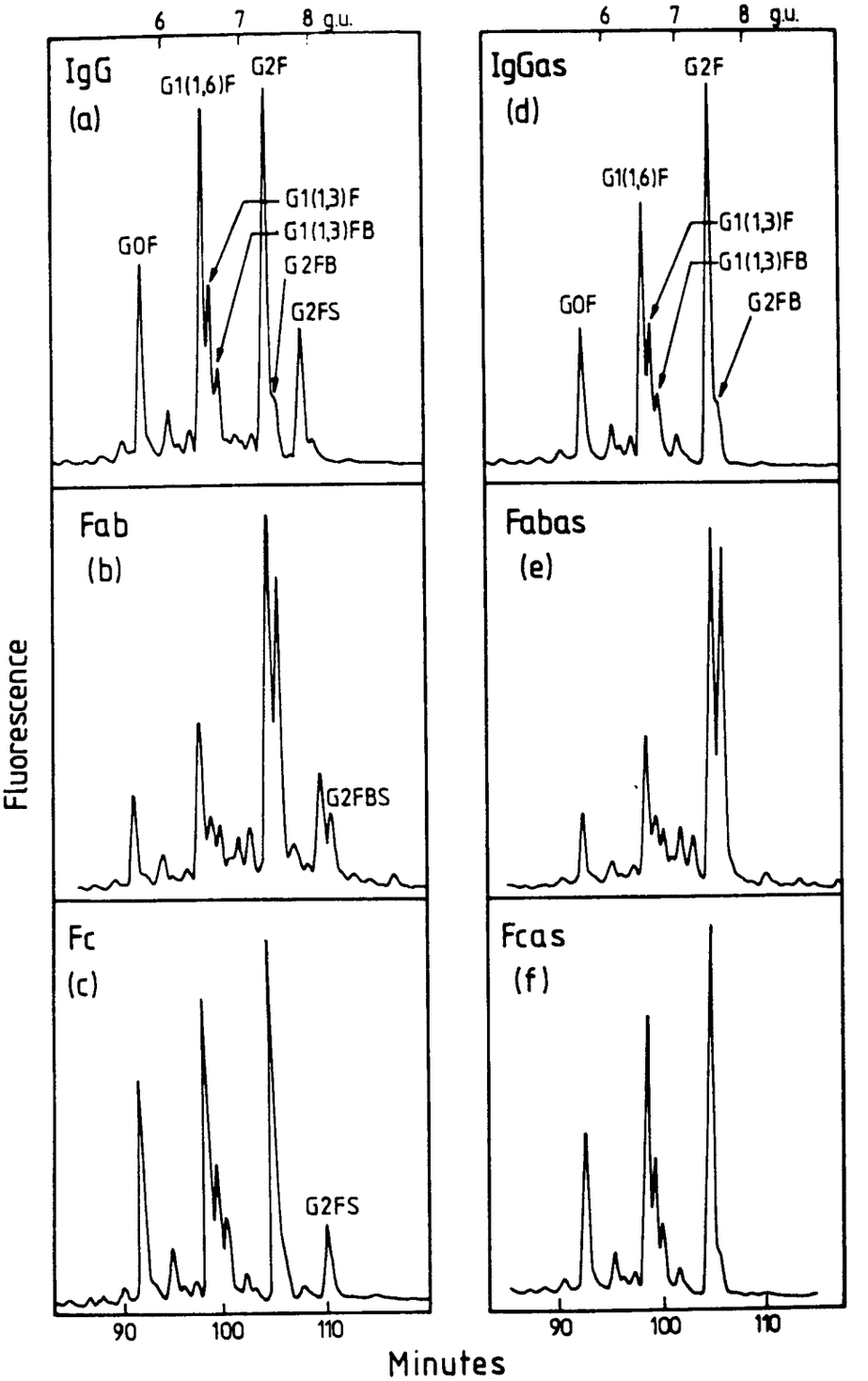


FIGURE 23. IgG oligosaccharide analysis. Panels a to c show the normal phase HPLC profiles of the fluorescently labeled oligosaccharides released from normal serum IgG, Fab, and Fc, respectively. Panels d to f contain the analyses of the desialylated glycan pools from the same samples.

point away from the surface of the protein and not be in a position to stabilize the quaternary structure of the CH2 domains. As in the case of IgA, this function may be performed by the interchain disulfide bond located at Cys260 in each domain.

5. Protein-Oligosaccharide Interactions and Processing in IgG

Protein-oligosaccharide interactions also affect the accessibility of the Fc glycans. IgG contains three sets of glycoforms (Figure 19A), which contain biantennary oligosaccharides that are associated with 0, 1, or 2 galactose residues (G0, G1, and G2 types). Nuclear magnetic resonance (NMR) relaxation studies suggest that, for G0 glycans, there is an increase in the solvent accessible surface of the protein (Figure 24) (Malhotra et al., 1995). The decreased interactions of the oligosaccharide with the protein surface, which result from the absence of galactose, lead to the displacement of the oligosaccharide from the position observed in the X-ray structure. This results in an increase in conformational freedom for the G0 oligosaccharides and exposure of the terminal GlcNAc residue. These analyses thus differentiate between two sets of glycoforms: the G2 and G1 glycoforms that are bound to, or have restricted motion relative to, the peptide surface, and the G0 glycoforms that have considerable conformational freedom independent of the protein. The exposure of the terminal β1-4GlcNAc residue on the 1,6 arm allows these glycoforms to interact with lectins, such as the *Psathyrella belutina* lectin (Tsuchiya et al., 1994) and mannose binding protein (MBP) (Figure 25) (Malhotra et al., 1995).

Interestingly, the ratio of α-1,3 to α-1,6 outer arm galactosylation was found to be 1:4 for the IgG glycans (Parekh, 1995; Wormald et al., submitted) (Figure 23). This is the reverse of the observed *in vitro* specificity (6:1) of Galβ1-4 transferase for Asn-linked oligosaccharides (Narashima et al., 1985) and suggests that *in vivo,* the 1,6 arm is more accessible to the β-galactosyl transferase than the 1,3 arm, and that the accessibility of the GlcNAc residues to the enzyme has been dramatically altered by the presence of the protein. In a recent study (Lund et al., 1996), the replacement of mainly hydrophobic amino acid residues involved in interactions with the Fc oligosaccharides with alanine resulted in increased galactosylation and sialylation of the sugars. This suggests that, in the mutants, the oligosaccharide has been released from the protein surface and is more accessible to the galactosyl and sialyl transferases than is the case with normal IgG.

6. The Location of the Glycosylation Sites on Immunoglobulins A1, M, D, E, and G

To date, only IgG and IgA (Mattu et al., submitted) are known to be glycosylated in the variable region of the Fab, and only IgA and IgD contain O-glycosylated hinge regions (see Section III.C.2). However, all immunoglobulin molecules contain N-glycosylation sites in the Fc region that are highly conserved within each immunoglobulin class. IgE and IgM are glycosylated in the CH1 domain, whereas IgG, IgA, IgD, IgE, and IgM all contain sugars in the CH2 domain. IgD, IgE, and IgM are glycosylated in the CH3 region, and IgA and IgM contain glycosylated tailpieces that are not immunoglobulin domains. Interestingly, although the C domains are structurally similar in all

FIGURE 24. Crystal structure of the IgG Fc CH2 domain comparing part of the solvent accessible surface in the G1 and G0 glycoforms. On the left, the CH2 domain contains a G1 oligosaccharide carrying a terminal galactose residue on the α-1,6 arm. The model structure shows the oligosaccharide interacting with the protein. On the right, the CH2 domain is shown with a G0 glycan. The decreased protein-oligosaccharide interactions, which result from the absence of the terminal galactose residue, allow the sugar to move within the space between the two CH2 domains. Intact IgGFc is shown in the top-left insert.

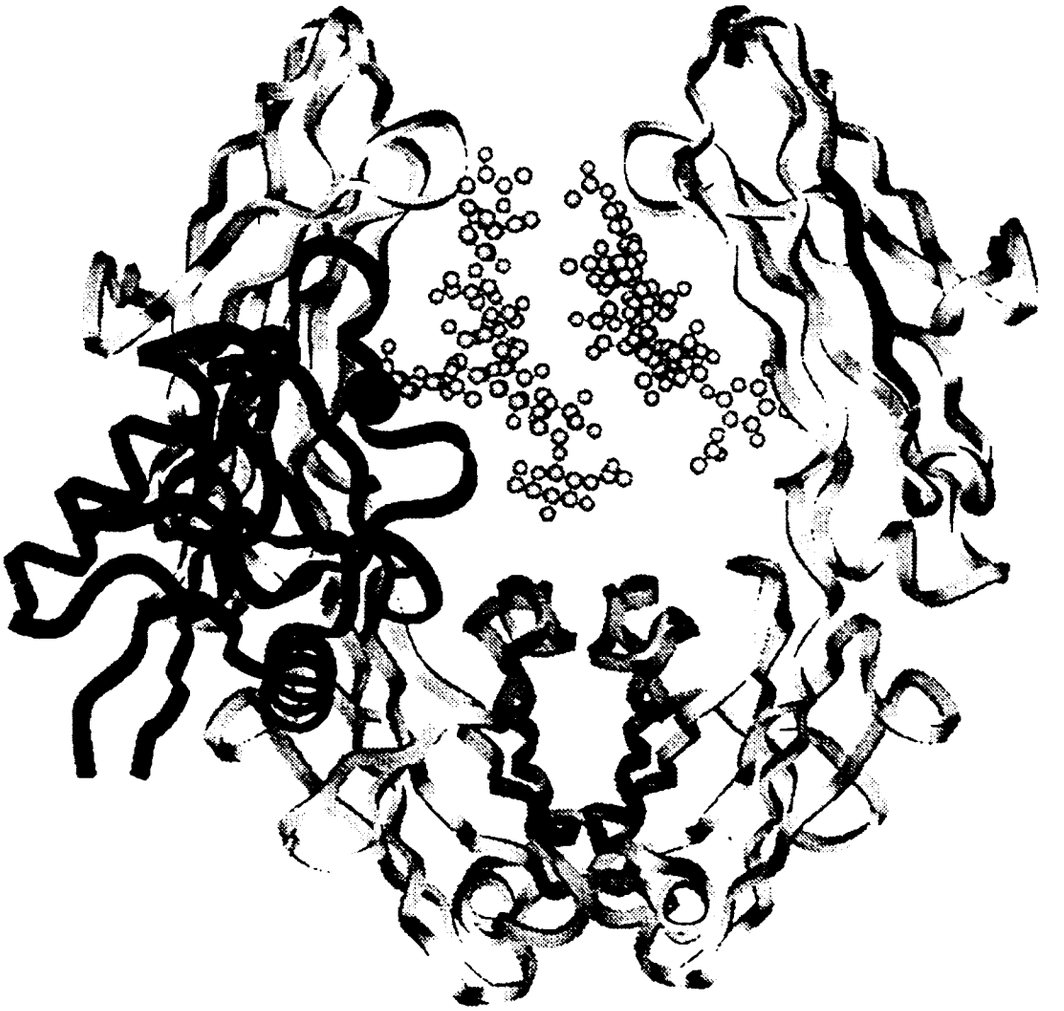


FIGURE 25. Theoretical model of the interaction between IgG Fc (light gray) and a single carbohydrate recognition domain (CRD) from MBP (black). The MBP is shown interacting with the α-1,6 arm terminal GlcNAc residue on a G0 oligosaccharide (white spheres). By an alternative displacement of the Fc oligosaccharide chain, the CRD can also interact with the α-1,3-arm nonreducing terminal GlcNAc residue.

immunoglobulins, the location of the glycosylation sites is not fully conserved between different members of the family, nor even in different domains in the same Ig molecule. The locations of the sugars on the various domains to which they are attached, the secondary structure at the sites, and the composition of the glycosylation sequons are shown in Figure 21 and in Table 1. There are a variety of numbering systems in use; this table uses the EU index numbers in which the sequences are aligned (U.S. Department of Health, 1991). Table 1 gives details of the site-specific glycosylation of these molecules. Apart from IgG and IgA, little information is available about the glycosylation of normal human serum immunoglobulins, and the site analysis of IgM was carried out on a murine IgM (Anderson et al., 1985).

7. What Determines the Glycosylation Pattern of Sugars Attached to Ig Domains?

Although the Fc regions of all the classes of immunoglobulins consist of C2-type immunoglobulin domains, the molecules are not uniformly glycosylated. In the following section we discuss some of the factors that may influence site occupancy and the final structure of the glycoform populations. Some of these factors are (1) the structure of the X amino acid in the glycosylation sequon (AsnXSer), (2) the accessibility of the amino acid side chain (i.e., is it pointing into the protein structure or away from it?), (3) the local protein secondary structure (e.g., β-strand, loop, or helix), (4) the quaternary structure of the fully assembled Ig, and (5) the accessibility of the developing glycan chain (i.e., is it sterically protected form the glycosylating enzymes by the 3D structure of the protein, and in particular is it able to take part in noncovalent interactions with the protein surface?). However, in using the available data to discuss this question, there are some caveats. In human serum, such comparisons are necessarily limited because cell-specific factors cannot be eliminated. Native serum immunoglobulins are polyclonal, and, also, they are derived from cells that have undergone proliferative and class-switching processes.

8. A Comparison of the N-Glycosylation of the CH2 Domains of IgG and IgA1

In IgA1, the N-glycosylation sites are located toward the C terminus of the domain at the C-terminal end of a β-strand on each of the two conserved Cα2 regions. As a consequence of the disulfide bonding arrangement, the amide side chains of Asn258LeuThr point away from the protein (Figure 17A). This is in contrast to IgG, where the sugars, also located in the Cγ2 region, but on a loop at Asn297SerThr, are closer to the amino terminus and are contained in the interstitial space between the two Cγ2 domains. The IgG sugars are involved in noncovalent interactions with the protein surface (Figure 24) (Padlan, 1991), which further limits their accessibility to the glycosyltransferases. In both molecules, all the sites on the Fc are fully occupied. However, after processing of the sugars in the Golgi, serum IgA1 contained 13% of triantennary glycans (Field et al., 1994), whereas IgG contained none, and 85% of IgG sugars were core fucosylated (Parekh et al., 1984) compared with only 34% of IgA1 sugars. Approximately 86% of IgA1 N-linked sugars were sialylated compared with 15% on IgG, and 24% of IgG sugars were bisected compared with 15.6 on IgA1. Interestingly, although the greater accessibility of the IgA sugars is reflected in increased sialylation and an increase in the number of antennae, the less accessible IgG sugars contain more core fucose. It is not known whether IgG-producing cells express higher levels of fucosyl transferase than IgA-producing cells or whether this finding is related to differences in the protein structures.

9. Glycosylation of Different Domains of IgM

IgM contains four glycosylation sites attached to three different Ig domains, CH1, CH2, and CH3 (Figures 21 and 22). Analysis of the total glycan pool from normal serum IgM indicated that the immunoglobulin contained mainly bi- and triantennary glycans as well as oligomannose structures

(Cahour et al., 1983). Similar sugars were identified in a site analysis of murine IgM that also contains four Ig-linked glycosylation sites (Table 1). The Cμ1 site (on a loop toward the middle of the domain) contained biantennary, sialylated glycans, whereas the Cμ2 (near the beginning of a β-sheet close to the N terminus of the domain) and Cμ3 (two sites close together on a β-sheet and a loop) both contained triantennary, sialylated glycans. Interestingly, the tailpiece, which does not have the structure of an Ig domain, contained only oligomannose sugars (Anderson et al., 1985), suggesting that in the pentameric IgM this site is protected from GlcNAc transferase I. The site-specific glycosylation of IgM suggests that the protein structure plays a role in determining the glycosylation processing, and that the position of the glycosylation site within the domain is more important than the secondary structure. This presumably also reflects the general accessibility of the sugars rather than restrictions arising from quaternary structure.

10. Glycosylation of IgD Domains

Immunoglobulin D (Figures 21 and 22) has three N-glycosylation sites (Table 1). One, in the Cδ2 domain, is located at the same position as the glycan in the CH2 region of IgG (on a loop). The other two are in the Cδ3 domain (one on a loop and the second at the beginning of a β-sheet near the C terminus and at a similar position to the IgM sugar in the CH2 domain). A site analysis of human serum IgD (Mellis and Baenziger, 1983) found that the Cδ2 domain contained unprocessed oligomannose sugars. These data are consistent with the location of this site (Asn297GlySer) between the two CH2 domains and suggest that, as is the case with IgG, the sugars may take part in noncovalent interactions with the protein. Although in IgG the site in the Cγ2 domain contains complex sugars, many are not fully galactosylated or sialylated (see Section IV.A.3). Asn387 in the Cδ3 domain contained biantennary complex glycans, some of which were sialylated and fucosylated. Asn437 contained biantennary sugars, but no fucose or sialic acid was detected, although, because Asn387 contained fucosylated structures, the enzymes were available in the cell. Interestingly, in IgM, this site contained sialylated sugars. This suggests that, in myeloma IgD, there may be lower levels of sialyl transferase and that, as a result of a competition for the enzyme, Asn387 may be preferentially processed compared with Asn437 because the sugars are relatively more accessible.

11. Glycosylation of IgE Domains

In the previous sections, we have discussed the influence of protein structure in determining the glycosylation of Ig domains. IgE offers an example in which the composition of the sequon and steric factors might be important in determining the efficiency of glycosylation.

Human IgE (Figures 21 and 22) consists of 14 domains, each with the characteristic immunoglobulin fold. Compared with the 12-domain structure of IgG, IgE has an additional pair of domains (Cε2) in the Fc region in place of the hinge in IgG. IgE contains seven potential N-glycosylation sites (Dorrington and Bennich, 1978; Baenziger et al., 1974a, b; Young et al., 1995). The only glycosylation analysis of IgE to date was carried out on the total glycan pool from a myeloma protein (Baenziger and Kornfeld, 1974a,b) and a

mixture of oligomannose and complex-type sugars were reported. No site analysis is available, but on average four carbohydrate chains per molecule were present, suggesting that there is variable occupancy in at least some of the potential glycosylation sites. In a study by Young et al. (1995), Asn275IleThr and Asn379aLeuThr were found to be partially occupied, Asn390HisSer was not glycosylated, and Asn401GlyThr was 100% glycosylated.

In general, sequons containing Leu may be poorly glycosylated (Shakin-Eshleman et al., 1996), and this may explain the partial occupancy of Asn379aLeuThr (middle of the β-strand). However, Asn 275IleThr (also on a β-strand), which is also partially occupied, is located at the same position in the CH2 domain as Asn379aLeuThr is in the CH3 domain. Because sequons containing Ile are not noted to be poorly glycosylated (Shakin-Eshleman et al., 1996), this may suggest that steric factors prevent these two sites from being fully occupied. The reason for the absence of sugars at Asn390HisSer (junction of a loop and a β-sheet) is not known. In principle, partial glycosylation at Asn379aLeuThr and full glycosylation at Asn401GlyThr could result in steric hindrance at Asn390HisSer. There is a site in a similar position in the CH1 domain of IgE at Asn164GlyThr. It would be interesting to know whether this is fully glycosylated. The fully occupied glycosylation site at Asn401GlyThr in the CH3 domain is on a loop in the same position as a glycosylated sequon, Asn401AlaThr on IgM. This indicates that in both cases this site is accessible to the dolichol phosphate precursor/oligosaccharyl transferase in the ER. In both IgG and IgD, the equivalent position in the CH2 domain contains the glycosylation site Asn297, which is fully occupied (Figure 21).

12. Role of Quaternary Structure in the Glycosylation of Igs

The quaternary structure may also be important in determining the glycosylation of Igs. This is highlighted by the differences in the glycosylation of the tailpieces of monomeric IgA1 (human) and pentameric IgM (murine). The glycosylation analysis of Fc from a chimera of mouse anti-NIP IgA1 antibody and human IgA1 in which the CH2 glycosylation site was deleted showed that the tailpiece contained both triantennary (60%) and biantennary sugars (40%) (Mattu, 1996). This suggests that the tailpiece site on monomeric IgA1 is accessible to all relevant processing enzymes. In contrast, the tailpiece of IgM (murine) contains oligomannose sugars (Man6, 45%; Man7, 10%; Man8, 45%) (Anderson et al., 1985). Because other sites on IgM contain complex-type sugars (Anderson et al., 1985), the murine IgM-producing cells must contain the necessary enzymes to develop complex glycans. Therefore, these data suggest that within the IgM pentamer, which is assembled in the ER, the glycans in the tailpiece are protected from all enzymes except the mannosidases. There may, however, be an advantage in retaining these oligomannose structures. When IgM pentamers interact with antigen, their conformation changes from a planar form to a 'table' or 'staple' form (Figure 26). In the staple form, the oligomannose sugars would become clustered together above the surface of the molecule in the center of the staple, where they may be available to take part in other events, such as activating complement. The mechanism for such an event may involve lectin recognition by, for example, the mannose binding protein. Interestingly, *in vitro*, some glycoforms of IgG that are characteristic of rheumatoid arthritis (IgG0) have

been shown to activate complement through the mannose-binding protein (Malhotra et al., 1995) (Figure 26).

B. Glycan Heterogeneity and the Immunoglobulin Superfamily

1. Protein-Specific Glycosylation within the Immunoglobulin Superfamily

In contrast to the immunoglobulins, most members of the IgSF are single-chain molecules containing Ig-like domains linked together. The structures that have been determined by X-ray crystallography or NMR show that they are linear and that the domains are unlikely to interact significantly with each other except at the domain interface.

A study of the glycosylation of these molecules (Rudd et al., in preparation) explored a number of questions relating to protein-specific glycosylation.

1. Does the glycosylation profile only reflect the tertiary structure of the domain to which the sugar is attached?
2. In proteins with more than one type of domain, does the structure of one domain influence the overall glycosylation?
3. Does the local protein structure, such as the location of glycosylation sites on sheets or loops, control the glycan processing?
4. Do some transferases compete for sugars more effectively at some glycosylation sites than others?
5. Is the transit time of soluble proteins an important feature in controlling glycosylation?
6. Does the attachment of a GPI anchor affect glycosylation?
7. When different leukocyte antigens are expressed on the same cell surface is the overall glycosylation dominated by the cell or do the differences in the structures of individual proteins ensure that a range of oligosaccharides are present?

The extent to which common features of tertiary structure control glycan processing was explored in the leukocyte antigens shown in Table 2. The antigens, which contain three different types of structural units, were all expressed in the same CHO cell line (CHO-K1) to eliminate differences that might arise from cell-specific glycosylation. They included members of the immunoglobulin superfamily, CD48, CD4, and CD2, and also CD5 and CD59, which contain scavenger receptor (ScR) domains and a Ly-6 domain, respectively (Figure 27).

The number of glycosylation sites associated with each protein ranged from one on sCD59 to four on CD48. Two of the glycans in CD2 are attached to C2 domains and one to a V domain. In CD48, three are in C domains and one is in a V domain, and in CD4, one is attached to a C and one to a V-type domain.

2. Glycan Heterogeneity and Domain Structure

CHO cells possess an efficient glycosylation machinery and contain all the enzymes present on human cells. However, the gene for α2,6 sialyltransferase in CHO cells is normally silent. The majority of the glycans associated with IgSF proteins expressed in CHO cells were of the complex biantennary core-fucosylated type, common on the C domains of human IgG and IgA1.

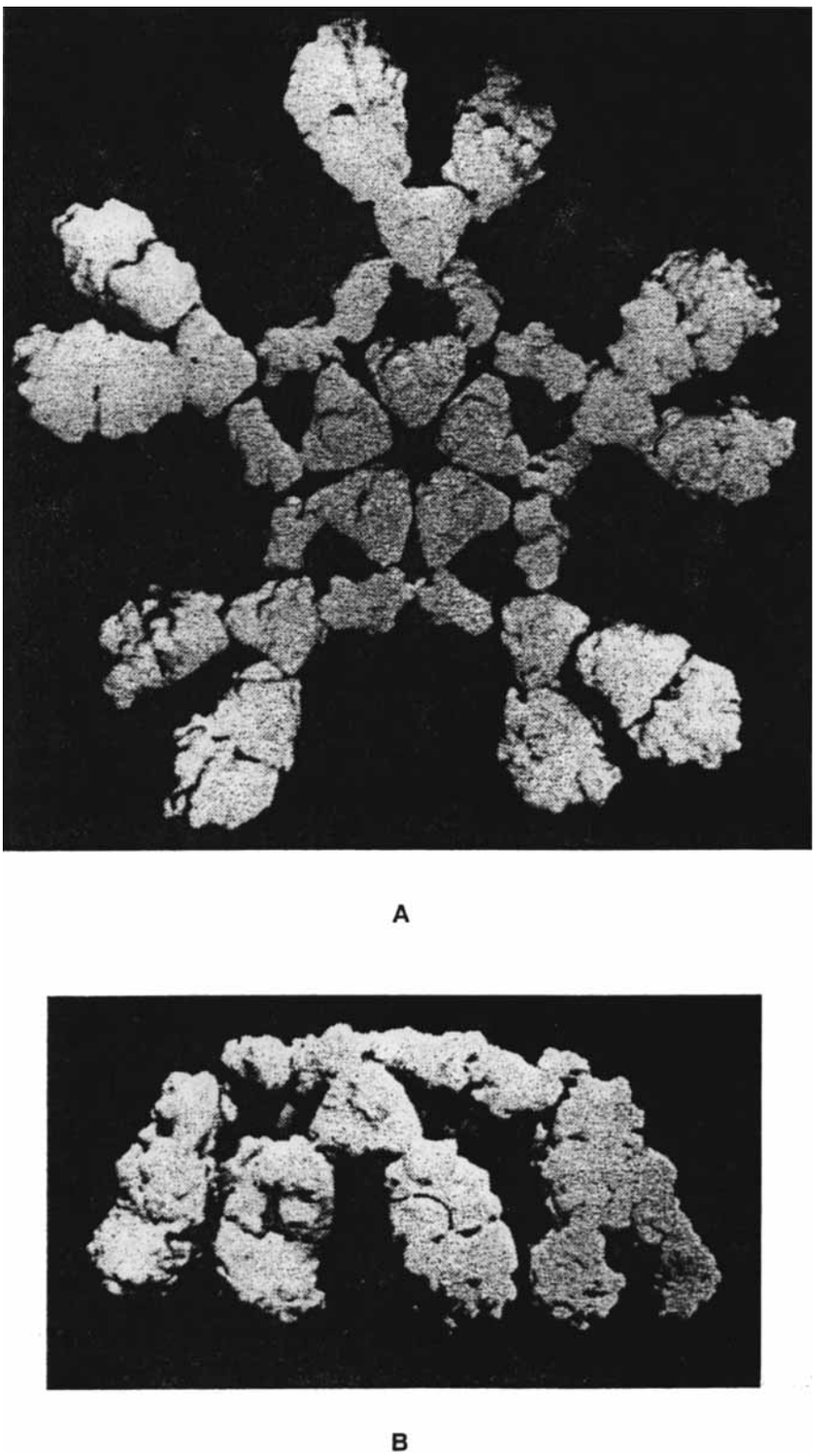


FIGURE 26. Model of the IgM pentamer in (A) the planar form and (B) the staple form. The planar form changes to the staple form after antigen binding. In the staple form, the oligomannose sugars in the tailpiece would become clustered together above the surface of the molecule in the center of the staple, where they may be available to take part in other events, such as complement activation. (From Feinstein, A. and Beale, D., 1977. *Immunochemistry,* (Glynn, L. E. and Stewart, M. W., Eds. New York: John Wiley & Sons. With permission.)

TABLE 2
The Location of Glycosylation sites on the leucocyte antigens (CD2, CD48, CD4, Thy-1)

Hu sCD2	Rat sCD2	Rat sCD48 (by sequence alignment)	Rat sCD4	Rat brain Thy-1
Asn65GlyThr; loop in cleft on V domain	Asn67LeuThr; loop in V domain	Asn13ValThr; loop in V domain	Asn159; middle of β-strand in C2 domain 2	Asn23AsnThr; junction of C-terminus of β-strand and loop in V domain
Oligomannose			Oligomannose	Oligomannose
Asn117ThrThr; loop in C2 domain	Asn77ValThr; (β-Strand in C2 domain)	Asn67ValSer; loop in C2 domain	Asn270; middle of β-strand in V domain 3	Asn74PheThr; junction of β-sheet and N terminal of loop in V domain
Complex (oligomannose)			Complex/hybrid	Complex/hybrid
Asn126GlyThr; loop in C2 domain	Asn84GlyThr; loop in C2 domain	Asn150LeuThr; loop in C2 domain		Asn98LysSer; middle of β-strand in V domain
Complex (oligomannose)				Complex/hybrid/ oligomannose
	Asn112AlaThr; loop in C2 domain	Asn167LysSer; β-sheet in C2 domain		

Note: These antigens are all members of the immunoglobulin superfamily. Where information is available the class(es) of the major oligosaccharides present at each site are given.

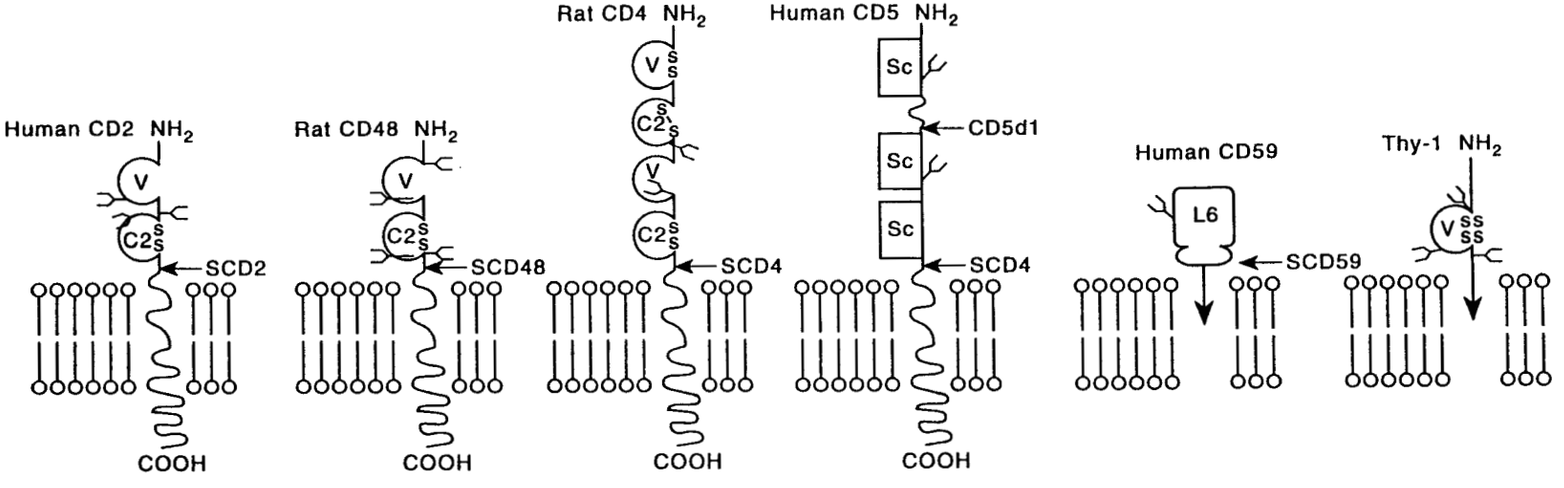


FIGURE 27. Schematic diagrams of some leukocyte cell surface antigens showing the N-glycosylation sites and proposed sites of O-glycosylation. Human CD2, rat CD48, rat CD4, and Thy-1 are drawn with circles representing the immunoglobulin superfamily (IgSF) domains and Y-shapes representing the N-linked glycans. The IgSF domains are designated as V or C2 on the basis of sequence analysis (Williams et al., 1989). CD5 contains three scavenger receptor domains (Sc), two N-glycosylation sites, and an O-linked site that may be located in the link region between domains 1 and 2 (Rudd et al., in preparation) . CD59 consists of a Ly domain and contains one N-linked glycosylation site and a variable number of O-glycans (Rudd et al., submitted). Recombinant proteins were expressed as soluble (S) forms with the anchor section deleted as indicated by the arrows.

However, in addition, the IgSF proteins (CD2, CD4, and CD48) contained a much wider range of structures than the immunoglobulins. These included hybrid structures (rat sCD48), polylactosamine structures (CD2, CD48), oligomannose sugars (CD2 and CD4), which in Igs have been noted only on the tailpiece of IgM and not on any of the Ig domains.

3. Glycan Heterogeneity and Local Protein Structure

The glycosylation sites in the leukocyte antigens CD2, CD4, CD5, CD48, and CD59 and their locations, whether on sheets, loops, or helices, are summarized in Table 2.

To compare the overall glycosylation of a set of proteins all expressed in the same cell line, the fluorescently labeled desialylated glycan pools were resolved by HPLC using the method of Guile et al. (1996). The results (Figure 12) showed that the overall glycosylation of each individual protein is different, indicating that each protein plays a role in determining its own glycosylation. Some interesting features of the glycosylation of these molecules are discussed below in the context of the proteins to which the sugars are attached.

1. Human CD2 (three glycosylation sites) and rat CD48 (four glycosylation sites) have many similarities in their protein sequence. However, while the pattern of complex and polylactosamine glycosylation is similar (Figure 12), only CD2 contains an oligomannose series. It has been shown (Recny et al., 1992) that Asn65 contains only oligomannose glycans. This may result from the location of Asn 65, which is attached to a loop on a V-type domain situated in a cleft in the protein, where it may be protected from GlcNAc transferase I.
2. In common with CD2, CD4 also contains one site (Asn159) that is glycosylated with oligomannose sugars (Ashford et al., 1993). Asn270 on CD4 (located in the middle of a β-strand on a C2 domain) contains complex and hybrid structures. Asn159 is also located in the middle of a β-strand but, like Asn65 on CD2, the glycosylation site is located in a V domain. The fact that these three sites all contain sugars with mannose branches suggests that none is freely accessible to the GlcNAc transferases and that Asn65 on CD2 and Asn159 on CD4 are totally inaccessible.
3. By comparison, the sugars attached to CD59 at the single glycosylation site located at the end of a β-strand on an Ly6 domain contains about 45% of polylactosamine-type structures. These data suggest that the location of a glycosylation site on a β-sheet (tertiary structure) does not restrict glycosylation per se. While the position (middle or end) on the β-strand may be a factor, in this case the overriding constraint is provided by the local 3D structure within the domain, or domains if a sugar is located at a domain interface.

4. Glycan Heterogeneity and Multiple Domains

There are very few examples of glycoproteins with single glycoforms (see Section IV.D). Remarkably, therefore, domain 1 of CD5 (CD5d1) (Figure 27), which has only one glycosylation site (Asn91CysSer), contains only the biantennary, core-fucosylated complex glycan (Figure 12) of which 83% was sialylated. No structural

reason for this restriction to the glycoform population can be proposed because the structure of the ScR domain is not yet known.

The glycosylation of a chimera consisting of human CD5/rat CD4d3 and 4 was compared with that of CD5d1 to investigate whether the range of sugars was increased by the presence of extra domains (Figure 12) (Rudd et al., 1996b, in preparation). The glycans associated with the four domains of rat soluble CD4 (expressed in CHO cells) (Figure 12) are heterogeneous and include 42% hybrid, 22% oligomannose, and only 14% of the biantennary, core-fucosylated complex glycan structures present on CD5d1. Domain 3 contained complex sugars with and without core fucose and hybrid sugars (Ashford et al., 1993). CD4 domain 3 and domain 4, which does not contain a glycosylation site, were incorporated into the chimera.

Interestingly, in the chimera, 95% of the glycans were of the same complex sugar type as CD5d1 and 100% were sialylated. By comparing the glycosylation of the chimera with that of Cd5d1 and native CD4 domains 3 and 4 (Figure 12) it was concluded that (1) there was no change in the type of complex sugar attached to CD5 domain 1 when this domain was part of the chimera, (2) the glycosylation of the scavenger domain 2 on CD5 in the chimera was identical to that on CD5 domain 1, and (3) no hybrid sugars were identified and all the complex sugars were fucosylated.

The addition of four domains to CD5d1 in CD5/CD4d3/4 therefore resulted in very little extra processing of the biantennary complex glycan that occupied the 100% of the glycosylation sites in CD5d1. In contrast to native CD4, the sugars attached to CD4 domains 3 and 4 in the chimera were all processed to complex biantennary sugars, and all were fucosylated. In this case, it appears that one domain in a multidomain protein may influence the glycosylation processing of other domains, perhaps by altering transit times or quaternary interactions.

5. Glycan Heterogeneity and Oligosaccharide Structure

In addition to the control exercised by the cell and the protein, the developing oligosaccharide chains may themselves control the efficiency of their interaction with some transferases. For example, analysis of the sugars attached to a number of leukocyte antigens revealed that shorter oligosaccharides (defined as one linear unit of the repeating disaccharide Galβ1,4GlcNAcβ1/arm) were more often sialylated than those containing more than one unit of lactosamine (Rudd et al., unpublished data). This is consistent with the findings of others (Merkle and Cummings, 1987; Edge and Spiro, 1985; and Fukuda et al., 1984) that indicate that sialylation is more likely to occur in the early stages of processing and that the incidence falls off as the chains become more elongated. Interestingly, shorter oligosaccharides more often contain α2,6-linked sialic acid, whereas the α-2,3 linkage is more common in the relatively less sialylated longer polylactosamine extensions (Merkle and Cummings, 1987), suggesting that the two sialyl transferases have a higher affinity for some oligosaccharide substrates than others.

C. Glycan Heterogeneity and the Local 3D Structure of the Protein

It appears from the first part of this review (Section III) that the functions of sugars may be divided into two groups: those that are general, where any sugar may suf-

fice, and those that require a specific glycan epitope at a particular site on a particular protein. To achieve the specificity required by the second group of functions, the glycosylation must not only be protein specific, but also site specific. In this section we examine the site-specific glycosylation of five different proteins: ribonuclease (RNase), the major histocompatability complex (MHC), the human leukocyte antigen (HLA), Thy-1, tissue plasminogen activator (t-PA), and Influenza A hemagglutinin (HA).

The presence of different glycan structures on different proteins from the same cell could be explained on the basis of different compartments and enzyme exposures. However, this rationale cannot explain differences in glycosylation at sites on a single protein. Analysis of the glycosylation at individual sites on the same molecule demonstrates that each site has a characteristic glycosylation pattern, and therefore that each of the initially identical oligosaccharide precursors in a protein represents a different substrate for the processing enzymes. Thus, in a given cell, the range of sugars that are attached at a glycosylation site reflect the primary, secondary, tertiary, and quaternary structure of the protein at the site as well as the glycosylation machinery in the cell in which it is expressed.

1. Site-Specific Processing of RNase

Human nonsecretory RNases from kidney, liver, and spleen have very similar, if not identical, protein backbones (Lawrence et al., 1993a). All are glycosylated with the same unusual tri, penta, and hexasaccharide structures (Lawrence et al., 1993b). This suggests that the glycosylation-processing pathways in the cells that express these enzymes are similar, and indeed it has been suggested that they may all be synthesized elsewhere (Glitz et al., 1993). In contrast, secretory RNases from human viscera and body fluids are organ specific (Yamashita, 1986; Hitoi et al., 1987).

In the human pancreas, Asn34 contains neutral hybrid-type sugars (Strecker and Montreuil, 1979), whereas Asn76 and Asn88 contain neutral complex-type glycans. In urine RNase, the sugars at all three sites are complex, and a proportion contain sialic acid (Ribo et al., 1994). This indicates that the 3D geometry around site 34 does not limit the access of the GlcNAc transferase III (assuming that this enzyme has the same structure in the different cell lines giving rise to these products). In addition, pancreatic RNase contains three to four times as much fucose as the urine form (Ribo et al., 1994). These data suggest that the processing of the glycans attached to secretory RNases is regulated at one level by the cell.

Each glycosylation site represents a different substrate for the enzymes and might therefore be characterized by different binding constants for the enzymes. This suggests that if the concentration of an enzyme or a sugar nucleotide donor falls below a critical level, a percentage of the sugars at sites with weaker binding characteristics may not be fully processed. For example, the existence of a hybrid structure at site 34 on pancreatic RNase suggests that the concentration of GlcNAc transferase III and/or UDPGlcNAc is too low to convert all the sugars at this site to complex-type structures. In contrast, under the same conditions, the sites at Asn76 and Asn88 bind sufficiently strongly to the GlcNAc transferase III to be fully processed to complex glycans. The final result of this process, which may be repeated at each stage of the processing pathway, is a heterogeneous mixture of glycoforms. However, it is now accepted that this heterogeneity is reproducible and not the result of a random pro-

cess. One of the early studies that showed this was that of the MHC antigens by Swiedler et al. (1985), which is discussed below.

2. Site-Specific Glycosylation of the Major Histocompatability Complex (MHC) and the Human Leukocyte Antigen (HLA)

A study by Swiedler et al. (1985) on the influence of peptide structure on class I and class II antigens was carried out using a murine B-cell lymphoma, AKTB-1β. This was the first study to be made of the site-specific glycosylation of several closely related, simultaneously synthesized, membrane proteins. The data indicated that there was a unique, reproducible distribution of glycans at each site resulting from variations in both sialylation and branching, even after many months of *in vivo* tumor passage. Importantly, this work indicated that heterogeneity is the result of a nonrandom process.

The analysis of the oligosaccharides on the HLA-DR and DC1 B-cell antigens (class II) expressed in (Epstein Barr virus) EBV transformed in human B-lymphoblastoid cell lines also demonstrated that glycosylation is site specific. A complex oligosaccharide was identified on the light chain, whereas the heavy chain contained both complex- and oligomannose-type structures (Shackelford and Strominger, 1983). Interestingly, the glycosylation of HLA class 1 molecules expressed in EBV-transformed cells was much more restricted, and the predominant structure at Asn86 was the biantennary complex glycan NA2F, with and without a bisecting GlcNAc residue (Figure 28) (Rudd et al., unpublished data).

3. Site-Specific Glycosylation of Thy-1

In contrast to the immunoglobulins that contain both C- and V-type domains, Thy-1 consists of only one domain, a V-type region that contains three glycosylation sites (Figure 21). A comparative analysis of the site-specific glycosylation patterns from rat (Asn23AsnThr, Asn74PheThr, Asn98LysSer), mouse (Asn23AsnThr, Asn75PheThr, Asn99LysSer), and human (Asn23ThrSer, Asn60PheThr, Asn100ValThr) neural Thy-1 shows that, although the domain accommodates all classes of glycans, there is site-specific processing (Williams et al., 1993). Despite differences in amino acid sequences, the pattern of N-glycosylation at corresponding sites was well conserved between the three species, suggesting a functional role for the sugars. This suggests that, within the context of neural cell glycosylation pathways, the local structure at the site, rather than the overall structure of the domain, controls the processing events.

For example, in rat brain Thy-1, site 23 (Asn-Asn-Thr) contains only oligomannose structures, whereas site 98 (Asn-Lys-Ser) is associated with complex, hybrid, and oligomannose structures, and site 74 (Asn-Phe-Thr) contains both complex and hybrid structures (Parekh et al., 1987). The local 3D structure of the protein determines the proximity of the protein surface to the glycans. This may be an important factor that determines the glycosylation pattern at individual glycosylation sites on proteins expressed in a single cell by controlling the affinity of different enzymes for the developing glycoforms.

It has been suggested that increasing complexity in glycan structure generally increases with the distance of the glycosylation site from the C terminus (Pollack and Atkinson, 1983). As is often the case in glycosylation, generalizations are difficult. For instance, in this case, human brain Thy-1 conforms to this pattern, although rat and mouse brain Thy-1 do not.

Although Thy-1 is one of the most abundant glycoproteins on mammalian neurons

HLA Class I

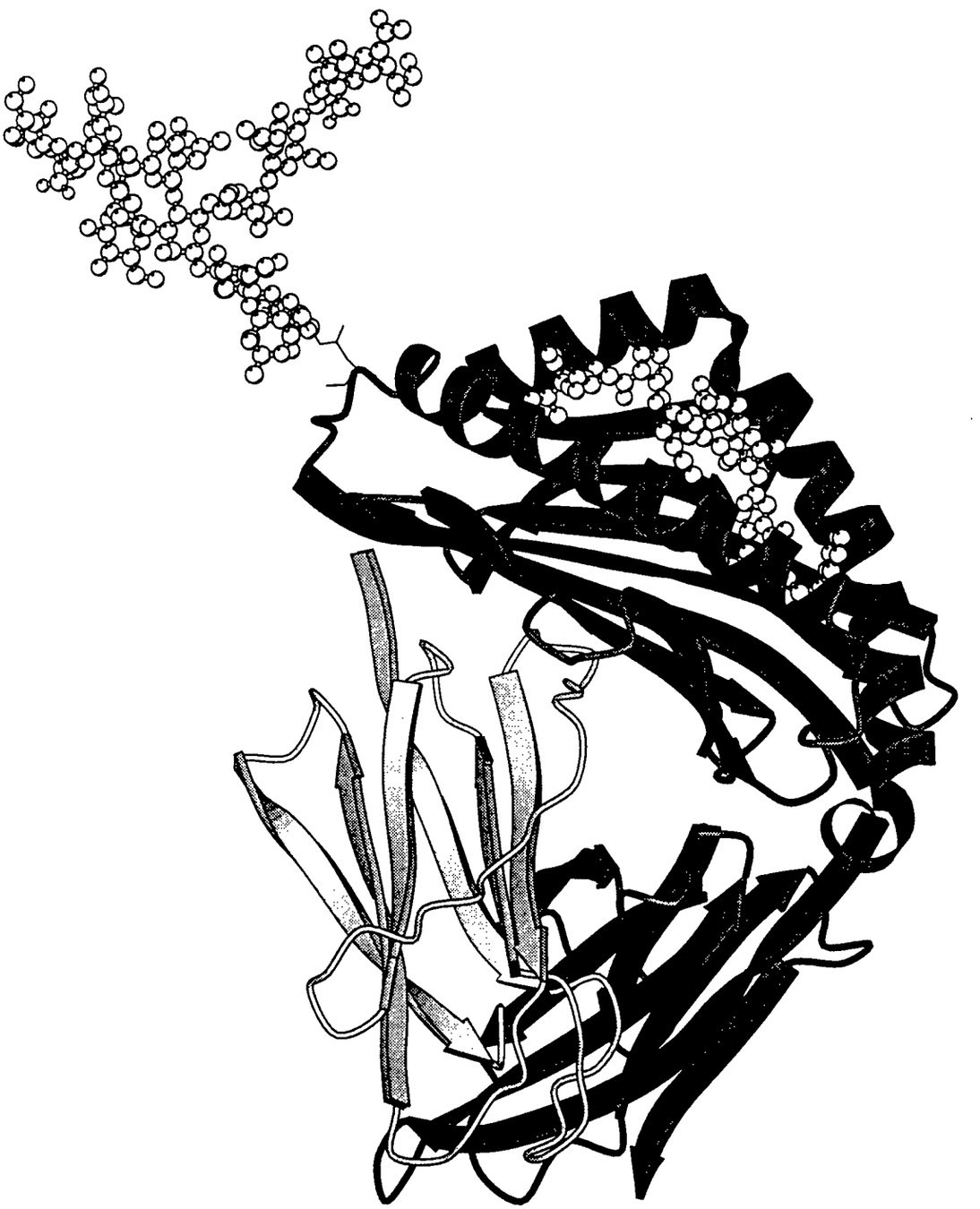


FIGURE 28. Molecular model of an HLA class I molecule expressed in an EBV-transformed cell line showing a peptide in the groove and a biantennary core-fucosylated complex glycan sugar at Asn86 (Rudd, P. M., Pazmany, L., Strominger, J. L., and Dwek, R. A., unpublished data).

the roles of both the protein and of its sugars remain elusive. A recent study by Dreyer et al. (1995), which identified an astrocyte receptor, suggested that neuronal Thy-1 binding to the receptor might actively promote neurite outgrowth or block an inhibitory effect of Thy-1.

4. Site-Specific Glycosylation of IgG

As a result of the different locations of the Fab and Fc sugars within the protein, IgG shows site-specific glycosylation. In the Fab, there is an increase in glycoforms containing G2-type (Section IV.A.3) sugars and in those containing bisecting GlcNAc compared with the Fc (Figure 23). This suggests that the glycans on the Fab are more accessible to galactosyl and GlcNAcV transferases that attach galactose and bisecting GlcNAc residues to the glycan chain. In contrast, the Fc oligosaccharides are partially protected by the protein structure, with which they interact, at the stages of the biosynthetic pathway when these enzymes act (see section). Interestingly, the proportion of G2 sugars which are sialylated is the same in the Fab as in the Fc, suggesting that the sialyl transferase is not inhibited by the protein structure in the Fc.

In normal serum IgG, approximately 60% of the Fab sugars are of the G2 type, compared with approximately 32% on the Fc. In contrast, the relative proportions of G1 on both Fab and Fc are similar (Figures 23B and C). In both cases, the ratio of G1(α1,6):G1(α1,3) is about 4:1 (Figure 22; Parekh, 1985). Interestingly, in chimeric mouse-human antibodies, the relative proportions of the arm-specific populations of G1 are related to IgG subclass, suggesting that subclass-related differences in the 3D structure of IgG may be important (Lund et al., 1993). Thus, in IgG1 and IgG4, the α-1,6 arm terminal GlcNAc appears to be in a more favorable conformation for interaction with the galactosyl transferase than in IgG2 and IgG3, where the α-1,3 arm is galactosylated preferentially.

In patients with RA, both serum and synovial IgG contain an increased proportion of glycans that lack galactose and terminate in *N*-acetylglucosamine (Parekh et al., 1984). The loss of galactose is limited to the Fc (Youings et al., 1996), suggesting that, if the effective enzyme levels fall in the cells associated with the production of IgG in these diseases (as has been suggested by Furukawa et al., 1990; Alavi et al., 1995), the more accessible Fab sugars will be galactosylated preferentially.

5. Site-Specific Glycosylation of Tissue Plasminogen Activator (t-PA)

t-PA (Figure 3A) shows species-, cell-, and site-specific patterns of glycosylation (Parekh et al., 1989a; Chan et al., 1991). Interestingly, at Asn117 (K1), oligomannose sugars are highly conserved between species (Howard et al., 1991), whereas the majority of the glycans at Asn184 (K2) and Asn448 (P domain) are complex and hybrid structures. This suggests that although Asn184 may not be fully accessible to the oligosaccharyl transferase, once the protein is fully folded, there is less restriction to the access of processing enzymes. At site 117, the reverse is true. The site is fully occupied, indicating that the Asn residue is accessible to the oligosaccharyl transferase. However, the sugars are not totally accessible to the mannosidase enzymes because oligomannose structures larger than Man5 were detected, and, in addition, they are inaccessible to GlcNAc transferase I because no hybrid or complex sugars are present at this site. This suggests that the

transient structures generated during the folding process (Section II.A.3) and the final tertiary and quaternary structure may all play a role in controlling the site occupancy and the class of oligosaccharides that are located at a glycosylation site.

In a study by Wilhelm et al. (1990), a series of t-PA mutants were expressed in CHO cells. These molecules were identical with respect to the location of the glycosylation sites within the domains, but differed in the order in which the domains were linked together. Interestingly, all the mutants except one lacked the oligomannose structures present at Asn117 on K1. The exception was a molecule in which the original domain arrangement was retained up to and through the glycosylation site at Asn117. This raises the possibility that the quaternary, rather than the tertiary, structure of t-PA is responsible for the restriction of processing at Asn117. Occupancy of site 184 affects the fine structure of the glycan population at site 448, demonstrating that glycosylation at one site can influence the processing at another. In Bowes melanoma type I t-PA, the major species at site 448 were neutral glycans of the complex or oligomannose type, and only 13% of the glycans were sialylated or sulfated complex or hybrid structures. In type II, however, 72% of the structures were sulfated complex glycans, and there were relatively few neutral sugars (Parekh et al., 1989a; Jaques et al., 1996). This suggests that in type I t-PA, the presence of a glycan at site 184 restricts the access of some of the glycosylation-processing enzymes at site 448, reducing their efficiency.

6. Site-Specific Glycosylation of Influenza A Hemagglutinin (HA)

The first study of the glycosylation of HA in mammalian (Madin-Derby bovine kidney, MDBK) cells showed that the conserved glycans at each of the four glycosylation sites on the HA1 subunit have distinct site-specific characteristics (Figure 8) (Mir-Shekari et al., in press), indicating the importance of the protein structure in controlling glycosylation. Asn21SerThr, situated on a loop, proximal to the membrane, was shown to contain tri- and tetraantennary glycans. Consistent with this was the absence of bisecting GlcNAc, which, when present, inhibits GlcNAc transferases IV and V that are responsible for branching (Harpaz and Schachter, 1980). The charge on the sugars at this site was derived mainly from sulfate substitutions that most commonly occur on outer-arm GlcNAc residues. In contrast, Asn65IleThr is situated in a buried helix and contains only neutral glycans of the oligomannose type. This is consistent with molecular modeling studies (Mir-Shekari et al., in press) that suggest that the sugars may no longer be accessible for processing after folding and trimerization of the HA polypeptides. The sugars at Asn129HisThr and Asn271AlaSer, which are located on a loop, are of the biantennary, bisected type. The three sites that contain complex-type sugars also contain α-galactose residues. The α1,3-galactosyl transferase, which is not found in human cells, competes with α2,6-sialyltransferase (Blanken and Van den Eijnden, 1985). The glycosylation processing of HA takes place on the fully assembled influenza virus. Sialidase is present as part of the virus and may cleave sialic acid residues from the sugars. The addition of sialic acid to complex glycans is normally a chain-termination step. Therefore, if terminal sialic acid is removed by the sialidase while the virus is still within the glycosylation pathway, the oligosaccharides attached to the viral surface may be extended further. In MBDK cells, this results in the preferential attachment of terminal α-galactose in MBDK cells.

In the case of HA, there is a correlation between the extent of processing that increases with the proximity of each glycosylation site to the membrane. All of the glycosyl transferases in the N-linked glycan-processing pathway to date are type II membrane proteins (Lowe, 1991). Once a polypeptide is folded, oligosaccharides proximal to the membrane may, in some cases, be able to interact better with the β-1,3GlcNAc and β-1,4Gal transferases involved in the addition of lactosamine units. Interestingly, large polylactosamine structures appear to be associated mainly with membrane-bound glycoproteins. For example, membrane-bound CD59 from erythrocytes or platelets is rich in polylactosamines that are present at only low levels in soluble CD59 secreted from CHO cells (Rudd, P. M., unpublished data).

D. Glycan Heterogeneity and the Function of Sugars

While all N-linked glycoproteins enter the oligosaccharide-processing pathway as a single $Glc_3Man_9GlcNAc_2$ glycoform, most emerge from the *trans*-Golgi as a heterogeneous mixture of glycosylated variants. This extensive heterogeneity poses major questions for glycobiology. For example, does each individual glycoform, or structurally related population of glycoforms, have a particular function? Alternatively, does extensive heterogeneity imply that the particular structures are relatively unimportant and that any range of glycans can fulfill the roles that sugars play in the life of a particular protein? Indeed, are there both general roles for glycosylation and roles for specific sugars? How can oligosaccharides alone function as specific recognition motifs when the same structures appear on many proteins? Is there any evidence that receptors recognize specific subsets of glycoforms or that different glycoforms of the same protein can be recognized by different receptors? Is heterogeneity compatible with a function for oligosaccharides in recognition?

In the next section, we discuss some roles for glycan heterogeneity in relation to the structure of the oligosaccharides and their presentation by the protein on the cell surface.

1. Glycan Heterogeneity and the Cell Surface

The lipid bilayer of most cells contains a densely packed array of glycoproteins, most of which are heavily glycosylated. The surface is therefore covered with a wide range of heterogeneous glycans. The analysis of soluble leukocyte antigens expressed in the same CHO cell line indicated that glycosylation is protein specific, suggesting that a single cell surface presents many different sugars. As yet, there is little information about the glycosylation patterns on the surface of cells, but some functions for the glycosylation of cell surface molecules are beginning to emerge.

2. Glycan Heterogeneity and the Orientation and Packing of the Cell Surface Protein CD59

Human erythrocyte CD59 consists of a mixture of at least 100 different glycoforms of which approximately 50% are biantennary complex type structures with lactosamine extensions (Figure 12) (Rudd et al., in preparation). The average length of the sugars attached to human erythrocyte CD59 is roughly of the same order of magnitude as the diameter of the globular protein (4.2 nm), and they are located on the opposite side of

the molecule to the binding site for C5β-8/C9 (Figure 15). The active site, which includes Trp40, Arg53, Glu56, and, possibly, Asp24, is on the membrane distal surface of the extracellular domain (Bodian et al., submitted). The glycans on CD59 do not appear to contribute to the affinity of CD59 for the C8 and C9 components of the C5β-9 complex (Ninomiya et al., 1992; Bodian et al., submitted), but may be expected to restrict the rotational freedom of the extracellular domain around axes parallel to the membrane. In turn, this would stabilize an exposed location for the active face. In addition, the bulky hydrophilic glycans would limit interactions with the lipid bilayer and in this way may facilitate the diffusion of the protein in the membrane. Figure 29 shows three differently glycosylated CD59 molecules inserted into the cell membrane. The heterogeneity of the sugars suggests that the glycans influence the geometry of the packing and that the glycans will prevent the aggregation of CD59 molecules on the cell surface. By limiting such protein-protein interactions, the glycans may influence the distribution of CD59 molecules at the cell surface because GPI-anchored proteins are believed to associate in microdomains in dynamic equilibrium with isolated individual molecules. The large N-glycans may also be important in preventing proteolysis of the extracellular domain because glycosylation has been shown to increase the dynamic stability of a protein, whereas different glycoforms variably increase its resistance to protease digestion (Rudd et al., 1994a).

3. Glycan Heterogeneity and the Cell Adhesion Molecules CD2 and CD48

In the immune system, adhesion of T cells to antigen-presenting cells is mediated by receptors on both cells. CD2 and LFA3 (CD48 in rat; CD58 in human) are present on T cells and antigen-presenting cells, respectively, and interact through homologous-binding surfaces (van der Merwe et al., 1995) (Figure 30). The dimensions of the sugars that make up the heterogeneous arrays of glycans associated with CD2 and CD48 are approximately the same size as the protein domains (about 4 nm). In CD48, the sugars are all located on one face of the molecule, and in neither molecule is the sugar close to the binding sites, which are located in the amino-terminal domains, distal to the membrane surfaces. The binding sites of both CD2 and LFA3 are confined to a region described by a cone of 52° (Dustin et al., 1996). The crystal structure suggests that for *cis*-interactions between CD2 molecules on the same cell to occur a confinement angle >65° would be required (Jones et al., 1992). Molecular models (Figure 30) indicate that the sugars at Asn126 on CD2 and Asn167 on CD48 are both located close to the membrane surface, where they may orient the binding site by limiting the conformational space available to the protein. This may provide a physical explanation for the confinement cone described by the CD2 that promotes *trans*-interactions with CD48 on other cells and discourages *cis*-interactions with other CD2 molecules on the same cell. In addition, CD2 molecules, which contain GPI anchors, cluster when presented with LFA3 receptors (Dustin et al., 1996). During clustering nonspecific protein interactions of CD2 may be inhibited by the heterogeneous arrays of sugars that variably shield the protein surfaces. A further role for sugars may also be to protect the protein surface from the activity of proteases.

CD2 shows site-specific glycosylation, and only oligomannose glycans are present at Asn65 (Recny et al., 1992), which is located at a cleft in the protein that may restrict access to GlcNAc transferase I. A specific role for sugars at Asn65 on CD2

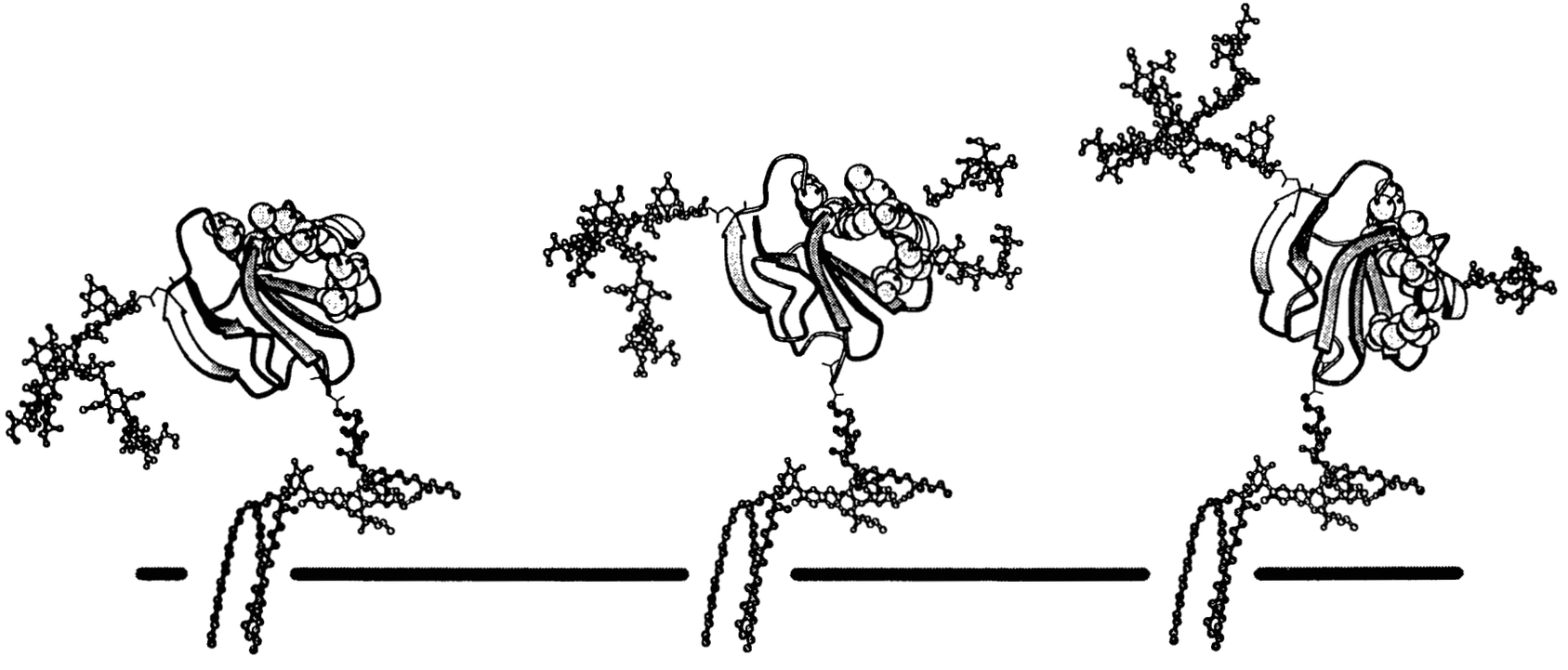


FIGURE 29. Schematic figure showing the effect of CD59 glycosylation on the flexibility of the protein relative to the GPI anchor. Three different glycoforms are shown. Both the O- and N-linked oligosaccharides (size range, 2 to 6 nm) attached to CD59 (diameter approximately 3 nm) restrict the conformational space available to the protein and limit its interaction with the lipid bilayer. The sugars may therefore orient the N-terminal active site of CD59 toward the C5b-9 complex, which is also inserted into the cell membrane. In addition, the heterogeneity of the sugars suggests that the glycans influence the geometry of the packing, and it is unlikely that CD59 molecules will form a regular array on the cell surface.

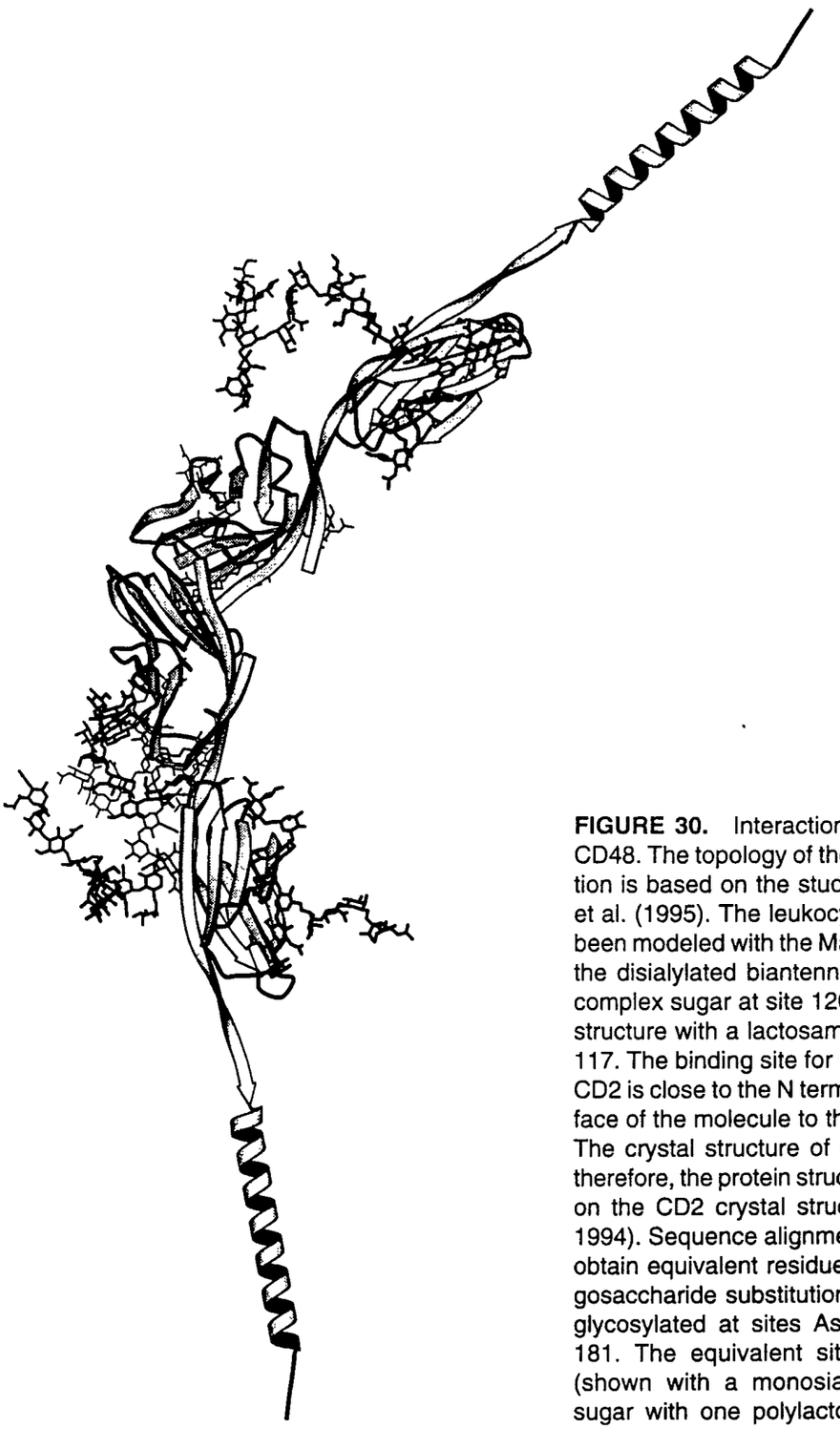


FIGURE 30. Interaction of rat CD2 and rat CD48. The topology of the CD2/CD48 interaction is based on the study by van der Merwe et al. (1995). The leukocyte antigen CD2 has been modeled with the Man6 glycan at site 65, the disialylated biantennary core-fucosylated complex sugar at site 126, and a triantennary structure with a lactosamine extension at site 117. The binding site for CD48 on monomeric CD2 is close to the N terminus on the opposite face of the molecule to the glycans at site 65. The crystal structure of CD48 is not known; therefore, the protein structure has been based on the CD2 crystal structure (Bodian et al., 1994). Sequence alignment has been used to obtain equivalent residues in CD2 for the oligosaccharide substitutions in CD48. CD48 is glycosylated at sites Asn16, 150, 164, and 181. The equivalent sites on CD2 are 13 (shown with a monosialylated biantennary sugar with one polylactosamine repeat), 67 (shown with a tetraantennary glycan), 150 and 167 (shown with a biantennary, core-fucosylated complex sugar)

may be to stabilize the protein by counterbalancing the clustering of positive charges created by five lysine residues situated on the surface of domain 1 (Wyss et al., 1995). Interestingly, only the $Man_1GlcNAc_2$ core motif of the N-linked sugar contributes to the interaction suggesting that any oligomannose structure (and possibly any complex sugar which could similarly occupy the restricted space available within the cleft in the protein) could fulfill the function. The sugar attached at the flexible link region (Asn117) in CD2 may be in a position to affect the orientation of the two domains, and again this function may not require a specific glycan structure.

E. Is Glycan Heterogeneity Compatible with Recognition?

Sugars may impart a secondary function to a protein that is over and above its normal activity, such as the potential for recognition of a glycoprotein through its carbohydrates. We discussed earlier the observation that glycoform homogeneity is rare (Section III.A.5). However, just as certain 'motifs' in a protein rather than the entirety may be involved in recognition events, so the same is true for oligosaccharides. On this basis, glycan heterogeneity is clearly compatible with recognition. This raises three further questions:

1. Is the recognition of an oligosaccharide motif independent of the 3D structure of the protein and glycoform?
2. How can oligosaccharides function as physiologically relevant recognition motifs when the same structures appear on many proteins?
3. How can physiologically relevant binding be achieved when recognition of an individual monosaccharide is usually characterized by a weak binding constant?

1. Presentation of Oligosaccharide Motifs and the 3D Structure of the Protein

In some cases, recognition of an oligosaccharide motif may be independent of the structure of the rest of the sugar to which it is attached and of the 3D structure of the individual protein. For example, many of the proteins that are reported to be apically secreted contain N-glycans. These include erythropoietin (Epo), clusterin (gp80), human corticosteroid binding globulin (hCBG), polymeric immunoglobulin receptor, neutral endopeptidase, decay accelerating factor, and placental alkaline phosphatase (reviewed in Scheiffele et al., 1995). This range of glycoproteins suggests that the pathway is independent of both the 3D structure of the protein and the sugars because proteins carrying oligomannose and complex-type sugars are found on the apical cell surface. This suggests that the recognition motif is the common tri-mannosyl core. One putative lectin that has been proposed is VIP36, which has been identified in apical transport vesicles. Interestingly, GPI-anchored proteins are also apically secreted and are sorted to glycolipid-enriched membrane subdomains during transport to the apical cell surface (Brown et al., 1992).

Other examples in which recognition of a glycan motif is independent of the 3D structure of the glycoprotein include the mannose-6-phosphate receptor and ERGIC-53. Glycoproteins that carry a mannose-6-phosphate residue are incorporated into the lysosome (Kornfeld, 1992), whereas glycoproteins that contain terminal mannose residues bind ERGIC-53 (Arar et al., 1995).

This lectin continuously recycles between the ER and the *cis*-Golgi and may be involved in the trafficking of glycoproteins between these two organelles. ERGIC-53 may also bind to the terminal mannose (D1) on the oligosaccharide $Man_9GlcNAc_2$ following removal of the glucose residues, thus protecting the sugar from further processing (Arar et al., 1995).

In some cases, the 3D structure of the protein can prevent the recognition of a glycan motif. For example, the trimannosyl core that is present on all N-glycans may be recognized by peptide N-glycanase (Suzuki et al., 1995). In general, peptide N-glycanases (PNGases) may function both as lectins that mediate glycoprotein transport and as enzymes that regulate the biosynthesis and catabolism of glycoproteins (Suzuki et al., 1994). When PNGase acts as an enzyme, it cleaves the sugar-amide linkage in glycoproteins containing N-glycans. In some cases, the 3D structure of the protein plays a role in protecting the bond. For example, peptide N-glycanase F cleaves the Asn18-sugar amide linkage in native CD59 and the Asn34-sugar amide linkage in denatured but not in native RNase B (Rudd, 1996). Molecular modeling suggests that the 3D structure of the RNase protein prevents the N-glycosidic linkage from interacting with the binding site of the deglycosylating enzyme (Kuhn et al., 1995), whereas in CD59 the access is unrestricted (Figure 31).

The recognition of glycan motifs may also depend on the structure of other regions of the oligosaccharide that may modulate the binding through noncovalent interactions with the receptor or by influencing hydration patterns. An example is galectin 3, which recognizes terminal galactose residues. Studies of the binding specificity (Sato and Hughes, 1992) showed that the inhibition of binding to asialo fetuin-sepharose by straight-chain polylactosamine structures of complex-type branched glycans increased in proportion to the number of Galβ1-3(4) units present. Although substitutions in the terminal residue did not affect binding, substitution of the penultimate *N*-acetylglucosamine residue did. This suggests that, although not directly recognized, the structure of the remainder of the sugar modulates the efficiency of the binding event.

Many lysosomal membrane and erythrocyte glycoproteins contain high levels of consecutive *N*-acetyllactosamine residues in polylactosaminoglycans, which are oligosaccharide motifs with a high affinity for galectins. For example, 50% of the glycoforms of CD59 are of the polylactosamine type (Rudd et al., 1977). Heterogeneity may therefore allow the recognition of subsets of glycoforms within a population.

Finally, we note that a single oligosaccharide may contain several different epitopes that can be recognized in much the same way as different regions in a protein may be recognized by different receptors. For example, the $Glc_1Man_9GlcNAc_2$ structure is recognized by calnexin, calreticulin, glucosyl transferase, α-mannosidase, endomannosidase I, and glucosidase II. In each of these interactions, only a small number of monosaccharides constitute the epitope required for binding. In the case of calnexin, the $Glc_1Man_9GlcNAc_2$ glycoform binds with higher affinity than the $Glc_1Man_5GlcNAc_2$ glycoform, suggesting that the structure of the remainder of the oligomannose glycan can modulate the binding.

2. Physiologically Relevant Binding

In general, monosaccharides are weakly recognized. Biological specificity usually arises from multivalent binding. Triggering

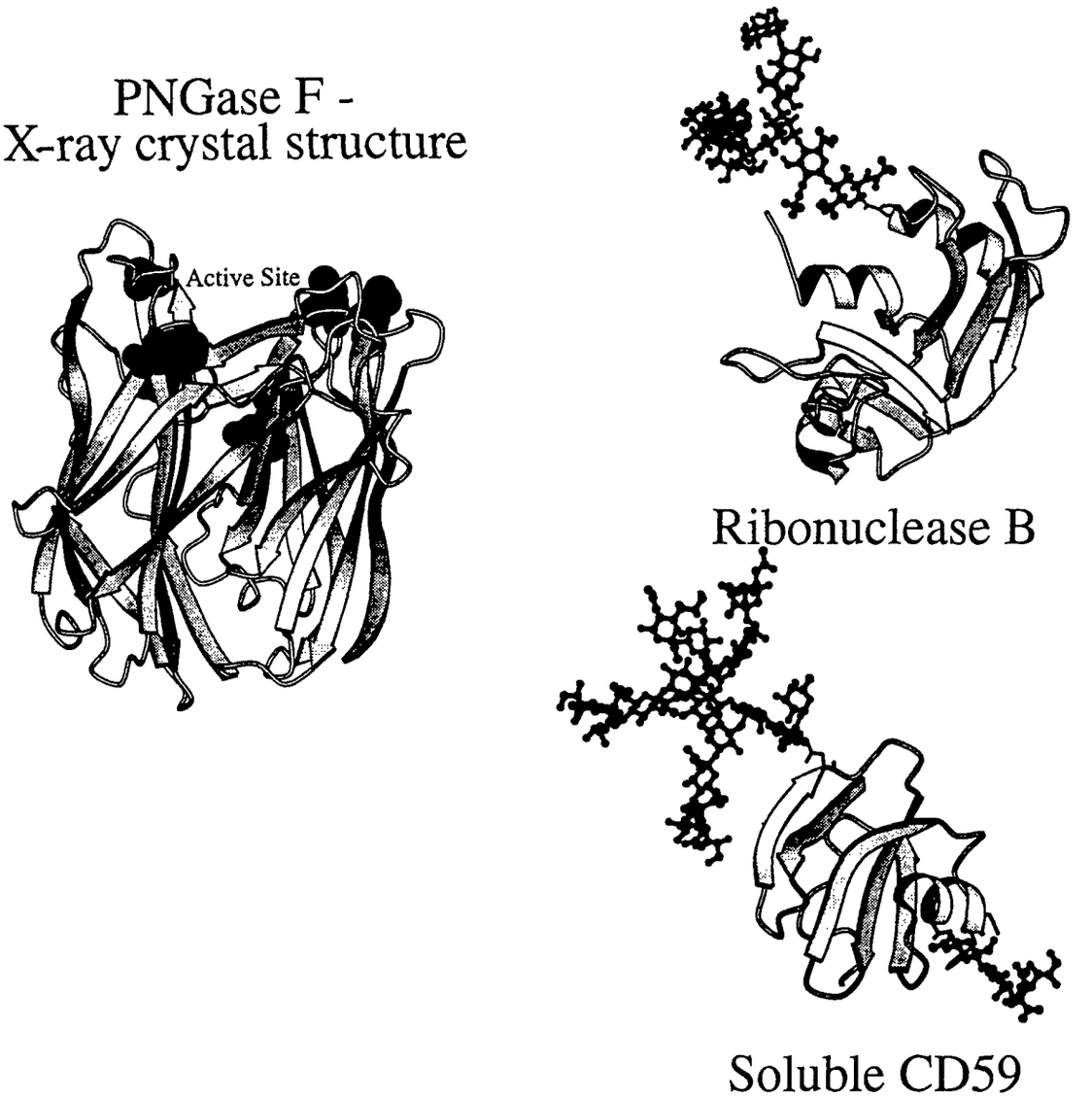
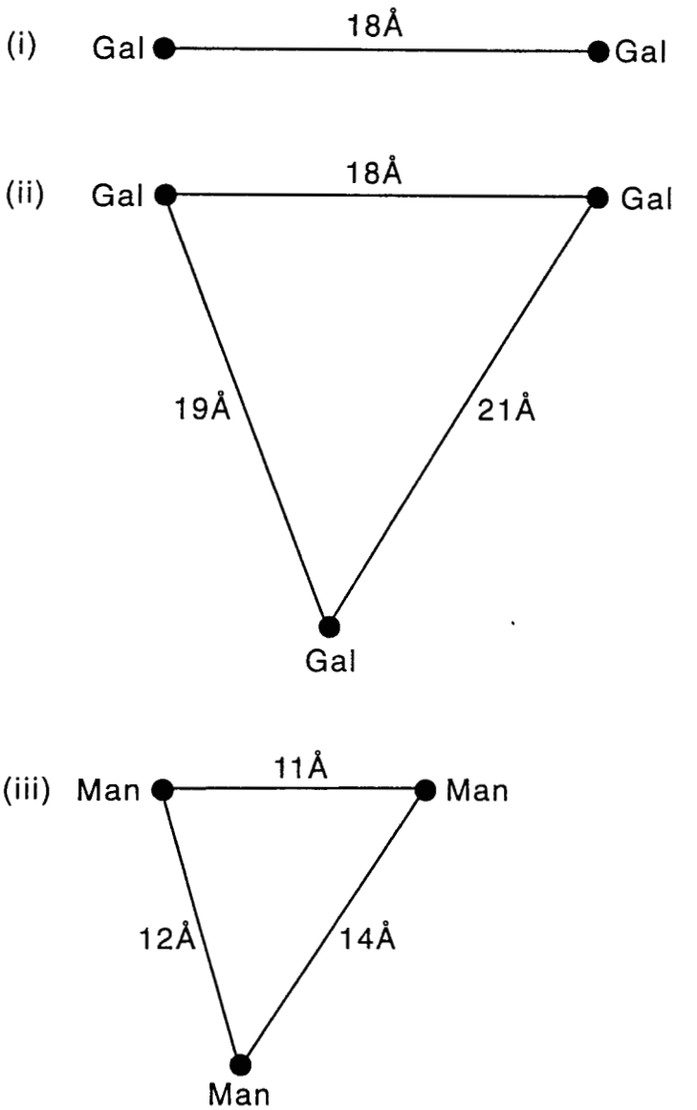


FIGURE 31. The crystal structures of CD59, RNase B, and PNGase F. The structure of PNGase F is based on the crystal structure (Norris et al., 1994). The protein is folded into two domains, each with an eight-stranded, antiparallel β-jelly roll configuration similar to the types of structures found in lectins. CD59 was modeled according to the NMR solution structure (Kieffer et al., 1994). RNase B was modeled according to the crystal structure (Williams et al., 1987). All three molecules are modeled to the same scale. The active site for PNGase F is in a cleft flanked by five tryptophan residues and one phenylalanine residue (Kuhn et al., 1995). In contrast to the CD59 asparagine-sugar amide linkage, which is susceptible to the enzyme, the same linkage in RNase is protected by the protein and not accessible.

of functions associated with the receptor requires the multivalent presentation of the sugars to be geometrically matched with binding sites on the multivalent receptor. For example, in the case of the asialoglycoprotein receptor, which is a C-type lectin, several identical residues on the same sugar may form the multivalent epitope (Fig-

ure 32). The receptor binds with high affinity to naturally occurring complex oligosaccharides found in desialylated glycoproteins (Lee et al., 1983). Although biantennary glycans bound to the receptor through terminal galactose residues, a triantennary structure containing three terminal β1-4galactose residues showed a significantly in-



A

FIGURE 32. **(A)** Diagram showing typical distances between terminal residues in energy-minimized conformations of the (1) bi- and (2) triantennary complex oligosaccharides A2G2 and A3G3, respectively. The triantennary glycan is a high-affinity ligand for the asialo glycoprotein receptor. (3) Typical distances between the terminal residues of Man_9, which is recognized by the mannose-binding protein for energy-minimized conformations. **(B).** Molecular model of the $Man_9GlcNAc_2$ sugar showing (left) the dimensions of the oligosaccharide chain and (right) two possible configurations of Man_9 showing the distances between the terminal residues. The figures on the left and top right are of energy-minimized structures.

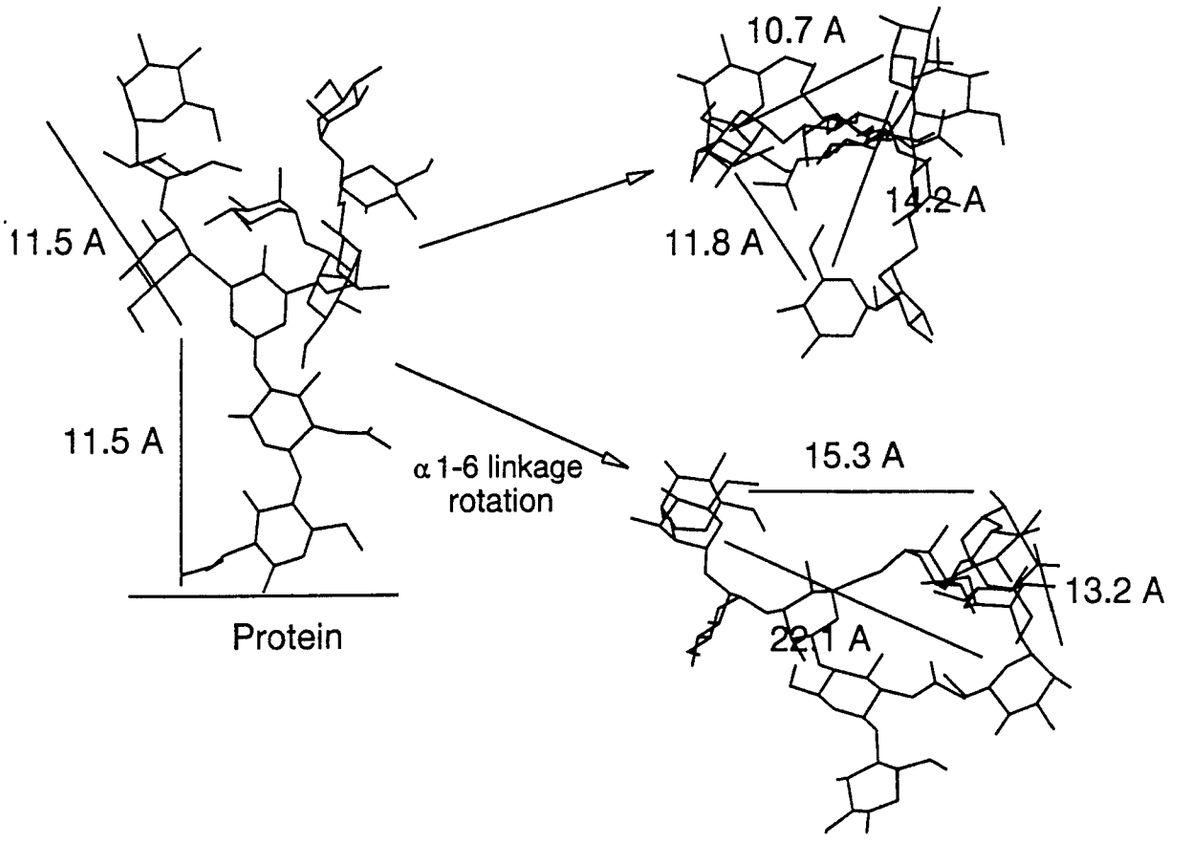


FIGURE 32B

creased affinity as a result of polyvalency. The 3D structure of the peptide was shown to have negligible influence on the binding, which involved a precise geometry between the trivalent ligand and the lectin (Rice et al., 1990). The finding that the affinity of triantennary sugars is greater than biantennary sugars, which is greater than the affinity of a monoantennary structure, suggests that a single oligosaccharide can bridge the gap between the three CRDs in one subunit of the asialoglycoprotein receptor. The distances between the terminal galactose residues on bi- and triantennary sugars in the minimum energy conformations are shown in Figure 32A and may give some insight into the distances between the CRDs.

In general, inspection of the geometry of the sugar epitope may be a useful means of obtaining information about the presentation of carbohydrate recognition domains. For example, the molecular models of the Man_9 structure (Figure 32B) show the distances between the terminal residues for two different conformations of the glycan. Terminal mannose residues are recognized by the C-type lectin, mannose binding protein (MBP), which is an example of a receptor that achieves specificity through multivalency (for reviews see Weis and Drickamer,1994; Hoppe and Reid, 1994). MBP specifically recognizes *cis*-hydroxyl groups on some terminal monosaccharide residues, including mannose. One molecule of serum MBP contains up to 18 identical CRDs organized into six subunits. The affinity of one CRD for a single terminal monosaccharide is relatively weak (kd, 1

mM) and for MBP to activate complement, multiple CRDs must interact with multiply presented ligands. The distance between the carbohydrate-binding sites is 44 Å on human MBP (Weis and Drickamer, 1994). Because the maximum distance between the mannose residues on an oligomannose structure is approximately 16 Å, the oligomannose structure cannot bind to more than one CRD within a subunit. It is the geometry of the oligomannose structures on the cell surfaces of various bacteria and fungi that allows multivalency to occur and that results in recognition of the foreign body by the MBP and neutralization by complement-mediated cell lysis or by opsonization.

Thus, multivalency may be a means of distinguishing between self and non-self. In contrast to the surfaces of bacteria and fungi, mammalian cells and proteins do not generally carry multiply presented glycan residues capable of interacting with the MBP in a physiologically relevant manner.

3. Recognition May Involve Subsites

For completion, we note that oligosaccharide binding can also be achieved by having a number of subsites that bind monosaccharides. In some cases, recognition requires the cooperative binding of several monosaccharides in the glycan, such as in a lysozyme, where there are six subsites (Perkins et al., 1981) or in the antibody recognition of *Salmonella typhimurium* (Cygler et al., 1991). This type of recognition involves precise 3D geometry of both the oligosaccharide and the protein to allow the interaction of complementary regions, which are then normally stabilized by hydrogen bonding and van der Waals forces.

Sialoadhesin and CD22 are members of a family of sialic acid-dependent cell adhesion molecules that are presented as multimers on the surface of macrophages and B cells, respectively. Both bind terminal disaccharide motifs by the cooperative binding of two monosaccharide residues in subsites. Sialoadhesin, which contains 17 Ig-like domains, recognizes sugars terminating in NeuAcα2-3Gal in N- and O-glycans through the amino-terminal V-set Ig-like domain 1. In contrast, CD22 contains seven Ig-like domains and binds NeuAcα2-6Gal in N-glycans. The binding site for CD22 is also contained in the terminal V-set domain 1, but the adjacent C2 set domain is also necessary and may be required for correct folding of domain 1 (Nath et al., 1995).

V. THE FUTURE

A. Glycan Heterogeneity and the Future of Technology

There have been many advances in the technology available for oligosaccharide analysis in the last 10 years. However, the complete analysis of the oligosaccharides attached to a glycoprotein is still time consuming, and the challenge to automate the process remains. In this respect, a strategy has been developed recently that involves a sensitive and reproducible HPLC that is capable of resolving subpicomolar quantities of heterogeneous mixtures of fluorescently labeled neutral and acidic glycans simultaneously and in their correct molar proportions (Guile et al., 1996). It has been shown that the elution positions of specific glycans can be predicted and that this is a valuable aid to interpreting the full glycosylation profiles of glycoproteins. The fine specificity of the resolution enabled the N-glycans released from human platelet

CD59 to be separated into more than 36 peaks (Figure 12). The reproducibility of the separation system, the predictability of glucose unit values, and the quantitative response of the detection system for individual fluorescently labeled glycans has also allowed the simultaneous analysis of pools of sugars using arrays of exoglycosidases (Figure 33).

In the future, this strategy will enable the rapid accumulation of computer-generated oligosaccharide databases for individual glycoproteins. This technology has also been used to distinguish between a range of autoimmune diseases by comparing the differences in the glycosylation of IgG (Figure 20) (Watson et al., in preparation) In the future, computer-generated one-, two-, and three-dimensional oligosaccharide maps (Nakagawa et al., 1995; Takahashi, 1995) will lead to a rapid means of comparing glycosylation in, for example, different physiological states, different diseases, and different stages of development and differentiation.

B. Glycan Heterogeneity and the Structure of Glycoproteins

It is clear that information is lost when the constituent sugars and protein that make up a glycoprotein are perceived as two separate entities. Although the analyses of the component parts may be undertaken separately, the full implications of glycosylation for the protein only become apparent when functional and structural studies take account of the whole molecule and its range of glycoforms. One difficulty in envisaging a glycoprotein as a single entity stems from the fact that the available methods for analyzing protein structure do not yield high-quality information about the sugars. A limitation of X-ray crystallography is that the mobility and heterogeneity of the sugar results in poor definition. NMR solution structures of glycoproteins do not have this problem; however, only the major glycoforms in a mixture can be studied.

One approach to the problem of heterogeneity would be to generate a database of structural units from which the oligosaccharides attached to a protein can be modeled and then built onto the protein structure. The justification for this stems from the fact that in contrast to proteins, where tertiary and quaternary interactions play a significant role, the main structural features of oligosaccharides are determined by their secondary structure, which involves interactions across individual glycosidic linkages.

For example, the conformation of the chitobiose core is independent of both the protein and the outer-arm sugars (Woods, 1995; Dwek, 1996). This observation suggests that oligosaccharide structures can be predicted from their primary sequences by linking together smaller structural units. This involves generating a database of ϕ and ψ angles (Figure 34) for a range of appropriate units. Molecular mechanical forcefields can then be applied to obtain energy-minimized structures in water, giving the average structure of a small oligosaccharide with a high degree of confidence (Woods et al., 1993).

Two other features that are important are the dynamics of the oligosaccharide itself and the flexibility of the protein side chain to which the sugars are attached. Features such as the space that the sugars occupy and their presentation can be examined in a whole range of glycoforms. Inspection of molecular models constructed in this way allows both long- and short-range interactions of the sugars with the protein surface to be postulated. In addition, the extent to which different sugars shield the surface of the protein or active sites may

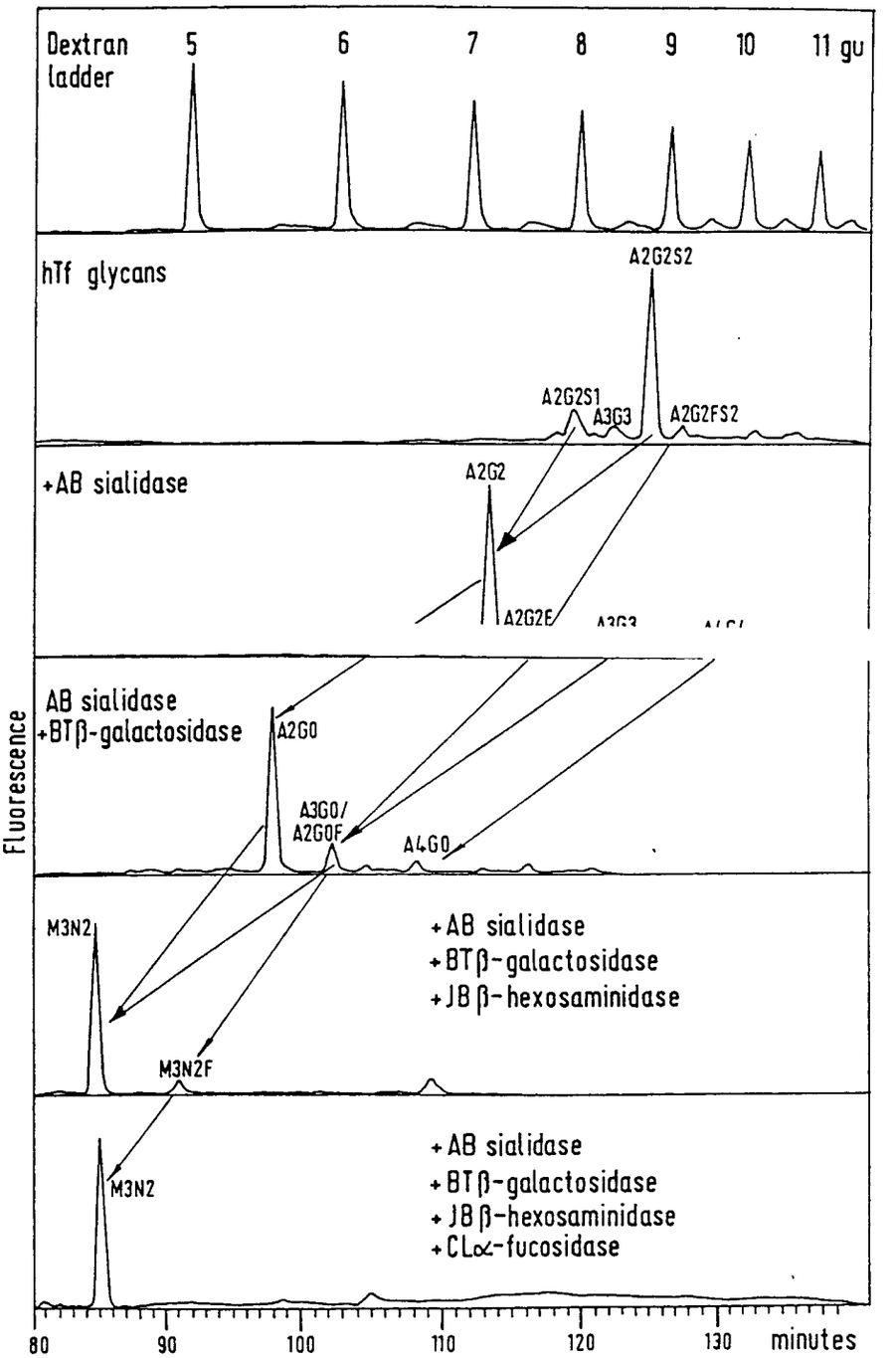


FIGURE 33. HPLC analysis of the products of the multiple exoglycosidase enzyme digests of the pool of fluorescently labeled sugars released from serum transferrin from a control individual. Five aliquots of the total glycan pool were digested simultaneously with the sets of enzyme arrays shown in the panels. The products of each of the digests were analyzed by normal phase HPLC and are shown in the figure. The peaks were assigned (lourin et al., 1996) and have been annotated. The arrows indicate the new elution positions of each glycan in the pool, indicating the effect of the additional enzyme in each digestion in the series.

be assessed. However, sugar linkages are flexible; therefore, the confidence in the predicted topology decreases to some extent for larger sugars that contain more linkages.

C. Molecular Modeling of Glycoproteins

Tissue plasminogen activator (Figure 3A) naturally exists in two forms. Type I contains sugars at Asn117, 184, and 448. In type II t-PA, the sugar at Asn184 is absent and the rate of plasminogen activation in the presence of fibrin is increased twofold. t-PA binds to terminal lysine residues in degraded fibrin via the lysine- binding site in K2. As can be seen in Figure 3A, molecular modeling suggested that glycosylation at site184 in type I t-PA may sterically hinder this interaction, thereby decreasing the rate of plasminogen activation (Parekh et al., 1989b). In addition, molecular modeling was used to explore the possible structure of a disulfide-bonded t-PA dimer that had been detected experimentally (Mori et al., 1995).

Bovine pancreatic RNase B is another example of a glycoprotein in which molecular modeling of oligosaccharides has been combined with 3D structural data to explore the interactions between the protein and the sugar and to provide a structural overview of a molecule with a well-defined protein structure (Rudd et al., 1995). Bovine pancreatic RNase B has one N-glycosylation site at Asn34 and consists of a mixture of five glycoforms RNase B Man5-9. There is a well-resolved 2.5-Å X-ray crystal structure of RNase B (Figure 35), although, typically, the electron density associated with the sugars was poorly defined (Williams et al., 1987). NMR studies have shown that the RNase sugars have independent motions and that the oligosaccharides do not interact with the protein surface (Joao et al., 1992). The oligosaccharides were modeled independently, and dynamics simulations carried out to map the trajectory of the sugars. A molecular model of the glycoprotein, obtained by combining all these data, was used to explore the effect of the flexibility of the amide side chain on space occupied by the oligosaccharides. The outcome of these combined dynamic effects is that a considerable area of the protein is protected by the sugar. The model was then used to explore the role of the oligosacccharides in the modulation of the activity of RNase. Important amino acid residues for the interaction of RNase with RNA are those within the active site of the enzyme and also a number that form salt bridges with phosphate groups in the backbone of RNA. There are five of these bridges, including one in the active site where cleavage of the phosphodiester bond takes place. In particular, a cluster of residues (Lys31 Lys37 Arg10 and Arg33) provides a strong local anion binding site where the 5′ terminal phosphate of RNA interacts with the enzyme (McPherson et al., 1986). The size of all the oligosaccharides associated with RNase is such that they can cover this cationic cluster, suggesting that they may hinder the formation of the salt bridge, reducing the overall fit between enzyme and substrate. This prediction was supported experimentally by the finding that the activity of unglycosylated RNase A was more than three times greater than RNase B (Rudd et al., 1995), indicating that the oligomannose sugars of RNaseB are a factor that modulates the rate at which RNA is degraded.

Another glycoprotein that contains oligomannose sugars is CD2. The glycosylation at site Asn65 contains only oligomannose sugars. The solution structure of domain 1 of recombinant human CD2 indicates that the core monosaccharides ($GlcNAc_2$) interact with the polypep-

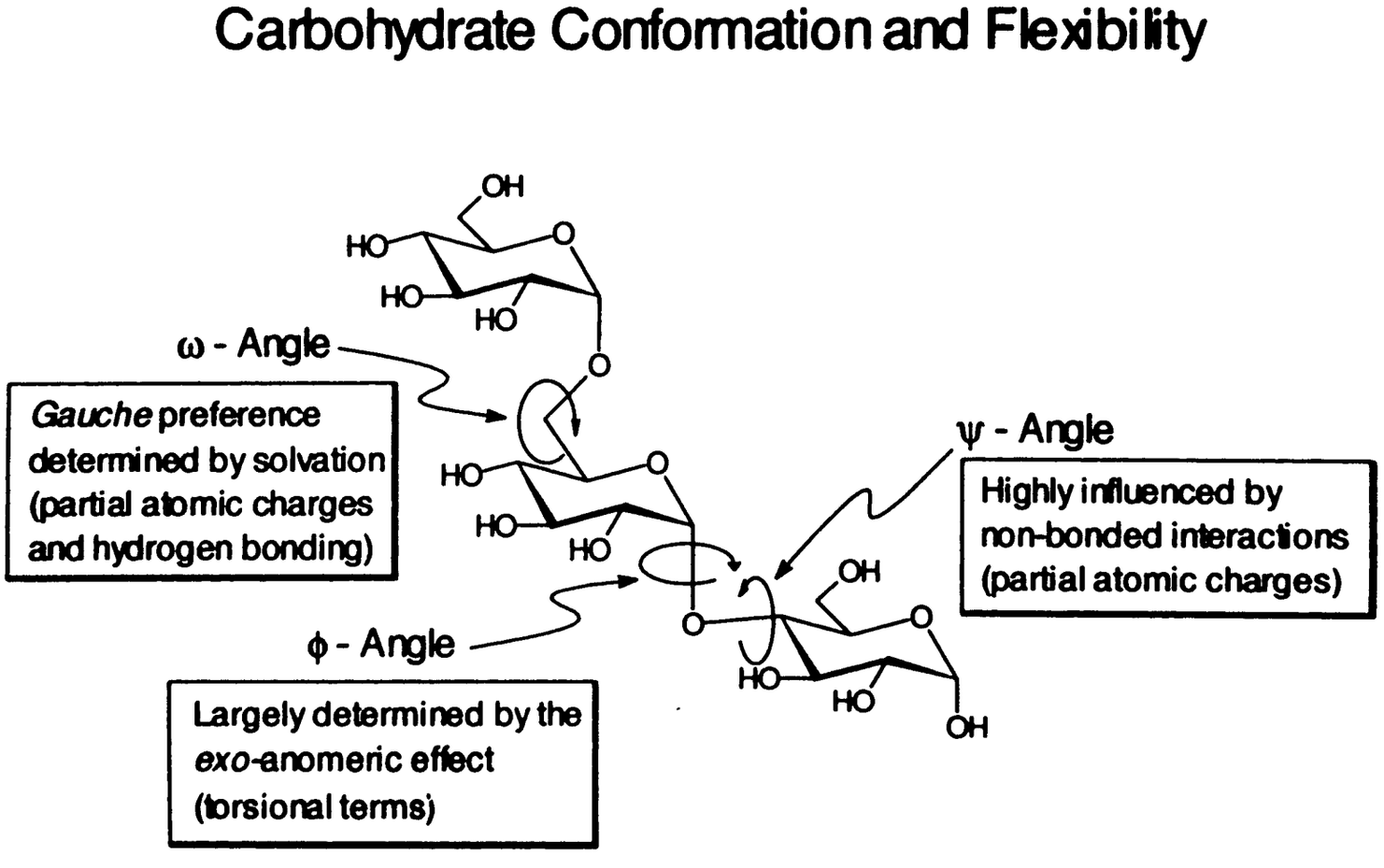


FIGURE 34. Definition of the angles of rotation involved in carbohydrate conformation and flexibility.

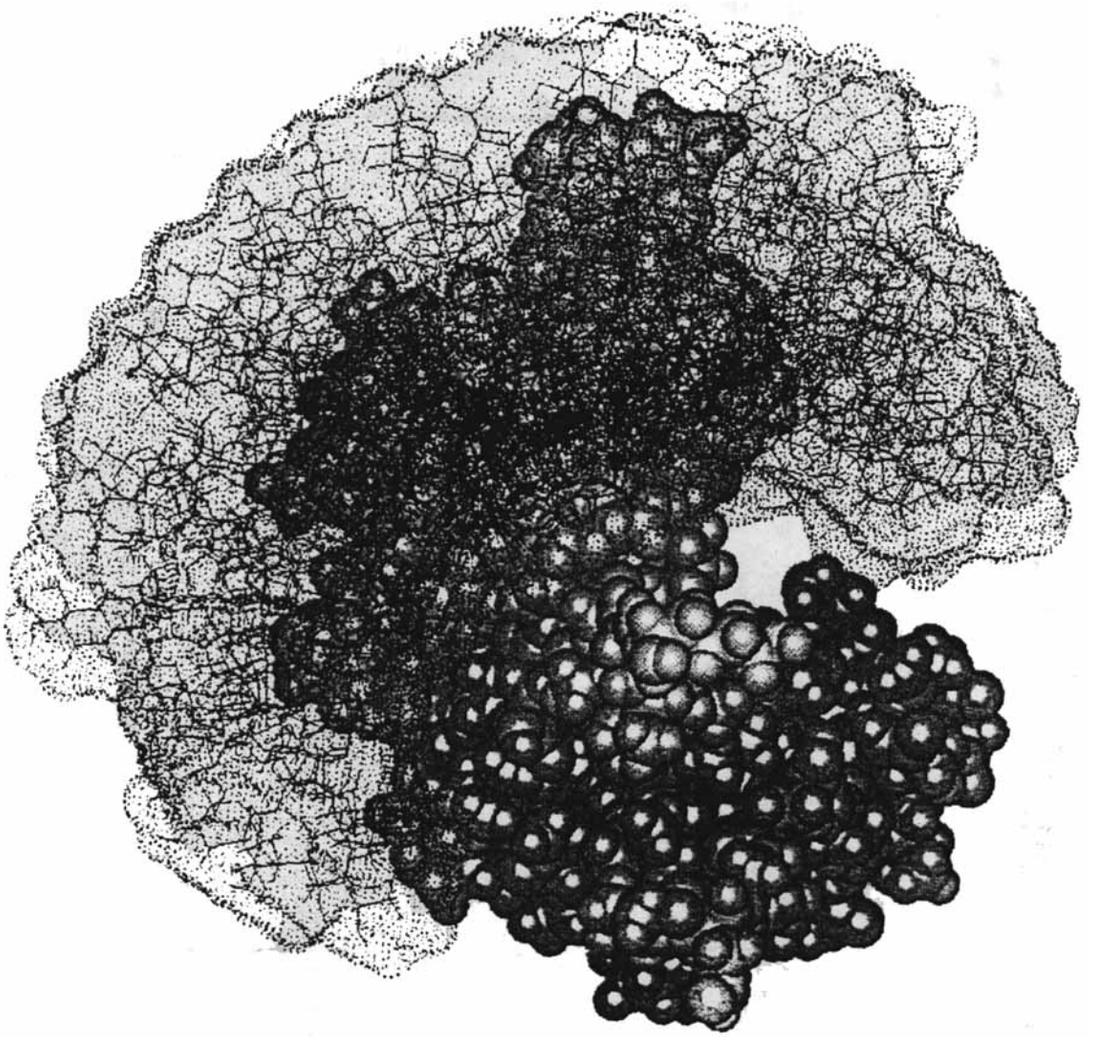


FIGURE 35. The Man9 glycoform of RNase B based on the 2.5 Å X-ray crystal structure with an overlay of 10 oligosaccharide conformations from a 750 ps MD trajectory of Man9 linked through Asn34. The side chain of Asn34 was maintained in the crystallographically determined orientation. In order to ensure a correct position for the reducing terminus, the oligosaccharides were overlaid on the first GlcNAc residue. All hydrogen atoms have been omitted for clarity and the total Van der Waals surfaces of the oligosaccharides are illustrated by dots.

tide, and, as a result, much of the remainder of the oligosaccharide is conformationally restricted. This is in contrast to the situation with RNase B, where the sugars have motion that is independent of the protein.

One function of the sugars in the CD2 domain appears to be to stabilize an exposed cluster of five positive charges from surface lysine residues, mainly through hydrogen bonds and van der Waals contacts (Wyss et al., 1995). If the central Lys61 is replaced by glutamic acid (as in rat CD2), the negative charge allows the formation of a salt bridge with Lys69, and rat CD2 is stable even when the sugar is removed. Interestingly, examination of the molecular model suggests that the sugars are located in a cleft in the protein. This feature, together

with the interactions of the chitobiose core with the protein, may hinder the access of GlcNAc transferase I to the developing glycan chain, thus providing an explanation for the site-specific glycosylation of CD2.

In the case of IgG, where protein-oligosaccharide interactions involve the outer-arm sugars, different glycoforms assume different orientations with respect to the protein. In the Fc region, the two conserved N-linked glycans occupy the space between the CH2 domains. Molecular modeling based on the crystal structure (Padlan, 1991) revealed significant differences in long-range interactions between the sugars and the CH2 region of the protein when different oligosaccharides were attached at Asn297. The crystal structure clearly showed extensive interactions between the 1,6-arm galactose of the biantennary complex sugar and the protein surface, indicating that this residue has a crucial role in maintaining the contacts between the sugar and the protein surface. Thus, glycoforms lacking terminal galactose were predicted to be more mobile than the full- length structures, thereby exposing part of the protein surface. Examination of the molecular modeling suggested that the oligosaccharide chains lacking galactose that terminate in GlcNAc may be available for recognition by lectins, such as MBP (Figure 26). Both of these predictions were confirmed experimentally (Wormald et al., submitted; Malhotra et al., 1995).

VI. CONCLUSION

Every cell contains a single biosynthetic glycosylation pathway through which all secretory proteins pass, the fine details of which may be adjusted in response to changes in the external or internal environment. Within the framework of the cell, individual glycoproteins can selectively modulate their interactions with the glycosylating enzymes by virtue of their unique primary, secondary, tertiary, and quaternary structures. Thus, the characteristic glycosylation profile of a single protein reflects the way in which that protein has interacted with the available glycosylation machinery. The common pathway allows each protein to select from many permutations of structures, and almost all glycoproteins emerge with a heterogeneous, but reproducible, mixture of glycans at each glycosylation site. It seems that there are two types of functions that sugars attached to glycoproteins can fulfill. The first can be achieved by a range of different sugars, whereas the second requires a specific glycan motif. The flexibility of the biosynthetic pathway is such that essential glycan recognition motifs can be conserved within a heterogeneous range of oligosaccharides, thus allowing the protein to retain the possibility of involving its sugars in either or both types of function. In some cases, the recognition of a glycan motif may be modulated by the 3D structure of both the protein and the remainder of the oligosaccharide.

Finally, it is clear that a full understanding of the implications of glycosylation for the structure and function of a glycoprotein can only be reached when a glycoprotein is viewed as a single entity and not only as the sum of its individual parts.

ACKNOWLEDGMENTS

We would like to acknowledge the valuable and interesting discussions that we have had with many people while writing this manuscript. In particular, we thank the following: Professors Richard Coico (IgD); Michael Ferguson (GPI anchor modifications); Royston Jefferis (IgG) and Ghislain Opdenakker (ribonuclease B, gelatinase B,

t-PA); Drs. A. Neil Barclay (leukocyte antigens), Terrence Butters (biosynthesis of oligosaccharides); A. Kristy Downing and Christopher Ponting (t-PA): Tim Elliott and Phil Wood (ER proteases); Elizabeth Hounsell (O-glycans); Taj Mattu (IgA); Anthony Merry (blood group sugars); Mark Wormald (molecular modeling, calnexin); Stefana Petrescu (tyrosinase); Brian Sutton (IgE); Glenys Tennant (SAP); Ms. Mercy Devashayim (Thy-1) and Mr. Anand Mehta (Hepatitis B virus).

We also thank Professor G. Opdenakker for critically reading the manuscript, Mr. Joshua Dwek for his valuable contributions to the molecular modeling, and Ms. Ruth Preston and Mr. John Freeman for their help with the figures.

REFERENCES

Alavi, A. and Axford, J. 1995. Beta 1,4-galactosyltransferase variations in rheumatoid arthritis. *Adv. Exp. Med. Biol.*, **376:** 185–92.

Allen, S., Naim, H. Y., and Bullied, N. J. 1995. Intracellular folding of tissue type plasminogen activator; effects of disulphide bond formation on N-linked glycosylation and secretion. *J. Biol. Chem.*, **270:** 4797–4804.

Amzel, L. M. and Poljak, R. J. 1979. Three-dimensional structure of immunoglobulins. *Annu. Rev. Biochem.*, **48:** 961–97.

Anderson, D. R., Atkinson, P. H., and Grimes, W. J. 1985. Major carbohydrate structures at five glycosylation sites on murine IgM determined by high- resolution 1H-NMR spectroscopy. *Arch. Biochem. Biophys.*, **243:** 605–618.

Arar, C., Carpentier,V., Le-Caer, J. P., Monsigny, M., Legrand, A., and Roche, A. C. 1995. ERGIC-53, a membrane protein of the endoplasmic reticulum-Golgi intermediate compartment, is identical to MR60, an intracellular mannose-specific lectin of myelomonocytic cells. *J. Biol. Chem.*, **270:** 3551–3.

Ashford, D. A., Alafi, C. D., Gamble, V. M., Mackay, D., Rademacher, T. W., Williams, P. J., Dwek, R. A., Barclay, A. N., Davis, S. J., Somoza, C., Ward, H. A. and Williams, A. F. 1993. Site-specific glycosylation of recombinant rat and human soluble CD4 variants expressed in Chinese hamster ovary cells. *J. Biol. Chem.*, **268:** 3260–3267.

Ashford, D. A., Dwek, R. A., Rademacher, T. W., Lis, H., and Sharon, N. 1991. The glycosylation of glycoprotein lectins. Intra- and inter-genus variation in N-linked oligosaccharide expression. *Carbohydr. Res.*, **213:** 215–27.

Baenziger, J. and Kornfeld, S. 1974. Structure of the carbohydrate units of IgA1 immunoglobulin. II. Structure of the O-glycosidically linked oligosaccharide units. *J. Biol. Chem.*, **249:** 7270–81.

Baenziger, J., Kornfeld, S., and Kochwa, S. 1974a. Structure of the carbohydrate units of IgE immunoglobulin. II. Sequence of the sialic acid-containing glycopeptides. *J. Biol. Chem.*, **249:** 1897–1903.

Baenziger, J., Kornfeld, S., and Kochwa, S. 1974b. Structure of the carbohydrate units of IgE immunoglobulin. I. Overall composition, glycopeptide isolation, and structure of the high mannose oligosaccharide unit. *J. Biol. Chem.*, **249:** 1889–1896.

Bause, E. 1983. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.*, **209:** 331–336.

Beintema, J. J., Weitzes, P., Weickman, J. L., and Glitz, D. G. 1984. The amino acid sequence of human pancreatic ribonuclease. *Anal. Biochem.*, **136:** 48–64.

Bergeron, J. J., Brenner, M. B., Thomas, D. Y., and Williams, D. B. 1994. Calnexin: a membrane-bound chaperone of the endoplasmic reticulum. *Trends Biochem. Sci.*, **19:** 124–8.

Bevilacqua, M. P. 1993. Endothelial-leukocyte adhesion molecules. *Ann. Rev. Immunol.*, **11:** 767–804.

Blanken, W. M. and Van den Eijnden, D. H. 1985. Biosynthesis of terminal Gal alpha 1-3Gal beta 1-4GlcNAc-R oligosaccharide sequences on glycoconjugates. Purification and acceptor specificity of a UDP-Gal:*N*-acetyllactosaminide alpha 1-3-galactosyltransferase from calf thymus. *J. Biol. Chem.*, **260:** 12927–34.

Bodian, D. L., Davis, S. J., Morgan, B. P., and Rushmere, N. K. 1997. Mutational analysis of the active site and antibody epitopes of the

complement-inhibitory glycoprotein, CD59. *J. Exp. Med.*, in press.

Bodian, D. L., Jones, E. Y., Harlos, K., Stuart, D. I., and Davis, S. J. 1994. Crystal structure of the extracellular region of the human cell adhesion molecule CD2 at 2.5A resolution. *Curr. Biol. Struct.*, **2:** 755–766.

Bole, D. G., Hendershot, L. M., and Kearney, J. F. 1986. Posttranslational association of immunoglobulin heavy chain binding protein with nascent heavy chains in nonsecreting and secreting hybridomas. *J. Cell. Biol.*, **102:** 1558–66.

Breuer, W. and Bause, E. 1995. Oligosaccharyl transferase is a constitutive component of an oligomeric protein complex from pig liver endoplasmic reticulum. *Eur. J. Biochem.*, **228:** 689–696.

Breukelman, H. J., Beintema, J. J., Confalone, E., Costanzo, C., Sasso, M. P., Carsana, A., Palmieri, M., and Furia, A. 1993. Sequences related to the ox pancreatic ribonuclease coding region in the genomic DNA of mammalian species. *J. Mol. Evol.*, **37:** 29–35.

Brown, D. A. and Rose, J. K. 1992. Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface.*Cell*, **68:** 533–544.

Bulleid, N. J., Bassell-Duby, R. S., Freedman, R. B., Sambrook, J. F., and Gething, M. J. 1992. Cell-free synthesis of enzymically active tissue-type plasminogen activator. Protein folding determines the extent of N-linked glycosylation. *Biochem. J.*, **286:** 275–280.

Cahour, A., Debeire, P., Hartmann, L., and Montreuil, J. 1983. Comparative study of the carbohydrate moieties of normal and pathological human IgM. *Biochem. J.*, **211:** 55–63.

Carson, D. D., Earles, B. J., and Lennarz, W. J. 1981. Enhancement of protein glycosylation in tissue slices by dolichol phosphate. *J. Biol. Chem.*, **256:** 11552–11557.

Cartron, J. P., Andreu, G., Cartron, J., Bird, G. W., Salmon, C., and Gerbal, A. 1978. Demonstration of T-transferase deficiency in Tn-polyagglutinable blood samples. *Eur. J. Biochem.*, **92:** 111–119.

Chan, A. N., Morris, H. R., Panico, M., Etienne, A. T., Rogers, M. E., and Gaffney, P., Creighton-Kempsford and Dell, A. 1991. A novel sialylated N-acetyl galactosamine-containing oligosaccharide is the major complex type structure present in Bowes melanoma tissue plasminogen activator. *Glycobiology*, **1:** 173–185.

Cygler, M., Rose, D. R., and Bundle, D. R. 1991. Recognition of a cell-surface oligosaccharide of pathogenic Salmonella by an antibody Fab Fragment. *Science*, **253:** 442–445.

Dahms, N. M. and Hart, G. W. 1986. Influence of quaternary structure on glycosylation. *J. Biol. Chem.*, **261:** 13186–13196.

Datta, A. K. and Paulson, J. C. 1995. The sialyltransferase 'sialyl motif' participates in binding the donor substrate CMP-NeuAc. *J. Biol. Chem.*, **270:** 1497–1500

Deeg, M. A., Humphrey, D. R., Yang, S. H., Ferguson, T. R., Reinhold, V. N., and Rosenberry, T. L. 1992. Glycan components in the glycoinositol phospholipid anchor of human erythrocyte acetylcholinesterase. Novel fragments produced by trifluoroacetic acid. *J. Biol. Chem,.* **267:** 18573–18580

De Graaf, T. W., Van der Stelt, M. E., Anbergen, M. G., and van Dijk, W. 1993. Inflammation-induced expression of Sialyl Lewis X containing glycan structures on α1-acid glycoprotein (orosomucoid) in human sera. *J. Exp. Med.*, **177:** 657–666.

Deisenhofer, J. 1981. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution. *Biochemistry*, **20:** 2361–2370.

De-Prisco, R., Sorrentino, S., Leone, E., and Libonati, M. 1984. A ribonuclease from human seminal plasma active on double-stranded RNA. *Biochim. Biophys. Acta*, **788:** 356–63.

Dorrington, K. J. and Bennich, H. H. 1978. Structure-function relationships in human immunoglobulin E. *Immunol. Rev.*, **41:** 3–25.

Dreyer, E. B., Leifer, D., Heng, J. E., McConnell, J. E., Gorla, M., Levin, L. A., Barnstable, C. J., and Lipton, S. A. 1995. An astrocytic binding site for neuronal Thy-1 and its effect on neurite outgrowth. *Proc. Natl. Acad. Sci. U.S.A.*, **92:** 11195–9.

Dulis, B. H., Kloppel, T. M., Grey, H. M., and Kubo, R. T. 1982. Regulation of catabolism of IgM heavy chains in a B lymphoma cell line. *J. Biol. Chem.*, **257:** 4369–4374.

Duncan, A. R., Woof, J. M., Partridge, L. J., Burton, D. R, and Winter, G. 1988. Localization of the binding site for the human high-affinity Fc receptor on IgG. *Nature*, **332:** 563–564.

Dustin, M. L., Ferguson, L. M., Chan, P. Y., Springer, T. A., and Golan, D. E. 1996. Visualization of CD2 interaction with LFA-3 and determination of the two-dimensional dissociation constant for adhesion receptors in a contact area. *J. Cell Biol.*, **132:** 456–474.

Dwek, R. A. 1996. Glycobiology: toward understanding the function of sugars. *Chem. Rev.*, **96:** 683–720.

Edge, A. S. B. and Spiro, R. G. 1985. Thyroid cell surface glycoproteins. Nature and disposition of carbohydrate units and evaluation of their blood group activity. *J. Biol. Chem.*, **260:** 15332–15338.

Elliott, T., Willis, A., Cerundolo, V., and Townsend, A. 1995. Processing of major histocompatibility class I-restricted antigens in the endoplasmic reticulum. *J. Exp. Med.*, **181:** 1481–1491.

Ellis, R. J. 1994. Roles of molecular chaperones in protein folding. *Curr. Opin. Struct. Biol.*, **4:** 117–122.

Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys, M. B., and Wood, S. P. 1994. Structure of pentameric human serum amyloid P component. *Nature (London)*, **367:** 338–345.

Esko, J. D. and Zhang, L. 1996. Influence of core protein sequence on glycosaminoglycan assembly. *Curr. Opin. Struct. Biol.*, **6:** 663–670.

Feinstein, A. and Beale, D. 1977. *Immunochemistry*. (Glynn, L. E. and Stewart, M. W., Eds.) New York: John Wiley & Sons.

Ferguson, M. A. J. 1991. Lipid anchors on membrane proteins. *Curr. Opin. Struct. Biol.*, **1:** 522–529.

Ferguson, M. A. J., Homans, S. W., Dwek, R. A., and Rademacher, T. W. 1988. Glycosylphosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane. *Science*, **239:** 753–759.

Field, M. C., Amatayakul-Chantler, S., Rademacher, T. W., Rudd, P. M., and Dwek, R. A. (1994) Structural analysis of the N-glycans from human immunoglobulin A1: comparison of normal human serum immunoglobulin A1 with that isolated from patients with rheumatoid arthritis. *Biochem. J.*, **299:** 261–275.

Fischer, P. B., Karlsson, G. B., Butters, T. D., Dwek, R. A., and Platt, F. M. 1996a. *N*-butyl deoxynojirimycin mediated inhibition of HIV entry correlates with changes in antibody recognition of the V1/V2 region of gp120. *J. Virol.*, **70:** 7143–7152.

Fischer, P. B., Karlsson, G. B., Dwek, R. A., and Platt, F. M. 1996b. *N*-butyl deoxynojirimycin-mediated inhibition of HIV entry correlates with impaired gp120 shedding and gp41 exposure. *J. Virol.*, **70:** 7153–7160.

Fletcher, C. M., Harrison, R. A., Lachman, P. J., and Neuhaus, D. 1994. Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. *Curr. Biol. Struct.*, **2:** 185–199.

Flynn, G. C., Pohl, J., Flocco, M. T., and Rothman, J. E. (1991) Peptide-binding specificity of the molecular chaperone BiP. *Nature*, **353:** 726–30.

Fukuda, M., Dell, A., Oates, J. E., and Fukuda, M. N. 1984. Structure of branched lactosaminoglycan, the carbohydrate moiety of band 3 isolated from human erythrocytes. *J. Biol. Chem.*, **259:** 8260–8273.

Fukuda, M. 1994. Cell surface carbohydrates: cell type-specific expression, in *Molecular Glycobiology*. (Fukuda, M. and Hindsgaul, O., Eds.) Oxford: IRL Press.

Fukuda, M. 1991. Lysosomal membrane glycoproteins. Structure, biosynthesis, and intracellular trafficking. *J. Biol. Chem.*, **266:** 21327–21342.

Furukawa, K., Matsuta, K., Takeuchi, F., Kosuge, E., Miyamoto, T., and Kobata, A. 1990. Kinetic study of a galactosyltransferase in the B cells of patients with rheumatoid arthritis. *Int. Immunol.*, **2:** 105–12.

Gavel, Y. and von-Heijne, G. 1990. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.*, **3:** 433–42.

Gerber, L. D., Kodukula, K., and Udenfriend, S. 1992. Phosphatidylinositol glycan (PI-G)-anchored membrane proteins. Amino acid requirements adjacent to the site of cleavage and PI-G attachment in the COOH-terminal signal peptide. *J. Biol. Chem.*, **267(17):** 12168–73.

Gerken, T. A., Butenhof, K. J., and Shogren, R. 1989. Effects of glycosylation on the conformation and dynamics of O-linked glycopro-

teins: carbon-13 NMR studies of ovine submaxillary mucin. *Biochemistry*, **28:** 5536–43.

Glitz, D. G., Kujawa, S., and Vosoghi, M. 1993. Immunogold location of human nonsecretory ribonuclease within cells and tissues. *Ribonucleases: Chemistry, Biology, Biotechnology*, 3rd Int. Meet., Capri, Italy, Abstr. L36.

Guile, G. G., Rudd, P. M., Wing, D. R., Prime, S. B., and Dwek, R. A. 1996. A rapid high-resolution method for separating oligosaccharide mixtures and analyzing sugarprints. *Anal. Biochem.*, **240:** 210–226.

Harpaz, H. and Schachter, H. 1980. Control of glycoprotein synthesis. Processing of asparagine linked oligosaccharides by one or more rat liver Golgi alpha-D-mannosidases dependent on the prior action of UDP-*N*-acetylglucosamine: alpha-D-mannoside beta 2-*N*-acetylglucosaminyltransferase I. *J. Biol. Chem.*, **25:** 4894–4902.

Heerman, K. H. and Gerlich, W. H. 1992. In: *Molecular Biology of Hepatitis B Virus*. (Machlachlan, A., Ed.) Boca Raton, FL: CRC Press.

Harris, R. J., Leonard, C. K., Guzzetta, A. W., and Spellman, M. W. 1991. Tissue plasminogen activator has an O-linked fucose attached to threonine-61 in the epidermal growth factor domain. *Biochemistry*, **30:** 2311–2314.

Harris, R. J. and Spellman, M. W. 1993. O-linked fucose and other posttranslational modifications unique to EGF modules. *Glycobiology*, **3:** 219–24.

Helenius, A. 1994. How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol. Biol. Cell.*, **5:** 253–265.

Hiraizumi, S., Takasaki, S., Ohuchi, N., Harada, Y., Nose, M., Mori, S., and Kobata, A. 1992. Altered glycosylation of membrane glycoproteins associated with human mammary carcinoma. *Jpn. J. Cancer Res.*, **83:** 1063–1072.

Hitoi, A., Yamashita, K., Niwata, Y., IrM., Kochibe, N., and Kobata, A. 1987. The carbohydrate moities of human urinary ribonuclease UL. Oligosaccharide microheterogeneity of the murine major histocompatibility antigens. *J. Biochem. Tokyo*, **101:** 29–41.

Homans, S. W., Ferguson, M. A., Dwek, R. A., Rademacher, T. W., Anand, R., and Williams, A. F. 1988. Complete structure of the glycosyl phosphatidylinositol membrane anchor of rat brain Thy-1 glycoprotein. *Nature (London)*, **333:** 269–272.

Hoppe, H. J. and Reid, K. B. 1994. Collectins — soluble proteins containing collagenous regions and lectin domains — and their roles in innate immunity. *Protein Sci.*, **3:** 1143–1158.

Hounsell, E. F., Davies, M. J., and Renouf, D. V. (1996) O-linked protein glycosylation structure and function. *Glycoconjugate J.*, **13:** 19–26.

Howard S. C., Wittwer, A. J., and Welply, J. K. 1991. Oligosaccharides at each site make structure dependent contributions to biological properties of human tissue plasminogen activator. *Glycobiology*, **1:** 411–417.

Hubbard, S. C. and Ivatt, R. J. 1981. Synthesis and processing of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.*, **50:** 555–584.

Hurtley, S. M. and Helenius, A. 1989. Protein oligomerization in the endoplasmic reticulum. *Annu. Rev. Cell Biol.*, **5:** 277–307.

Imperiali, B. and Rickert, K. W. 1995. Conformational implications of asparagine linked glycosylation. *Proc. Nat. Acad. Sci. U.S.A.*, **92:** 97–101.

Iourin, O., Mattu, T. S., Mian, N., Kier, G., Winchester, B., Dwek, R. A., and Rudd. P. M. 1996. The identification of abnormal glycoform populations of serum transferrin from carbohydrate deficient glycoprotein syndrome type I by capillary electrophoresis. *Glycoconjugate J.*, **13:** 1031–1042..

Itzkowitz, S. H., Bloom, E. J., Kokal, W. A., Modin, G., Hakamori, S., and Kim, Y. S. 1990. Sialosyl-Tn. A novel mucin antigen associated with prognosis in colorectal cancer patients. *Cancer*, **66:** 1960–1966.

Jaeken, J., Carchon, H., and Stibler, H. 1993. The carbohydrate-deficient glycoprotein syndromes: pre-Golgi and Golgi disorders. *Glycobiology*, **3:** 423–428.

Jaques, A. J., Opdenakker, G., Rademacher, T. W., Dwek, R. A., and Zamze, S. E. 1996. The glycosylation of Bowes melanoma tissue plasminogen activator: lectin mapping, reaction with anti-L2/HNK-1 antibodies and the presence of sulfated/glucuronic acid containing glycans. *Biochem. J.*, **316:** 427–437.

Joao, H. C., Scragg, I. G., and Dwek, R. A. 1992. Effect of glycosylation on protein conformation and amide proton exchange rates in RNase B. *FEBS Lett.*, **307:** 343–346.

Jolley, R. L., Evans, L. H., Makino, N., and Mason, H. S. 1974. Oxytyrosinase. *J. Biol. Chem.*, **249:** 335–345.

Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K., and Stuart, D. I. 1992. Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature (London)*, **360:** 232–238.

Kasturi, L., Eshleman, J. R., Wunner, W. H., and Shakin-Eshleman, S. H., 1995. The hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence N-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. *J. Biol. Chem.*, **270:** 14756–14761.

Kelly, R. B. 1990. Cell biology. Tracking an elusive receptor. *Nature*, **345:** 480–481.

Kerwin, J. L., Tuininga, A. R., and Ericsson, L. H. 1994. Identification of molecular species of glycerophospholipids and sphingomyelin using electrospray mass spectrometry. *J. Lipid Res.*, **35:** 1102–1114.

Killian, M. and Reinholdt, J. 1986. *Medical Microbiology*, pp. 173–208. (Easmon, C. S. F. and Jeljaszewics, J., Eds.) London: Academic Press.

Kieffer, B., Driscoll, P. C., Campbell, I. D., Willis, A. C., van der Merwe, P. A., and Davis, S. J. 1994. Three-dimensional structure of the extracellular region of the complement regulatory protein CD59, a new cell surface protein domain related to snake venom neurotoxins. *Biochemistry*, **33:** 4471–4482.

Kitagawa H. and Paulson J. C. 1994. Cloning of a novel α2,3 sialyl tranferase that sialylates glycoprotein and glycolipid carbohydrate groups. *J. Biol. Chem.*, **269:** 1394–401.

Klausner, R. D. and Sitia, R. 1990. Protein degradation in the endoplasmic reticulum. *Cell*, **62:** 611–614.

Kobata, A., Mizuochi, T., Endo, T., and Furukawa, K. 1989. Function and pathology of the sugar chains of human immunoglobulin G. *Ciba Found. Symp.*, **145:** 224–35.

Kornfeld, S. 1992. Structure and function of the mannose 6-phosphate/insulinlike growth factor II receptors. *Annu. Rev. Biochem.*, **61:** 307–330.

Kornfeld, R. and Kornfeld, S. 1985. Assembly of asparagine linked oligosaccharides. *Annu. Rev. Biochem.*, 631–644.

Kuhn, P., Guan, C., Cui, T., Tarentino, A. L., Plummer, T. H. Jr., and van Roey. 1995. Active site and oligosaccharide recognition residues of peptide *N*-4- (*N*-acetyl-β-D-glucosaminyl) asparagine amidase F. *J. Biol. Chem.*, **49:** 29493–29497.

Laskin, J. D. and Piccinini, L. A. 1986. Tyrosinase isozyme heterogeneity in differentiating B16/C3 melanoma. *J. Biol. Chem.*, **261:** 16226–35.

Lawrence, C. W., Little, P. A., Little, B. W., Miller, M. J., Bazel, S., and Alhadeff, J. A. 1993a. Human nonsecretory ribonucleases. I. Purification, peptide mapping and lectin blotting analysis of the kidney, liver and spleen enzymes. *Glycobiology*, **3:** 241–248.

Lawrence, C. W., Little, P. A., Little, B. W., Glushka, J., van Halbeek, H., and Alhadeff, J. A. 1993b. Human nonsecretory ribonucleases. II. Structural characterization of the N-glycans of the kidney, liver and spleen enzymes by NMR spectroscopy and electrospray mass spectrometry. *Glycobiology*, **3:** 249–259.

Leathem, A. J. and Brooks, S. A. 1987. Predictive value of lectin binding on breast cancer recurrence and survival. *Lancet*, **1:** 1054–1056.

Leatherbarrow, R. J. and Dwek, R. A. 1984. Binding of complement subcomponent C1q to mouse IgG1, IgG2a and IgG2b: a novel C1q binding assay. *Mol. Immunol.*, **21:** 321–327.

Lee, Y. C., Townsend, R. R., Hardy, M. R., Lonngren, J., Arnarp, J., Haraldsson, M., and Lonn, H. 1983. Binding of synthetic oligosaccharides to the hepatic Gal/GalNAc lectin. Dependence on fine structural features. *J. Biol. Chem.*, **258:** 199–202.

Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. J., Thomas, J. N., and Gregory, T. J. 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.*, **265:** 10373–82.

Letourneur, O., Sechi, S Willette-Brown, J., Robertson, M. W., and Kinet, J. P. 1995. Glycosylation of human truncated FcεRI chain is necessary for efficient folding in the endoplasmic reticulum. *J. Biol. Chem.*, **270:** 8249–8256.

Livi, G. P., Lillquist, J. S., Miles, L. M., Ferrara, A., Sathe, G. M., Simon, P. L., Meyers, C. A., Gorman, J. A., and Young, P. R. 1991. Secretion of N-glycosylated interleukin-1b in *Saccharomyces cerevisia* using a leader peptide from *Candida albicans*. Effect of N-glycosylation on biological activity. *J. Biol. Chem.*, **266:** 15348–15355.

Lippincott-Schwartz, J., Bonifacino, J. S., Yuan, L. C., and Klausner, R. D. 1988. Degradation from the endoplasmic reticulum: disposing of newly synthesized proteins. *Cell*, **54:** 209–20.

Lowe, J. B. 1991. Molecular cloning, expression, and uses of mammalian glycosyltransferases. *Semin. Cell. Biol.*, **2:** 289–307.

Lund, J., Takashi, N., Nakagawa, H., Goodall, M., Bentley, T., Hindley, S. A., Tyler, R., and Jefferis, R. 1993. Control of IgG/Fc glycosylation: a comparison of oligosaccharides from chimeric human/mouse and mouse subclass immunoglobulin Gs. *Mol. Immunol.*, **30:** 741–748.

Lund, J., Takahasi, N., Pound, J., Goodall, M., and Jefferis, R. 1996. Multiple interactions of IgG with its core oligosaccharide can modulate recognition by complement and human FcγR1, and influence the synthesis of its oligosaccharide chains. *J. Immunol.*, in press.

Malhotra, R., Wormald, M. R., Rudd, P. M., Fischer, P. B., Dwek, R. A., and Sim, R. B. 1995. Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose binding protein. *Nature Med.*, **1:** 237–241.

Masure, S., Nys, G., Fiter, P., Van Damme, J., and Opdenakker, G. 1993. Mouse gelatinase B. c-DNA cloning, regulation of expression and glycosylation in WEH1-3 macrophages and gene organization, *Eur. J. Biochem.*, **218:** 129–141.

Masure, S., Paemen, L., Van Aelst, I., Fiten, P., Proost, P., Van Damme, J., and Opdenakker, G. in press. Mouse gelatinase B. Production and characterisation of recombinant active enzyme from eukaryotic cells and *in vivo* effects after intravenous administration, *Eur. J. Biochem.*, in press.

Mattu, T. S. 1996. Oxford: Ph.D. dissertation.

Mayor, S., Menon, A. K., and Cross, G. A. 1991. Transfer of glycosyl-phosphatidylinositol membrane anchors to polypeptide acceptors in a cell-free system. *J. Cell Biol.*, **114(1):** 61–71.

McConville, M. J. and Ferguson, M. A. J. 1993. The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. *Biochem. J.*, **294:** 305–324.

McPherson, A., Brayer, G., Cascio, D. and Williams, R. 1986. The mechanism of binding of a polynucleotide chain to pancreatic ribonuclease. *Science*, **232:** 765–8.

Mehta, A., Lu, X., Block, T., Blumberg, B., and Dwek, R. 1996. Alterations in the unique M glycan prevent Hepatitis B virus formation — a potential therapeutic target. *Proc. Natl. Acad. Sci., U.S.A.*

Mellis, S. J. and Baenziger, J. U. 1983. Structures of the oligosaccharides present at the three asparagine-linked glycosylation sites of human IgD. *J. Biol. Chem.*, **258:** 11546–56.

Merkle, R. K. and Cummings, R. D. 1987. Relationship of the terminal sequences to the length of the poly-*N*-acetyllactosamine chains in asparagine-linked oligosaccharides from mouse lymphoma cell line BW5147. *J. Biol. Chem.*, **262:** 8179–8189.

Mestecky, J. and Kilian, M. 1985. Immunoglobulin A (IgA). *Methods Enzymol.*, **116:** 37–75.

Mir-Shekari, Y., Schultze, I. Harvey, D. J., Dwek, R. A., and Ashford, D. A. in press. A site-specific N-linked carbohydrate analysis of the hemagglutinin from influenza A virus receptor binding variants grown in a mammalian cell line. *J. Biol. Chem.*

Mizuta, K., Awazu, S., Yasuda, T., and Kishi, K. 1990. Purification and characterization of three ribonucleases from human kidney: comparison with urine ribonucleases. *Arch. Biochem. Biophys.*, **281:** 144–151.

Mononen, I. and Karjalainen, E. 1984. Structural comparison of protein sequences around potential N-glycosylation sites. *Biochem. Biophys. Acta*, **788:** 364–367.

Mori, K., Dwek, R. A., Downing, A. K., Opdenakker, G., and Rudd, P. M. 1995. The activation of type 1 and type 2 plasminogen by type I and type II tissue plasminogen activator. *J. Biol. Chem.*, **270:** 3261–3267.

Mulks, M. H. and Shoberg, R. J. 1994. Bacterial immunoglobulin A1 proteases. *Methods Enzymol.*, **235:** 543–554.

Muller, G., Ruppert, S., Schmid, E., and Schutz, G. 1988. Functional analysis of alternatively spliced tyrosinase gene transcripts. *EMBO J.*, **7:** 2723–2730.

Nakada, H., Inoue, M., Numata, Y., Tanaka, N., Funakoshi, I., Fukui, S., Mellors, A., and Yamashina, I. 1993. Epitopic structure of Tn glycophorin A for an anti-Tn antibody (MLS 128). *Proc. Natl. Acad. Sci. U.S.A.*, **90:** 2495–2499.

Nakagawa, H. K., Kawamura, Kato, K., Shamada, I., Arata, Y., and Takahashi, N. 1995. Identification of neural and sialyl N-linked oligosaccharide structures from human serum glycoproteins using three kinds of high-performance liquid chromatography. *Anal. Biochem.*, **226:** 130–138.

Narasimhan, S., Freed, J. C., and Schachter, H. 1985. Control of glycoprotein synthesis. Bovine milk UDP galactose: *N*-acetylglucosamineβ-4-galactosyltranferase catalyses the preferential transfer of galactose to the GlcNAcβ1,2Manα1,3-branch of both bisected and nonbisected complex biantennary asparagine-linked oligosaccharides. *Biochemistry*, **24:** 1694–1700.

Nath, D., van der Merwe, Kelm, S., Bradfield, P., and Crocker, P. R. 1995. The amino-terminal immunoglobulin-like domain of sialoadhesin contains the sialic acid binding site. *J. Biol. Chem.*, **270:** 26184–26191.

Natsuka, S. and Lowe, J. B. 1994. Enzymes involved in mammalian biosynthesis. *Curr. Opin. Struct. Biol.*, **4:** 683–691.

Nehre, K., Hagen, F. K., and Tabak, L. A. 1996. Charge distribution of flanking amino acids influences O-glycan acquisition *in vivo*. *J. Biol. Chem.*, **271:** 7061–7065.

Nilsson, I. M. and von Heijne, G. 1993. Determinations of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem.*, **268:** 5798–5801.

Ninomiya, H., Stewart, B. H., Rollins, S. A., Zhao, J., Bothwell, A. L. M., and Sims, P. J. 1992. Contribution of the N-linked carbohydrate of erythrocyte antigen CD59 to its complement inhibitory activity. *J. Biol. Chem.*, **267:** 8404–8410.

Nishimura, H., Takao, T., Hase, S., Shimonishi, Y., and Iwanaga, S. 1992. Human factor IX has a tetrasaccharide O-glycosidically linked to serine 61 through the fucose residue. *J. Biol. Chem.*, **267:** 17520–5.

Norris, G. E., Stillman, T. J., Anderson, B. F., and Baker, E. N. 1994. The three-dimensional structure of PNGase F, a glycosyl-asparaginase from *Flavobacterium meningosepticum*. *Curr. Biol. Struct.*, **2:** 1049–1059.

Nose, M. and Wigzell, H. 1983. Biological significance of carbohydrate chains on monoclonal antibodies. *Proc. Nat. Acad. Sci. U.S.A.*, **80:** 6632–6636.

Otteken, A. and Moss, B. 1996. Calreticulin interacts with newly synthesized human immunodeficiency virus type 1 envelope glycoprotein, suggesting a chaperone function similar to that of calnexin. *J. Biol. Chem.*, **271:** 97–103.

Ou, W. J., Cameron, P. H., Thomas, D. Y., and Bergeron, J. M. M. 1993. Assocation of folding intermediates with calnexin during protein maturation. *Nature (London)*, **364:** 771–776.

Padlan, E.A. 1994. Anatomy of the antibody molecule. *Mol-Immunol.*, **31:** 169-217.

Padlan, E. 1991. Biological significance of carbohydrate chains on monoclonal antibodies. *Proc. Natl. Acad. Sci. U.S.A.*, **80:** 6632–6636.

Parekh, R. B., Dwek, R. A., Sutton, B. J., Fernandes, D. L., Leung, A., Stanworth, D., Rademacher, T. W., Mizuochi, T., Taniguchi, T., Matsuta, K., Takeuchi, F., Nagano, Y., Miyamoto, T., and Kobata, A. 1985. Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature (London)*, **316:** 452–457.

Parekh, R. B. 1985. Ph.D. dissertation, Oxford.

Parekh, R. B., Tse, A. G. C., Dwek, R. A., Williams, A. F., and Rademacher, T. W. 1987. Tissue-specific N-glycosylation, site specific oligosaccharide patterns and lentil lectin recognition of rat Thy-1. *EMBO J.*, **6:** 1233–1244.

Parekh, R. B., Dwek, R. A., Thomas, J. R., Opdenakker, G., Rademacher, T. W., Wittwer, A. W., Howard, S. H., Nelson, R., Siegel, N., Jennings, M. G., Harakas, N. K., and Feder, J. 1989a. Cell-type specific and site specific N-glycosylation of type I and type II human tissue plasminogen activator. *Biochemistry*, **28:** 7644–7661.

Parekh, R. B., Dwek, R. A., Rudd, P. M., Thomas, J. R., Rademacher, T. W., Warren, T., Wun,

T. C., Hebert, B., Reitz, B., Palmier, M., Ramabhadran, T., and Tiemeir, D. C. 1989b. N-glycosylation and *in vitro* enzymatic activity of human recombinant tissue plasminogen activator expressed in chinese hamster ovary cells and a murine cell line. *Biochemistry*, **28:** 7670–7679.

Pepys, M. B., Rademacher, T. W., Amatayakul-Chantler, Williams, P., Noble, G. E., Hutchinson, W. L., Hawkins, P. N., Nelson, S. R., Gallimore, J. R., Herbert, J., Hutton, T., and Dwek, R. A. 1994. Human serum amyloid amyloid P component is an invariant constituent of amyloid deposits and has a uniquely homogeneous glycostructure. *Proc. Natl. Acad. Sci. U.S.A.*, **91:** 5602–5606.

Perkins, S. J., Johnson, L.N., Phillips, D. C., and Dwek, R. A. 1981. The simultaneous binding of lanthanide and *N*-acetylglucosamine inhibitors to hen egg-white lysozyme in solution by 1H and 13C nuclear magnetic resonance. *Biochem. J.*, **193:** 573–88.

Perez-Vilar, J., Hidalgo, J., and Velasco, A. 1991. Presence of terminal N-acetylgalactosamine residues in subregions of the endoplasmic reticulum is influenced by cell differentiation in culture. *J. Biol. Chem.*, **266:** 23967–23976.

Petrescu, S. M., Petrescu, A. J., Dwek, R. A., and Platt, F. M. in press. Inhibition of N-glycans processing in B16 melanoma cells results in inactivation of tyrosine but does not prevent its transport to the melanosome. *J. Biol. Chem.*

Petrescu, A., Butters, T. B., Petrescu, S., Reinkensmeir, G., Dwek, R. A., Platt, F. M., and Wormald, M. R., Submitted.

Plaut, A. G. 1983. The IgA1 proteases of pathogenic bacteria. *Annu. Rev. Microbiol.*, **37:** 603–22.

Pohl, G., Kallstrom, M., Bergsdorf, N., Wallen, P., and Jornvall, H. 1984. Tissue plasminogen activator: peptide analyses confirm an indirectly derived amino acid sequence, identify the active site serine residue, establish glycosylation sites, and localize variant differences. *Biochemistry*, **23:** 3701–3707.

Pollack, L. and Atkinson, P. H. 1983. Correlation of glycosylation forms with position in amino acid sequence. *J. Cell Biol.*, **97:** 293–300.

Rabouille, C. and Spiro, R. G. 1992. Nonselective utilization of the endomannosidase pathway for processing glycoproteins by human hepatoma (HepG2) cells. *J. Biol. Chem.*, **267:** 11573–11578.

Rademacher, T. W., Homans, S. W., Parekh, R., and Dwek, R. A. 1985. Immunoglobulin G as a glycoprotein. *Biochem. Soc. Symp.*, **51:** 131–148.

Recny, M. A., Luther, M. A., Knoppers, M. H., Neinhardt, E. A., Khandekar, S. S., Concino, M. F., Schimke, P. A., Francis, M. A., Moebius, U., Reinhold, B., Reinhold, V. N., and Reinherz, E. L. 1992. N-glycosylation is required for human CD2 adhesion functions. *J. Biol. Chem.*, **267(31):** 22428–22434.

Ribo, M., Beintema, J. L., Ossett, M., Fernandez, E., Bravo, J., De Llorens, R., and Cuchillo, C. M. 1994. Heterogeneity in the glycosylation pattern of human pancreatic ribonuclease. *Biol. Chem. Hoppe Seyler*, **375:** 357–363.

Rice, K. G., Weisz, O. A., Barthel, T., Lee, R. T., and Lee, Y. C. 1990. Defined geometry of binding between triantennary glycopeptide and the asialoglycoprotein receptor of rat heptocytes. *J. Biol. Chem.*, 265: 18429–18434.

Roitt, I. M., Brostoff, J., and Male, D. 1985. *Immunology*, chap. 5, New York: Gower.

Roth, J., Taatjes, D. J., Weinstein, J., Paulson, J. C., Greenwell, P., and Watkins, W. M. 1986. Differential subcompartmentation of terminal glycosylation in the Golgi apparatus of intestinal absorptive and goblet cells. *J. Biol. Chem.*, **261:** 14307–14312.

Rudd, P. M., Leatherbarrow, R. J., Rademacher, T. W., and Dwek, R. A. 1991. Diversification of the IgG molecule by oligosaccharides. *Mol. Immunol.*, **28:** 1369–1378.

Rudd, P. M., Scragg, I. G., Coghill, E., and Dwek, R. A. 1992. Separation and analysis of the glycoform populations of ribonuclease B using capillary electrophoresis. *Glycoconj. J.*, **9:** 86–91.

Rudd, P. M., Joao, H. C., Coghill, E., Fiten, P., Saunders, M. R., Opdenakker, G., and Dwek, R. A. 1994a. Glycoforms modify the dynamic stability and functional activity of an enzyme. *Biochemistry*, **33:** 17–22.

Rudd, P. M., Fortune, F. M. Patel, T., Parekh, R. B., Dwek, R. A., and Lehner, T. 1994b. IgA binding to T-cell surface receptor involves "O"-linked sugars from the hinge region of IgA1. *Immunology*, **83:** 99–106.

Rudd, P. M. 1995. Ph.D. dissertation. Open University.

Rudd, P. M., Woods, R. J., Wormald, M. W., Opdenakker, G., Downing, A. K., Campbell, I. D., and Dwek, R. A. 1995. The effects of variable glycosylation on the functional activities of ribonuclease, plasminogen and tissue type plasminogen activator. *Biochem. Biophys. Acta,* **1248:** 1–10.

Rudd, P. M., Morgan, B. P., Wormald, M. R., Harvey, D. J., van der Berg, C. W., Davis, S. J., Ferguson, M. A. J., and Dwek, R. A. in press. The glycosylation of the complement regulatory protein, human erythrocyte CD59. *J. Biol. Chem.*, in press.

Rudd, P. M., Barklay, A. N., Brown, M. H., Davies, S. J., and Dwek, R. A. 1996a. In preparation.

Rudd, P. M., Brown, M. H., Barclay, A. N., and Dwek, R. A. 1996b. In preparation.

Sato, S. and Hughes, R. C. 1992. Binding specificity of a baby hamster kidney lectin for H type I and II chains, polylactosamine glycans, and appropriately glycosylated forms of laminin and fibronectin. *J. Biol. Chem.*, **267:** 6983–90.

Schachter, H. 1986. Biosynthetic controls that determine the branching and microheterogeneity of protein-bound oligosaccharides. *Biochem. Cell. Biol.*, **64:** 163–181.

Scheiffele, P., Peranen, J., and Simons, K. 1995. N-glycans as apical sorting signals in epithelial cells. *Nature*, **378:** 96–98.

Shackelford, D. A. and Strominger, J. L. 1983. Analysis of the oligosaccharides on the HLA-DR and DC1 B cell antigens. *J. Immunol.*, **130:** 274–282.

Shakin-Eshleman, S. E., Spitalnik, S. L., and Kasturi, L. 1996. The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. *J. Biol. Chem.*, **271:** 6363–6366.

Sharon, N. and Lis, H. *Glycoproteins: Structure and Function in Methods of Glycoconjugate Analysis,* (S. U. H.-J., Gabius, Ed.) Weinheim: Chapman and Hall, in press.

Shogren, R., Gerken, T. A., and Jentoft, N. 1989. Role of glycosylation on the conformation and chain dimensions of O-linked glycoproteins: light-scattering studies of ovine submaxillary mucin. *Biochemistry,* **28:** 5525–36.

Singh, I., Doms, R.W., Wagner, K. R., and Helenius, A. 1990. Intracellular transport of soluble and membrane bound glycoproteins: folding, assembly and secretion of anchor-free influenza hemagglutinin. *EMBO J.*, **9:** 631–639.

Smith, B. O., Downing, A. K., Driscoll, P. C., Dudgeon, T. J., and Campbell, I. D. 1995. The solution structure and backbone dynamics of the fibronectin type-1 and epidermal growth factor-like pair of modules of tissue-type plasminogen-activator. *Structure,* **3:** 3823–3833.

Spiro, R. G., Zhu, Q., Bhoyroo, V., and Soling, H. D. 1996. Definition of the lectin-like properties of the molecular chaperone, calreticulin, and demonstration of its co-purification with endomannosidase from rat liver Golgi. *J. Biol. Chem.*, **271:** 11588–11594.

Strecher, G. and Montreuil, J. 1979. Glycoproteines et glycoproteinoses. *Biochemie,* **61:** 1199–1246.

Suzuki, T., Kitajim, K., Inoue, Y., and Inoue, S. 1995. Carbohydrate-binding property of peptide: N-glycanase from mouse fibroblast L-929 cells as evaluated by inhibition and binding experiments using various oligosaccharides. *J. Biol. Chem.*, **270:** 15181–15186.

Suzuki, T., Kitajima, K., Inoue, S., and Inoue, Y. 1994. Does an animal peptide: N-glycanase have the dual role as an enzyme and a carbohydrate-binding protein? *Glycoconj. J.*, **11:** 469–76.

Swiedler, S. J., Freed, J. H., Tarentino, A. L., Plummer, T. H., Jr., and Hart, G. W. 1985. Oligosaccharide microheterogeneity of the murine major histocompatibility antigens. Reproducible site-specific patterns of sialylation and branching in asparagine-linked oligosaccharides. *J. Biol. Chem.*, **260:** 4046–4054.

Takahashi, N., Nakagawa, H., Fujikawa, K., and Tomiya, N. 1996. Three-dimensional elution mapping of pyridylaminated N-linked neutral and sialyl oligosaccharides. *Anal. Biochem.*, **226:** 139–146.

Tatu, U., Hammond, C., and Helenius, A. 1995. Folding and oligomerization of influenza hemagglutinin in the ER and the intermediate compartment. *EMBO J.*, **14:** 1340–1348.

Treumann, A., Lifely, M. R., Schneider, P., and Ferguson, M. A. J. 1995. Primary structure of CD52. *J. Biol. Chem.*, **270:** 6088–6099.

Tsuchiya, N., Endo, T., Matsuta, K., Yoshinoya, S., Aikawa, T., Kosuge, E., Takeuchi, F.,

Miyamoto, T., and Kobata, A. 1989. Effects of galactose depletion from oligosaccharide chains on immunological activities of human IgG. *J. Rheumatol.*, **16:** 285–290.

Tsuchiya, N., Endo, T., Matsuta, K., Yoshinoya, S., Takeuchi, F., Nagano, Y., Shiota, M., Furukawa, K., Kochibe, N., Ito, K., et al. 1993. Detection of glycosylation abnormality in rheumatoid IgG using *N*-acetylglucosamine-specific *Psathyrella velutina* lectin. *J. Immunol.*, **151:** 1137–46.

van der Merwe, P. A., McNamee, P. N., Davies, E. A., Barclay, A. N., and Davis, S. J. 1995. Topology of the CD2-CD48 cell-adhesion molecule complex: implications for antigen recognition by T cells. *Curr. Biol.*, **5:** 74–84.

Van Schaftingen, E. and Jaeken, J. 1995. Phosphomannomutase deficiency is a cause of carbohydrate-deficient glycoprotein syndrome type I. *FEBS Lett.*, **337:** 318–320.

Vassilakos, A., Cohen-Doyle, M. F., Peterson, P. A., Jackson, M. R., and Williams, D. B. 1996. The molecular chaperone calnexin facilitates folding and assembly of class I histocompatibility molecules. *EMBO J.*, **15:** 1495–506.

Wada, I., Rindress, D., Cameron, P., Ou, W. J., Doherty, J. J., II, Louvard, D., Bell, A. W., Dignard, D., Thomas, Q. Y., and Bergeron, J. J. M. 1991. SSR alpha and associated calnexin are major calcium binding proteins of the endoplasmic reticulum membrane. *J. Biol. Chem.*, **266:** 19599–19610.

Wada, Y., Nishikawa, A., Okamoto, N., Inui, K., Tsukamoto, H., Okada, S., and Taniguchi, N. 1992. Structure of serum transferrin in carbohydrate-deficient glycoprotein syndrome. *Biochem Biophys Res Commun.*, **189:** 832–6.

Wang, J., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L., and Harrison, S. C. 1990. Atomic structure of a fragment of CD4 containing two immunoglobulin-like domains. *Nature (London)*, **348:** 411–418.

Wang, W-C., Lee, N., Aoki, D., Fukuda, M. N., and Fukuda, M. 1995. The poly-*N*-lactosamines attached to lysosomal membrane glycoproteins are increased by the prolonged association with the Golgi complex. *J. Biol. Chem.*, **266:** 23185–23190.

Ware, F. E., Vassilakos, A., Peterson, P. A., Jackson, M. R., Lehrman, M. A., and Williams, D. 1995. The molecular chaperone calnexin binds Glc1Man9GlcNAc2 oligosaccharide as an initial step in recognizing unfolded glycoproteins. *J. Biol. Chem.*, **270:** 4697–4704.

Weis, W. I. and Drickamer, K. 1996. Structural basis of lectin-carbohydrate recognition. *Ann. Rev. Biochem.*, **65:** 441–473.

Watson, M., Rudd, P. M., Bland, M., Dwek, R. A., and Axford, J. S. In preparation.

Weis, W. I. and Drickamer, K. 1994. Trimeric structure of a C-type mannose-binding protein. *Structure*, **2:** 1227–1240.

Weinstein, J., de-Souza-e-Silva, U., and Paulson, J. C. 1982. Sialylation of glycoprotein oligosaccharides N-linked to asparagine. Enzymatic characterization of a Gal beta 1 to 3(4)GlcNAc alpha 2 to 3 sialyltransferase and a Gal beta 1 to 4GlcNAc alpha 2 to 6 sialyltransferase from rat liver. *J. Biol. Chem.*, **257:** 13845–53.

Wilhelm, L., Lee, S. G., Kaylan, N. K., Cheng, S. M., Wiener, F., Pierchala, W., and Hung, P. P. 1990. Alterations in the domain structure of tissue type plasminogen activator change the nature of asparagine glycosylation. *Biotechnology*, **8:** 321–325.

Williams, A. F. 1987. A year in the life of the immunoglobulin superfamily. *Immunol. Today*, **8:** 298–303.

Williams, A. F. and Barclay, A. N. 1988. The Immunoglobulin superfamily — domains for cell surface recognition. *Ann. Rev. Immunol.*, **6:** 381–405.

Williams, A. F., Parekh, R. B., Wing, D. R., Willis, A. C., Barclay, A. N., Dalchau, R., Fabre, J. W., Dwek, R. A., and Rademacher, T. W. 1993. Comparative analysis of the N-glycans of rat, mouse and human Thy-1. Site specific oligosaccharide patterns of neural Thy-1, a member of the immunoglobulin superfamily. *Glycobiology*, **3:** 339–348.

Williams, A. F., Davis, S. J., He, Q., and Barclay, A. N. 1989. *Cold Spring Harbour Symp., Quant. Biol.*, **56:** 637–647.

Williams, D. B. and Lennarz, W. J. 1984. Control of asparagine-linked oligosaccharide chain processing: studies on bovine pancreatic ribonuclease B. *J. Biol. Chem.*, **259:** 5105–5114.

Williams, R. L., Greene, S. M., and McPherson, A. 1987. The crystal structure of ribonuclease B at 2.5 Å resolution. *J. Biol. Chem.*, **263:** 16020–16031.

Wilson, I. A., Skehel J. J., and Wiley, D. C. 1989. Structure of the hemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature (London)*, **289:** 366–373.

Woods, R. J. 1995. Three-dimensional structures of oligosaccharides. *Curr. Opin. Struct. Biol.*, **5:** 591–598.

Woods, R. J., Wormald, M. R., Edge, C. J., and Dwek, R. A. In preparation.

Wormald, M. R., Rudd, P. M., Harvey, D. J., Chang, S-C., Scragg, I., and Dwek, R. A. in press. Levels of terminal galactose modulate the dynamic motion of IgGFc oligosaccharides. *Biochemistry*.

Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanandam, A. R., Smolyar, A., Reinherz, E. L., and Wagner, G. 1995. Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science*, **269:** 1273–1278.

Yamashita, K., Ideo, H., Ohkura, T., Fukushima, K., Yusa, I., Ohno, K., and Takeshita, K. 1993. Sugar chains of serum hTf from patients with carbohydrate deficient glycoprotein syndrome. *J. Biol. Chem.*, **268:** 5783–5789.

Yamashita, K., Hitoi, A., Irie, M., and Kobata, A. 1986. Fractionation by lectin affinity chromatography indicates that the glycosylation of most ribonucleases in human viscera and body fluids is organ specific. *Arch. Biochem. Biophys.*, **50:** 263–266.

Young, R. J., Owens, R. J., Mackay, G. A., Chan, C. M., Shi, J., Hide, M., Francis, D. M., Henry, A. J., Sutton, B. J., and Gould, H. J. 1995. Secretion of recombinant human IgE-Fc by mammalian cells and biological activity of glycosylation site mutants. *Protein Eng.*, **8:** 193–199.

Youings, A., Chang, S-C., Dwek, R. A., and Scragg, I. G. 1996. Site-specific glycosylation changes on human immunoglobulin G in pregnancy and rheumatoid arthritis. *Biochem. J.*, **314:** 621–630.

Zapun, A., Petrescu, S. M., Rudd, P. M., Dwek, R. A., Thomas, D., and Bergeron, J. J. M. 1997. Conformation-independent binding of monoglucosylated ribonuclease B to calnexin. *Cell*, **88:** 29–38.

ADDED IN PROOF

3832–3846 Reprinted from The Journal of Physical Chemistry, 1995, 99.

Molecular Mechanical and Molecular Dynamical Simulations of Glyproteins and Oligosaccharides. 1. GLYCAM_93 Parameter Development

Robert J. Woods, Raymond A. Dwek, and Christopher J. Edge

Glycobiology Institute, Department of Biochemistry, The University of Oxford, South Parks Road, Oxford, OX1 3QU, U.K.

Beri Fraser-Reid

The Department of Chemistry, Paul M. Gross Chemical Laboratory, Duke University, North Carolina 27706

N. Sharon and H. Lis

Glycoproteins: Structure and function methods of glycoconjugate analysis, in Glycosciences, Oabius, S., Ed., Chapman & Hall, Weinheim, 1997