# Glycobiology: Toward Understanding the Function of Sugars

Raymond A. Dwek

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

*Received July 20, 1995 (Revised Manuscript Received October 28, 1995)*

Downloaded via UNIV OF MANITOBA on October 25, 2024 at 11:23:47 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

## Contents

1. Introduction and Background 683
1.1. Biological Macromolecules 683
1.2. Glycoproteins 684
1.3. Glycobiology 684
1.4. Technology Developments 684
1.5. There Is No Single Unifying Function for Oligosaccharides 684
2. What Does a Typical Glycoprotein Look like? 685
2.1. Implication of the Conformations and Dynamics of Protein Surface Oligosaccharides in Protein Function 685
2.2. Molecular Dynamics (MD) Simulation of $Man_9GlcNAc_2OH$ 686
2.3. The Core Conformation of the Oligosaccharide in N-Linked Glycoproteins Is Independent of the Protein 686
2.4. Outer Arm Conformation of the Oligosaccharide 687
2.5. Dynamic Sugar Model of RNase B 687
2.6. Consequences of Protein Side-Chain Flexibility on the Presentation of the Oligosaccharide 687
2.7. Biological Implications 688
3. Some Factors Which Control Protein Glycosylation 688
3.1. The Primary Peptide Structure Determines the Number and Location of Potential Glycosylation Sites 688
3.2. Structure and Diversity of N-Linked Glycans 689
3.3. Structure of O-Linked Glycans 689
3.4. Cell Type Influences Glycosylation 689
3.5. The 3D Structure of the Protein Influences the Extent and Type of Glycosylation 690
3.6. Generation of Glycoforms 691
4. Oligosaccharide Technology 692
4.1. Release of Glycans from Glycoprotein 692
4.2. Labeling of the Released Glycans To Enable Their Detection in Subsequent Procedures 693
4.3. Profiling the Pool Glycans To Determine the Types of Glycans Present and Their Relative Molar Properties 693
4.4. Structural Analysis 694
5. Characteristics of Protein Glycosylation 695
5.1. Importance of the Overall Protein Conformation in Determining Glycosylation 695
5.2. Effect of Local Protein Conformation—Glycosylation Shows Site Specificity 695
5.3. Glycosylation Is Protein-Specific, Site-Specific, and Tissue/Cell-Specific 697
6. Glycosylation Site Occupancy Can Modulate Enzyme Activities 698
6.1. The Multimolecular Interaction of tPA with Plasminogen and Fibrin Is Modulated by Glycosylation 699
6.2. Variable Glycosylation Site Occupancy on Carbohydrate-Deficient Glycoprotein Syndrome (CDGS) 700
7. Some Structural Roles for Oligosaccharides 701
7.1. Glycosyl-Phosphatidylinositol (GPI) Anchors 701
7.2. Structure/Function Relationships in IgG 702
8. Oligosaccharide Recognition 703
8.1. Specific Interactions with Animal Lectins 703
8.2. Neural Glycosylation and Recognition 705
8.3. Major Histocompatibility Complex (MHC) Restricted Recognition of Glycopeptides by T-Cells 706
8.4. Recognition of Oligosaccharides by Stimulated T-Cells 707
9. Glycosylation in Disease 707
9.1. The IgG Molecule 707
9.1.1. Site-Specific Glycosylation of IgG 707
9.1.2. IgG Glycoforms Associated with Rheumatoid Arthritis 708
9.1.3. Glycosylation Changes on the IgG Molecule Are "Disease Restricted" and Are an Important Factor in Rheumatoid Arthritis 708
9.1.4. Structural Changes in IgG Fc on Loss of Galactose 709
9.1.5. Functional Implications of IgG Glycoforms 710
9.1.6. Modeling of the Possible Interaction between Agalactosyl IgG Fc and MBP 711
9.1.7. $Ca^{2+}$-Dependent Binding of MBP to IgG Is Mediated by the Agalactosyl Fc Glycoforms 712
9.1.8. MBP Activation of Complement by Agalactosyl IgG Glycoforms 712
9.1.9. MBP and Agalactosyl IgG Are Present in Synovial Fluid 713
10. Glycosylation Inhibitors as Antiviral Agents 713
10.1. Glycosphingolipids 715
10.2. Glycosphingolipid Storage Disorders 717
11. Concluding Remarks 717
12. Acknowledgments 718
13. References 718

## 1. Introduction and Background

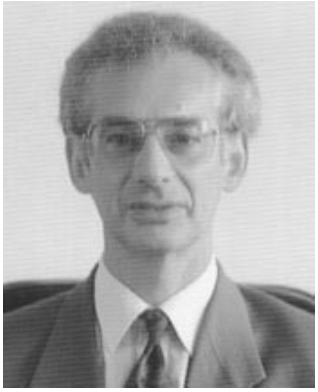
### 1.1. Biological Macromolecules

Four major classes of macromolecules in biology are DNA, proteins, carbohydrates, and lipids. Carbohydrates differ from the other two classes of biological polymers in two important characteristics: they can be highly branched molecules, and their monomeric

0009-2665/96/0796-0683$25.00/0 © 1996 American Chemical Society



FORMYCON-1086



Raymond A. Dwek *b* 1941, M.A., M.Sc., D.Phil., D.Sc., C.Chem. F.R.S.C. Research lecturer in Physical Chemistry, Christ Church, Oxford 1966–68, in Biochemistry 1975–76, lecturer in Inorganic Chemistry, Christ Church, Oxford 1968–1975, in Biochemistry Trinity College, Oxford 1976–84; Fellow Exeter College, Oxford 1976–; Director the Glycobiology Institute 1988–; Professor of Glycobiology University of Oxford 1988–; Visiting Royal Society Research Fellow Weizmann Institute, Rehovot, Israel 1969; Departmental Demonstrator Biochemistry Department, Oxford 1969–74; Royal Society Locke Research Fellow 1974–76; Visiting Professor Duke University, NC 1968; University of Trieste, Italy 1974; University of Lund, Sweden 1977; Institute of Enzymology, Budapest, Hungary 1980; Member MRC AIDS Antiviral Steering Committee 1987, Oxford Enzyme Group 1971–88; founder member Oxford Oligosaccharide Group 1983; Director and Founding Scientist Oxford GlycoSystems Ltd. 1988, Honorary Life/Founder Member Swedish Biophysical Society 1979, member European Molecular Biological Organization (EMBO) 1988–; Wellcome Trust Award for Research in Biochemistry Related to Medicine; Judge of the Millennium Fund Competition, The Daily Telegraph 1994; Advisory Panel, The Foundation for Ethnobiology 1994; Scientific Advisory Board Hepatitis Foundation, USA 1994, Carbohydrate Bioengineering Meeting 1995; Books: Nuclear Magnetic Resonance (NMR) in Biochemistry. Clarendon Press O.U.P. 1973. Physical Chemistry Principles and Problems for Biochemists. (jtly) O.U.P. 1975 3rd Edition 1983. NMR in Biology. (jtly) Academic Press. London and New York 1977. Biological Spectroscopy. (jtly) Benjamin Cummings 1984.

units may be connected to one another by many different linkage types. Proteins and nucleic acids are almost exclusively linear and they have only a single type of linkage between (amide bonds for proteins and 3′−5′ phospho diester bonds for nucleic acids). This complexity allows carbohydrates to provide almost unlimited variations in their structures. Although carbohydrates can be present without being attached to other molecules, the majority of carbohydrates present in cells are attached to proteins or lipids and the terminology glycoprotein and glycolipids is used to reflect this. Further, the attached carbohydrate is often referred to as an oligosaccharide.

## 1.2. Glycoproteins

Glycoproteins are fundamental to many important biological processes including fertilization, immune defense, viral replication, parasitic infection, cell growth, cell−cell adhesion, degradation of blood clots, and inflammation. Glycoproteins and glycolipids are major components of the outer surface of mammalian cells. Oligosaccharide structures change dramatically during development, and it has been shown that specific sets of oligosaccharides are expressed at distinct stages of differentiation. Further, alterations in cell surface oligosaccharides are associated with various pathological conditions including malignant transformation.

## 1.3. Glycobiology

Glycobiology deals with the role of carbohydrates in biological events. Two main themes have emerged: (1) How do carbohydrates influence the properties of the proteins to which they are attached? (2) How are carbohydrates involved in recognition events?

The use of new technology has shown that at each glycosylation site on a protein, there is a *set* (or ensemble) of glycosylated structures. This has led to the concept of a glycoprotein being defined as a set of glycoforms. These all have the same amino acids but differ in the sequence or position of the attached sugars. It is the populations in this set of glycoforms that change under a variety of conditions such as disease.

## 1.4. Technology Developments

In the early 1980s the determination of carbohydrate sequences was very difficult and carried out by very few laboratories—mainly in Japan by Akira Kobata and colleagues. Some 10 years ago the determination of the sequences of the oligosaccharides associated with a glycoprotein could take up to one year. As in all newly developing fields, technology plays a crucial role so that it is now possible to accomplish this in 1−2 weeks. Some idea of the new technologies and approaches for characterizing micromole amounts of glycans is shown in Figure 1. In general, oligosaccharides are released by hydrazinolysis to produce a pool of glycans. These glycans are then fluorescently labeled. A variety of methods, including HPLC, allow profiling of the mixture and separation of individual glycans. These can then be fragmented with *arrays* of enzymes. The fragments are then combined, and the measurement of the subsequent profile allows the structure to be worked out.

## 1.5. There Is No Single Unifying Function for Oligosaccharides

One of the aims of this review is to give some indication of the general principles of glycosylation which have emerged as a result of developments in carbohydrate technology. Another, is to describe how individual sugars or sets of glycoforms affect the properties of the proteins to which they are attached. What is quite clear, is that there is no single function for oligosaccharides. Perhaps their major function is to serve as recognition markers. Additionally, oligosaccharides can modify the intrinsic properties of proteins to which they are attached by altering the stability, protease resistance, or quaternary structure. The large size of oligosaccharides may allow them to cover functionally important areas of proteins, to modulate the interactions of glycoconjugates with other molecules, and to affect the rate of processes which involve conformational changes. Glycosylation is highly sensitive to alterations in cellular function, and abnormal glycosylation is diagnostic of a number of disease states including rheumatoid arthritis and cancer. The control of glycosylation by the cell affords, in principle, a means of putting the same recognition markers on quite different proteins without having to code the infor-

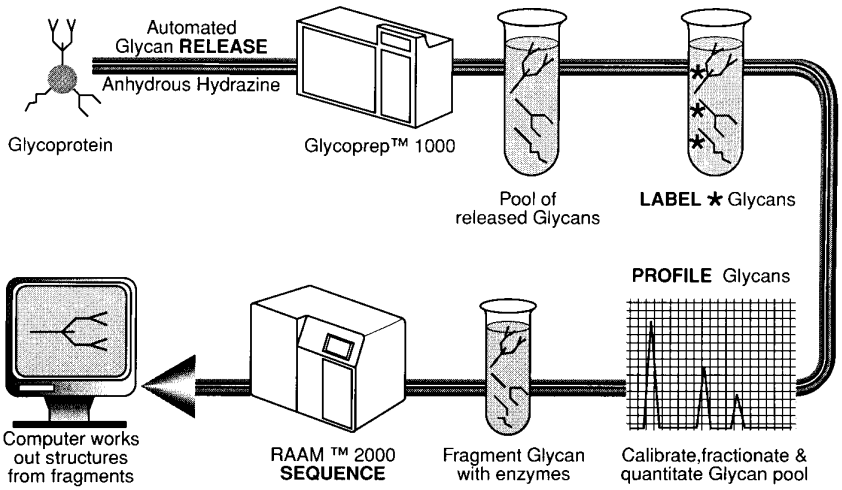



**Figure 1.** An overall process for release, labeling, and sequencing of nanomolar amounts of a glycan from a glycoprotein.

mation into the DNA of that protein. Site-specific glycosylation of a protein also suggests that the 3D structure of the protein plays a role in determining the extent and type of its own glycosylation.

In the last two years there have been two outstanding experimental contributions to the glycobiology field, which although outside the scope of this review, have great significance in understanding the function of oligosaccharides. The first is the important demonstrations of the significance of glycosylation in development by two sets of workers using gene knockout experiments.[1,2] In these, deletion of a transferase gene in developing mice resulted in embryonic lethality at about 10 days of gestation. The embryos had defects in neural tube closure and an overall stunting of growth. These types of experiments open up exciting possibilities for studying the role of glycosylation in development and indeed, in a host of other situations. The second development concerns the recognition of oligosaccharides by the serum lectin, the mannose-binding protein (MBP). This protein plays a role in immune defense and in general recognizes oligosaccharides on invading pathogens. What becomes clear from the X-ray data[3,4] on the MBP trimer is that the actual geometry of presentation of the oligosaccharides to the receptor is important in order to trigger a biological response. In general, multiple interactions with multivalent targets are required for physiologically relevant binding. Thus it is possible to have the same sugar on a whole host of different proteins, none of which may present the oligosaccharide with the correct geometry to the receptor. In this way it is possible to distinguish oligosaccharides which are "self" from those which are "nonself".

## 2. What Does a Typical Glycoprotein Look like?

### 2.1. Implication of the Conformations and Dynamics of Protein Surface Oligosaccharides in Protein Function

The majority of cell surface and secreted proteins are glycosylated, with carbohydrates covalently attached through either a nitrogen atom (supplied by the amino acid asparagine) or an oxygen atom (supplied by serine or threonine). The carbohydrate moiety of a glycoprotein may participate directly in recognition events,[5] but it may also modify the properties of the protein.[6,7] The large size of the carbohydrates[8,9] is probably the most significant factor in modifying the properties of the proteins to which they are attached. It should be noted that the distance across a carbohydrate residue (from O-1 to O-4) is approximately 5.4 Å and that the first three residues of the core of an N-linked oligosaccharide extend approximately 16 Å from head to tail. When the dynamic motions of the carbohydrate are taken into account it becomes apparent that large areas of the protein surface may be shielded by a relatively small oligosaccharide. As an N-linked oligosaccharide will typically have two or three outer arms each consisting of three or four sugar residues, the possibility that the oligosaccharide may cover an extremely large area of the protein surface is clear.

Furthermore, comparatively small motions of the protein−carbohydrate linkage, when combined with the rigidity of the carbohydrate core, will amplify the motion of the terminal arms of the oligosaccharide. This enables the carbohydrate to span an even larger area of the protein and may have a dramatic effect on the accessibility of the protein in intermolecular interactions. Accurate quantitation of these properties necessitates a knowledge of both the 3D structures of the carbohydrate and protein and, more importantly, of the time evolution of these properties. For many proteins conformational information may be obtained from crystallographic methods. However, the oligosaccharides present on glycoproteins appear much less amenable to these techniques. Moreover, for a given time period the carbohydrate exhibits greater dynamic fluctuations than the protein. NMR measurements offers insight into these dynamics, but NMR data alone are frequently insufficient to determine uniquely the conformations of the oligosaccharide. It is often necessary to combine NMR data with molecular dynamic simulations. To illustrate this, consider the enzyme ribonuclease (RNase) for which both NMR data and computer simulations have been carried out.

RNase is an example of a protein which exists *in vivo* in both non-glycosylated and glycosylated forms, A and B, respectively. RNase B has only a single

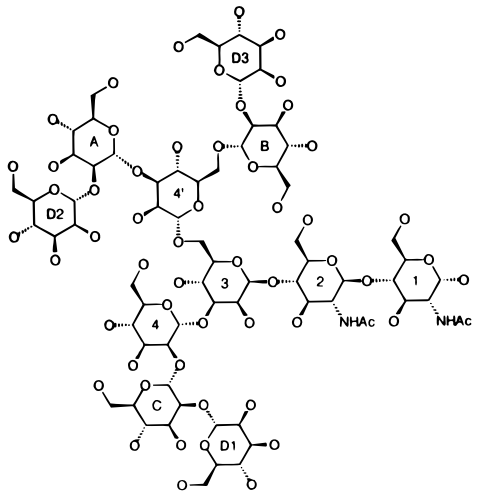



**Figure 2.** A schematic representation of $Man_9GlcNAc_2$-OH. The hydrogen atoms have been omitted for clarity.

N-linked glycosylation site at Asn-34. In bovine pancreatic RNase B, the glycosylation is characterized by a set of oligomannose $Man_{5\text{-}9}GlcNAc_2$ glycoforms (see section 4).

Despite the existence of a well-resolved X-ray crystal structure of RNase B the poor definition of the electron density associated with the oligosaccharide has prohibited any determination of the sugar conformation.[10] While NMR spectroscopy has been widely applied in the conformational analysis of proteins, including RNase A,[11−14] and RNase B,[15] unambiguous conformational determinations of oligosaccharides are less common. The ambiguity in oligosaccharide conformational analysis arises from a characteristic paucity of nuclear Overhauser effects (NOEs) between sugar residues. Typically two NOEs are observed across a given glycosidic linkage, and in oligomannose structures this has been shown to be consistent with more than one conformation.[16] A computer simulation of the dynamic properties of the oligosaccharide offers an additional approach to the conformational analysis.

## 2.2. Molecular Dynamics (MD) Simulation of $Man_9GlcNAc_2OH$

The application of MD techniques to proteins is typically part of the refinement protocol in X-ray crystallography.[17] In contrast, MD simulations of oligosaccharides are often applied in conjunction with NMR refinement. This difference leads to unique requirements for the simulations of oligosaccharides. In order to compare MD-generated data with NMR-derived data the duration of the simulation should be sufficient to sample adequately the conformational space of the macromolecule. An MD trajectory from an unrestrained simulation that is in agreement with the NMR-derived data, provides strong support for both the structure and the computational method. However, the determination of an appropriate duration for an unrestrained simulation requires careful analysis of the MD data for each system under examination.

The MD simulation of Man-9 (see Figure 2) used a novel parameter set for oligosaccharides and glycoproteins (GLYCAM_93) with the AMBER molecular mechanical force field.[18] The simulation was initiated with an energy minimized conformation in which the values for the glycosidic torsion angles were consistent with reported intraresidue NOE data.[16,19−23] In order to emulate as closely as possible the conditions of the NMR experiment, the oligosaccharide was immersed in a theoretical box of water molecules[24] and the simulation performed for 750 ps. The temperature was maintained at 300 K throughout the isobaric simulation. During the simulation, each of the glycosidic angles in Man-9 was monitored to determine whether any significant variations occurred.

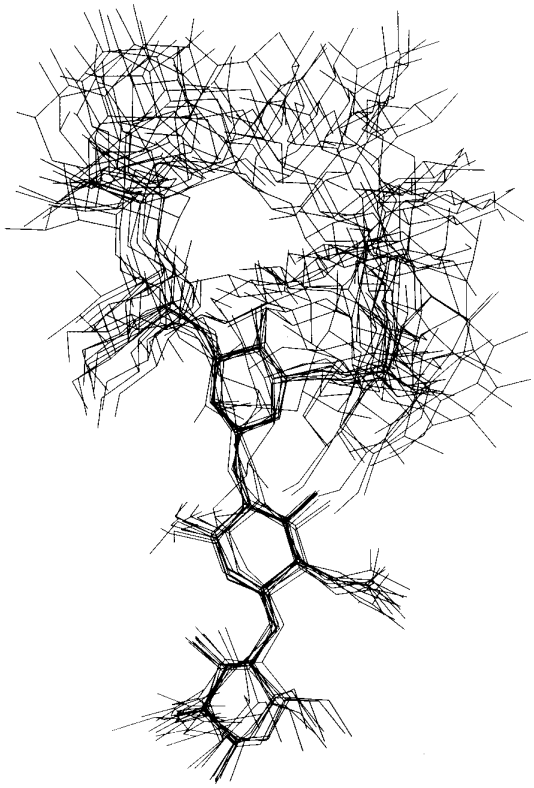


**Figure 3.** A least-squares overlay of rings 1−3 from 10 snapshots from the trajectory of Man-9 each separated in time by 15 ps.

## 2.3. The Core Conformation of the Oligosaccharide in N-Linked Glycoproteins Is Independent of the Protein

During the simulation the core residues were found to maintain a relatively constant conformation, despite average fluctuations in the individual glycosidic torsion angles of approximately 15°. An overlay based on a least-squares fitting of the non-hydrogen atoms of rings 1−3 of 10 randomly selected snapshots from the simulation is presented in Figure 3. The constancy of the conformation of these residues is readily apparent.

Although in general the oligosaccharides of glycoproteins are not well resolved in X-ray crystal structures, there are a few examples in which the residues are clearly defined. These are those in the lectin *Erythrina corallodendron*,[25] the serine protease human leukocyte elastase (HLE),[26] the Fc domain of human $IgG_1$,[27] and a variant surface glycoprotein

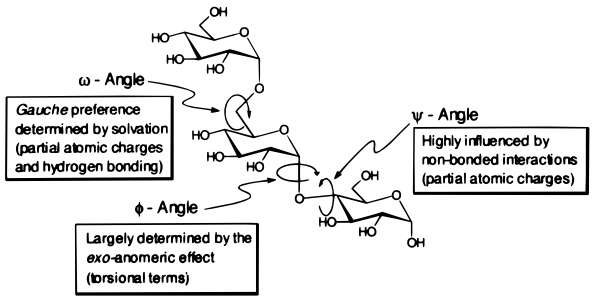



**Figure 4.** Definition of the angles of rotation of carbohydrate conformation and flexibility.

from *Trypanosoma brucei*.[28] Although the oligosaccharides present in the crystal structures varied in sequence and by the presence or absence of a fucose residue attached to the first GlcNAc residue, the core conformations were remarkably similar to each other. (A least-squares fitting of the ring atoms of core residues 1 and 2 for the average MD conformation with each of the glycoprotein cores gives atomic displacements of only 0.37−0.55 Å.) These similarities lead to the conclusion that the conformation of the di-*N*-acetylchitobiose core in N-linked glycoproteins is independent of the protein and would also be that present of the free sugar.

## 2.4. Outer Arm Conformation of the Oligosaccharide

The conformations of the remaining glycosidic linkages from the MD simulations are in good agreement with values of the glycosidic angles derived from previous NMR studies[16,19−23,29] and from MD simulations of related mannobiosides.[29,30] Interestingly, an analysis of the hydrogen bond and van der Waals interactions in the mannobioside simulations show that the Man(α1−2)Man linkage prefers a stacked conformation, in which the hydrophobic faces of the two sugars are in close proximity ($\varphi = -41°$, $\psi = 47°$). In contrast, the (α1−3)Man linkage prefers an extended conformation that optimizes interresidue and sugar−solvent hydrogen bonding ($\varphi = -56°$, $\psi = -32°$). The different conformations reflect not only the effects of linkage position, but also the attempt to maximize hydrophobic and hydrophilic interactions. As observed experimentally, the *gauche* (*gt*) orientation of the (α1−6)Man linkage in the disaccharide was the favored orientation during the simulation.[30] This *gauche* preference was attributed to bulk solvation effects (Figure 4).

## 2.5. Dynamic Sugar Model of RNase B

A structural model for RNase B can be constructed from the crystal structure of the protein and the simulation data for Man-9. While the sugar is not resolved in the crystal structure, the side chain of Asn-34 was well defined. In each of the glycoprotein crystal structures discussed above as well as in those of glycopeptides[31,32] the Asn−GlcNAc linkage displays the same conformation. This conformation has been reported also to be present in solution.[31,33] Given the similarity of the C1−N−C=O atomic sequence of an *N*-linkage to a Cα−N−C=O peptide bond, its preference for planarity and a torsion value of 0° is not surprising. Provided with this observation it is possible to link the dynamic model of the free sugar (each snapshot overlaid on GlcNAc-1) to the side chain of Asn-34 in the crystal structure of RNase B to generate the structure shown in Figure 5. This model indicates that a considerable area of the protein surface is shielded by the sugar.

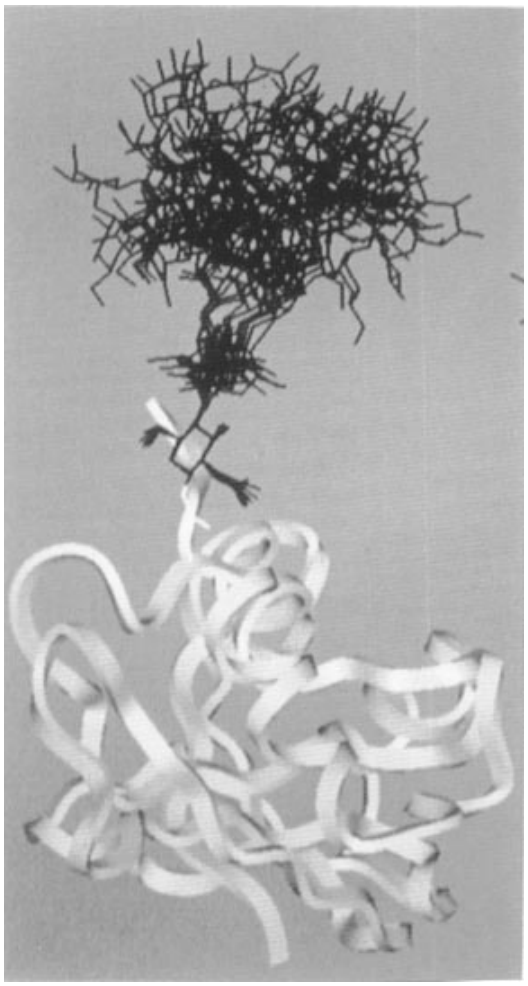


**Figure 5.** The Man-9 glycoform of RNase B based on the 2.5 Å X-ray crystal structure with an overlay of 10 oligosaccharide conformations from a 750 ps MD trajectory of Man-9 linked through Asn-34. The side chain of Asn-34 was maintained in the crystallographically determined orientation. In order 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.

## 2.6. Consequences of Protein Side-Chain Flexibility on the Presentation of the Oligosaccharide

Motion of the Asn−GlcNAc linkage can alter the presentation of the sugar relative to the protein surface. For example, consider the case when the side chain angles $\chi_1$ and $\chi_2$ (see Figure 6) are varied over an arbitrary range of ±30° from the crystallographic values 30° and −152°, respectively). For illustrative purposes the average oligosaccharide conformation is used and the variations of ±30° (in 15° increments) lead to no unfavorable van der Waals contacts between the oligosaccharide and the protein. Figure 7 shows that even this small movement has

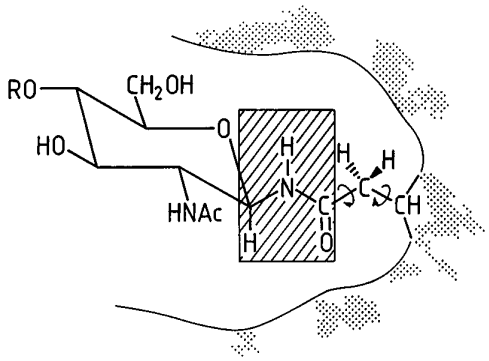



**Figure 6.** Schematic diagram of the Asn−GlcNAc linkage indicating the planarity of the C1−N−C═O glycosidic linkage and the flexible side chain angles $\chi_1$ and $\chi_2$.

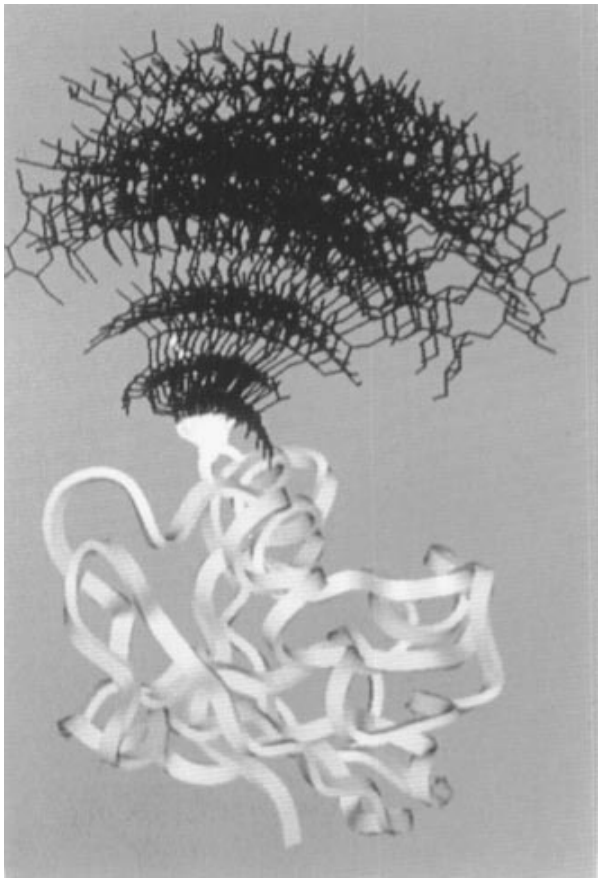


**Figure 7.** The effect of flexibility of the Asn-34 side chain on the orientation of the oligosaccharide in the Man-9 glycoform of RNase B. The $\chi_1$ and $\chi_2$ angles of the side chain of Asn-34 were varied by ±30° in 15° intervals from the crystallographically determined orientation. The 25 resulting orientations are displayed. All hydrogen atoms have been omitted for clarity.

a profound effect on the volume of space potentially available to be occupied by the oligosaccharide. In the absence of specific protein−sugar interactions this model must represent a comparatively modest estimate of the extent of the protein surface enveloped by the oligosaccharide.

### 2.7. Biological Implications

The functional activity of RNase in its interaction with double-stranded RNA is decreased by glycosylation (see section 6). One explanation of this is that the oligosaccharide sterically hinders the binding of RNA to the enzyme and is discussed in more detail in section 6.

In general, it may be concluded that the sugar moiety of a glycoprotein may have a significant effect on the properties of the protein. It is apparent that the molecular volume occupied by the sugar is large and therefore able to shield a large section of the protein surface. When the dynamic nature of the oligosaccharide and the flexibility of the asparagine side chain are also taken into account the portion of the protein surface covered by the sugar is even more extensive. This indicates that the sugar may interfere with the normal functioning of the protein, including regions of the protein that are considerably removed from the actual linkage site. Since the conformation of the N-glycosidic linkage is both rigid and planar, the actual conformational space available to an N-linked oligosaccharide in a glycoprotein may depend to a large extent on the flexibility of the asparagine side chain within the local environment of amino acids.[33]

## *3. Some Factors Which Control Protein Glycosylation*

Some rules have emerged with respect to the factors which control the attachment of oligosaccharides to potential glycosylation sites and the subsequent enzymatic modifications of the glycan chains. While the potential oligosaccharide processing pathways[34] available to a nascent protein are dictated by the cell in which it is expressed, its final glycosylation pattern is also the result of constraints imposed by the 3D structure of the individual protein.

### 3.1. The Primary Peptide Structure Determines the Number and Location of Potential Glycosylation Sites

The two main classes of glycosidic linkages to proteins (Figure 8) involve either oxygen in the side chain of serine, threonine, or hydroxylysine (O-linked glycans) or nitrogen in the side chain of asparagine (N-linked glycans). To be glycosylated, an asparagine residue must form part of the tripeptide AsnXSer where X is any amino acid apart from proline, although the presence of this sequon is not in itself sufficient to ensure glycosylation. The role of the peptide sequence in directing O-glycosylation is less clear, but a Pro residue, at −1 and +3, may make it favorable. Recently, a consensus sequence (Cys-X-X-Gly-Gly-Ser/Thr-Cys) has been found to correlate with O-fucosylation in epidermal growth factor domains.[35−37] Other O-linked glycans include those linked through hydroxylysine are found in collagen and also O-linked GlcNAc residues which are found in the nucleoplasmic and cytoplasmic compartments of cells.[38] Physiologically this O-GlcNAc modification is highly labile and seems to be abundant in all eukaryotes.[39]

A third type of linkage to proteins has been found for an increasing number of cell surface proteins, which are known to be inserted into the lipid bilayer via a glycophosphatidylinositol (GPI) anchor.[40] Only six amino acids serve as a GPI attachment site; these

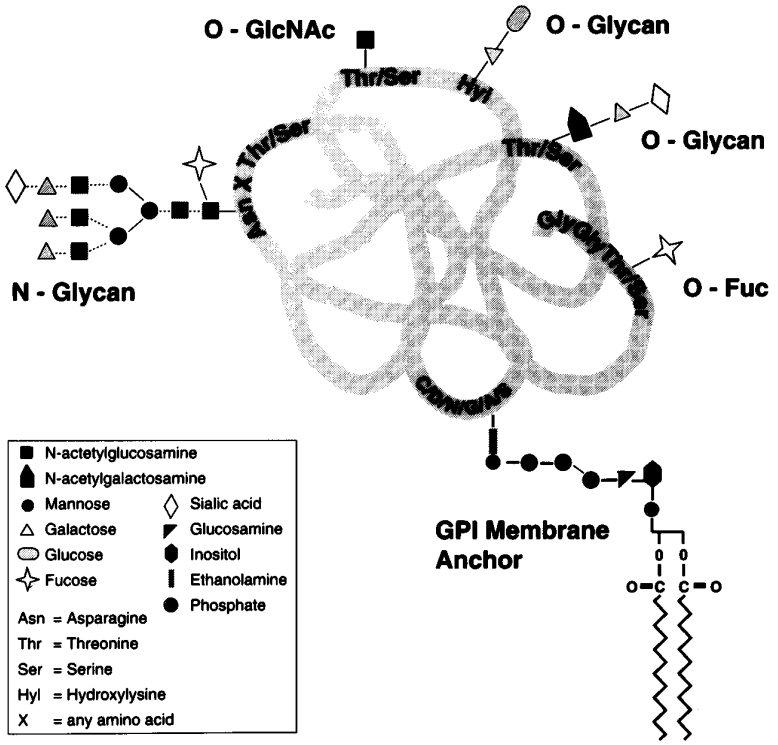



**Figure 8.** A schematic representation of the main forms of attachment of glycans to a polypeptide. Several glycans may be attached to a single polypeptide and some potential sites may remain unoccupied.

are Cys, Asp, Asn, Gly, Ala, and Ser (CDNGAS).[41] The amino acids Gly, Ala, and Ser predominate at the +1 positions and are obligatory at +2 positions.

### 3.2. Structure and Diversity of N-Linked Glycans

All N-linked glycans contain the pentasaccharide Manα1−6(Manα1−3)Manβ1−4GlcNAcβ1−4GlcNAc as a common core. On the basis of the structure and the location of glycan residues added to the trimannosyl core, N-linked oligosaccharides can be classified into four main groups.[34,42] These are oligomannose (high mannose), complex, hybrid, and poly-*N*-acetyllactosamine (Figure 9).

Oligomannose-type glycans contain only α-mannosyl residues attached to the trimannosyl core. Complex-type glycans contain no mannose residues other than those in the trimannosyl core, but have "antennae" or branches with *N*-acetylglucosamine residues (Figure 10a) at their reducing termini attached to the core. The number of antennae normally ranges from two (biantennary) to four (tetraantennary), but a pentaantennary structure has been reported in hen ovomuvoid.[42] While various monosaccharides can be found in the antennae, the presence or absence of fucose and a "bisecting" GlcNAc on the core contributes to the enormous structural variation of complex-type glycans (Figure 10b). Indeed complex-type N-glycans show the largest structural variation in the subgroups resulting mainly from the combinations of different numbers of antennae and variations of monosaccharides in the outer chains. Some of the outer chain structures found in complex-type sugar chains are shown in Figure 10b.

The hybrid-type N-glycans have the characteristic features of both complex-type and high mannose-type glycans. One or two α-mannosyl residues are linked to the Manα1−6 arm of the trimannosyl core (as in the case of oligomannose-type glycans) and usually one or two antennae (as found in complex type glycans) are linked to the Manα1−3 arm of the core.

The fourth group is the poly-*N*-acetyllactosamine *N*-glycans containing repeating units of (Galβ1−4GlcNAcβ1−3-) attached to the core. These repeats are not necessarily uniformly distributed on the different antennae and the lactosamine repeat may also be branched. Poly-*N*-acetyllactosamine extensions are most frequently found in tetraantennary glycans.[43]

### 3.3. Structure of O-Linked Glycans

In contrast to N-linked glycans, O-linked glycans do not share a common core structure. They are based on a number of different cores.[44] So far they can be categorized into at least six groups according to different core structures (Figure 11). These cores can be elongated to form the backbone region by addition of Gal in β1−3 and β1−4 linkages, and GlcNAc in β1−3 and β1−6 linkages. Although the glycans are often linked to serine or threonine residues through GalNAc, the linkages may be through other residues e.g. fucose. We should also note that single glycans such as fucose or GlcNAc may be O-linked to the peptide backbone. O-GlcNAc can also be β-linked as found in cytoplasmic and nucleoplasmic proteins.[38]

### 3.4. Cell Type Influences Glycosylation

The type of cell has a major role in determining the extent and type of glycosylation, which is both species and tissue specific.[9] Oligosaccharides are formed on an "assembly line". For protein-bound and

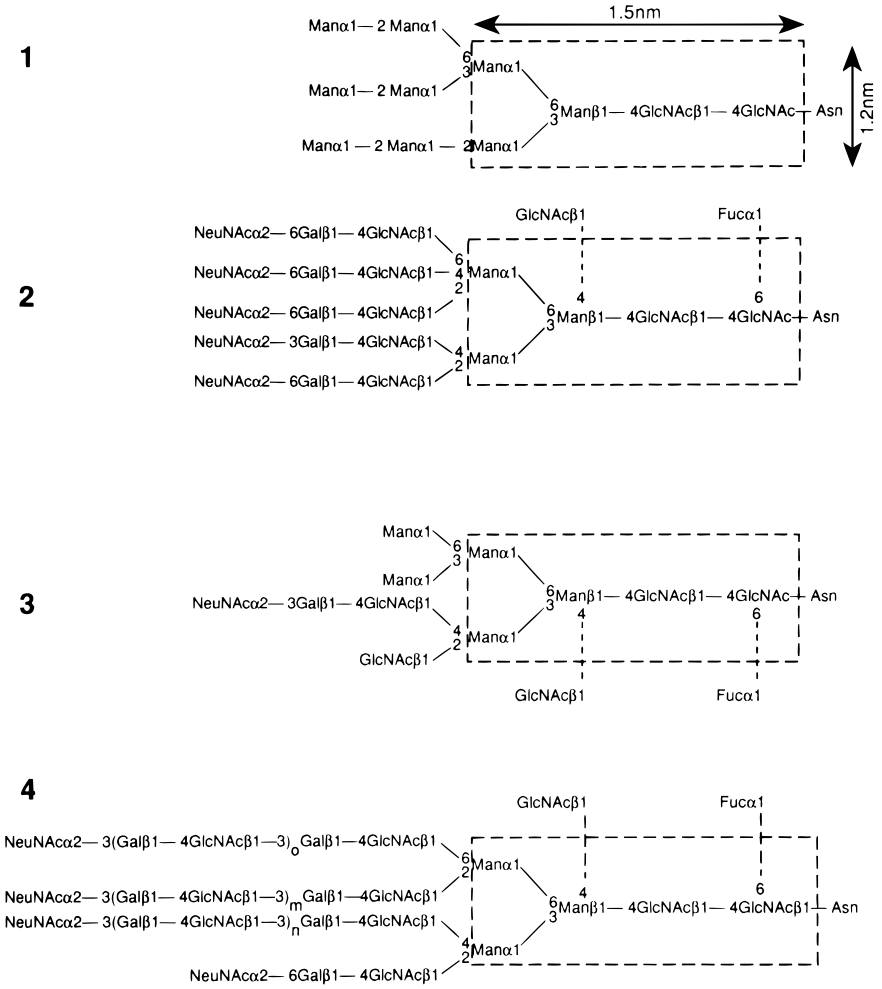



**Figure 9.** Four groups of N-linked glycans: (1) oligomannose (high-mannose); (2) complex; (3) hybrid; (4) poly-*N*-acetyllactosamine ($o > m > n$). The structure within the box contains the pentasaccharide core common to all N-linked glycans.

lipid-bound oligosaccharides this is the endoplasmic reticulum (ER) and the Golgi apparatus.[34] A series of membrane-bound glycosidases and glycosyltransferases act sequentially on the growing oligosaccharide as it moves through the lumen of the ER and Golgi apparatus. Many different enzyme reactions (typically eight for a complex oligosaccharide such as those on IgG) are involved in the processing pathways. Each individual enzyme reaction may not go to completion, giving rise to *glycoforms* or glycosylated variants of the polypeptide. The type of enzymes (glycosidases and transferases), their type, concentrations, kinetic characteristics, and compartmentalization, reflect both the external and internal environment of the individual cell in which the protein is glycosylated. This explains why the glycosylation patterns of natural glycoproteins may be influenced by physiological changes such as pregnancy and also by some diseases which may affect one or more of the enzymes in the cell. For example, in IgG isolated from rheumatoid arthritis patients the galactosyl transferase activity may be decreased. This results in an alteration of the glycoform populations in the Fc reflecting an increase in the proportion of agalactosyl N-linked glycans.

The glycosylation of recombinant glycoproteins can be sensitive to changes in manufacturing conditions, such as the glucose concentration of the culture medium.[45] The glycosylation pattern basically reflects the type of cell used in the expression system and the use of different cell lines can result in significant glycosylation differences. For example, there are differences in the branching structure in the complex type oligosaccharides. These arise from the different expression of the GlcNAc transferases shown in Figure 12.[46,47] The oligosaccharide structures at Asn-289 of recombinant human plasminogen expressed in both chinese hamster ovary (CHO) and *Mamestra brassicae* cell lines[175,176] have been compared with those associated with human serum plasminogen.[177] While human plasminogen contains only the variably sialylated complex biantennary glycan, $Gal_2$, $Man_3$, $GlcNAc_4$, both the CHO and the insect cell trimer express an additional range of glycoforms, which includes oligomannose structures, and, in the CHO cell line, tetraantennary complex glycans.

### 3.5. The 3D Structure of the Protein Influences the Extent and Type of Glycosylation

Although the same glycosylation machinery is available to all the proteins which are translated in a particular cell and use the secretory pathway, it

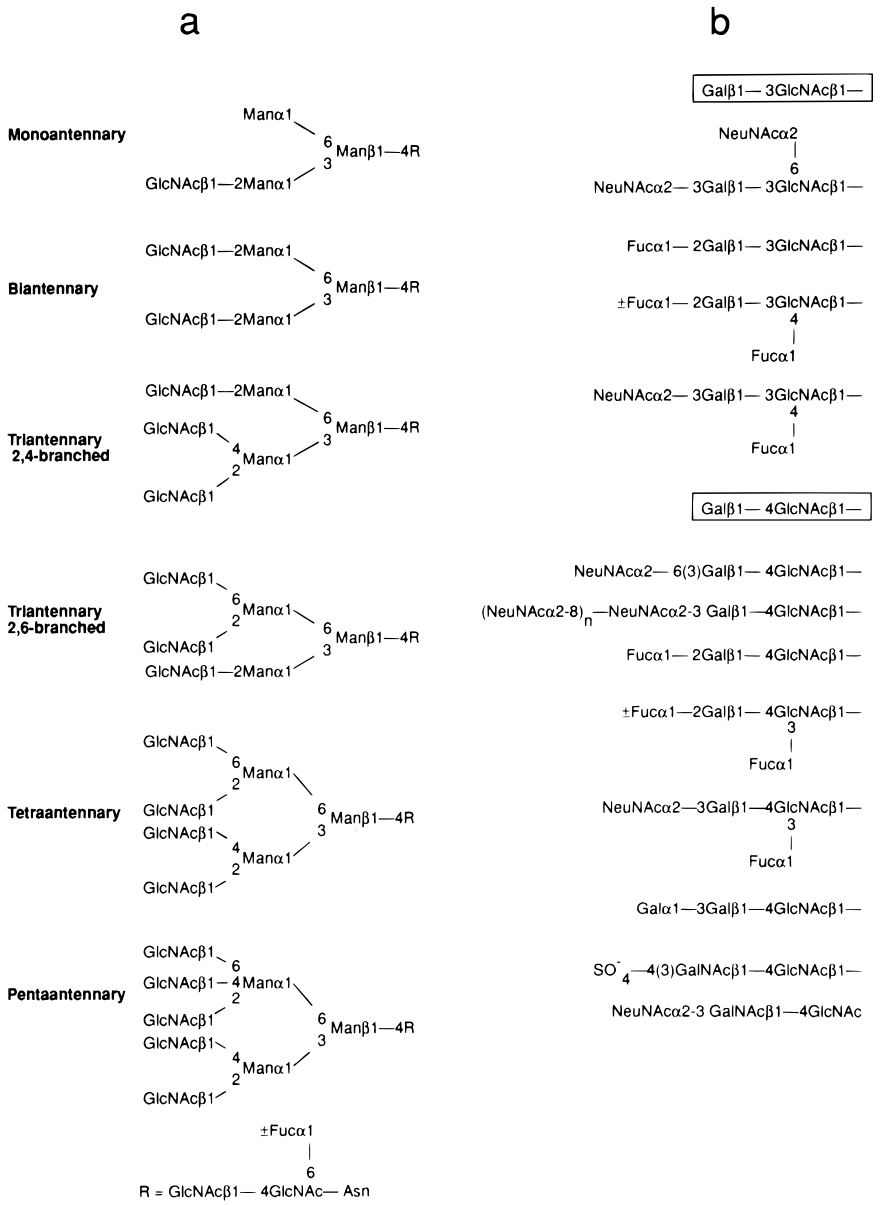



**Figure 10.** Two major elements that create the diversity of structures of complex-type sugar chains: (a) branching differential and (b) variations in chain structures.

has been estimated that between 10% and 30% of potential glycosylation sites are not occupied.[48,49] Moreover site analysis has shown that the distribution of different classes of N-linked oligosaccharide structures is frequently specific for each site on a protein. In the case of rat brain Thy-1 (Figure 13), for example, site 23 contains only oligomannose structures, site 74 has only complex and hybrid, while all three classes of glycans are present at site 98.[50,51]

The 3D structure of the individual protein clearly has a role in determining the type and extent of its glycosylation. A number of mechanisms may be involved. These include the following:

(i) The position of the glycosylation site in the protein. N-Linked sites at the exposed turns of $\beta$-pleated sheets, which are sometimes close to proline residues, are normally occupied while those near the C-terminus are more often vacant.

(ii) Access to the glycosylation site on the developing oligosaccharide. This may be sterically hindered by the local protein structure or by protein folding which may compete with the initiation of N-glycosylation.

(iii) Interaction of the developing oligosaccharide with the protein surface. This may result in a glycan conformation which may alter the accessibility to specific glycosyltransferases or glycosidases.

(iv) Interaction of the glycosyl enzymes with the protein structure. This can lead to site-specific processing.

(v) Glycosylation at one site in a multiglycosylated protein. This may sterically hinder events at a second site on the same molecule.

(vi) The interaction of protein subunits to form oligomers. This may prevent glycosylation or restrict the glycoforms at individual sites.

### 3.6. Generation of Glycoforms

The initial event in N-linked glycosylation is the cotranslational transfer, to an asparagine residue

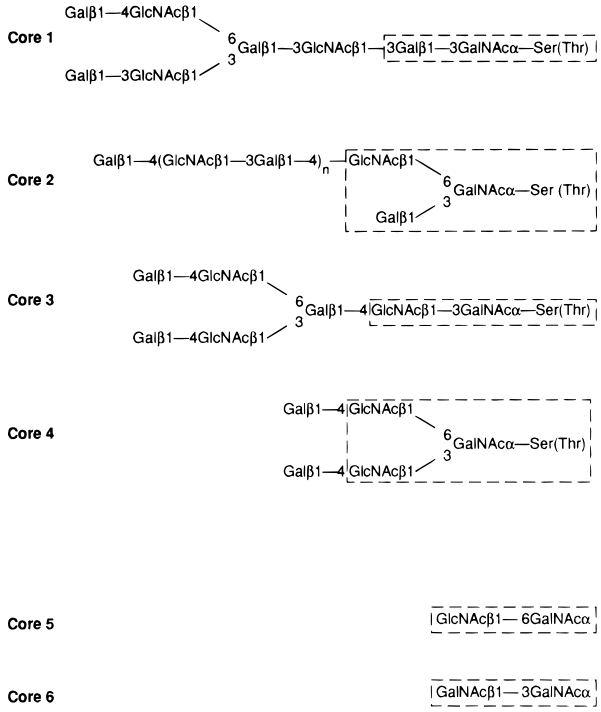



**Figure 11.** Six types of core structures (boxed) among those found in O-linked glycans.

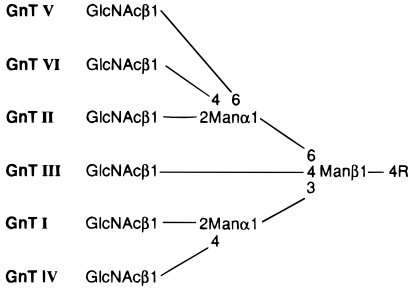



**Figure 12.** The "branching" GlcNAc-transferases. Five antennae can be initiated on the Manα1−3Man$\beta$1−4GlcNAc$\beta$1−4GlcNAc$\beta$-Asn core of N-glycans by the actions of GlcNAc-transferases I, II, IV, V, and VI. A "bisecting" GlcNAc can be added by GlcNAc-transferase III.

within a glycosylation sequon, of the dolichol-linked $Glc_3Man_9GlcNAc_2$ oligosaccharide to the nascent polypeptide chain (Figure 14a). A series of trimming events then occurs. First, glucosidase I hydrolyses, the outermost glucose (α1−2) residue, followed by the removal of the remaining two α1,3-glucose residues by glucosidase II. These reactions are reasonably rapid (on the order of minutes) and the protein is assumed to be fully folded by this stage. Subsequent enzyme reactions may clearly be influenced by the 3D structure of the protein in respect of the accessibility of the individual enzymes. The routing of glycoproteins within the cell, the compartmentalization of trimming enzymes with different specificities, and the competing secretion pathways are also important factors controlling the biosynthesis of N-linked oligosaccharides.

The synthesis of O-linked oligosaccharides is entirely a post translational event with a series of enzymes acting sequentially on the fully folded protein. Many of the factors discussed above will also apply. Initially O-linked oligosaccharides are co-

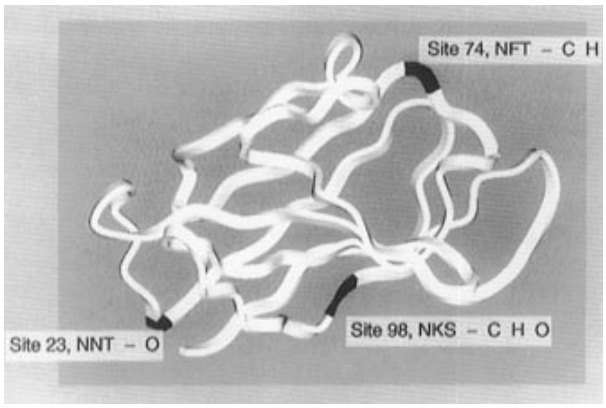



**Figure 13.** A molecular model of rat brain Thy1 demonstrating that glycosylation is site specific. Site 23 (Asn23-AsnThr) contains only oligomannose structures, site 74 (Asn74PheThr) has only complex and hybrid, while all three glycans are present at site 98 (Asn98LysSer).

valently attached through an O-glycosidic monosaccharide, and a serine or threonine. Some of the enzymes that act subsequently may be showed by both the N- an O-linked biosynthetic pathways.

In summary we have seen that both the cell and the protein influence glycosylation and that competing reactions and pathways give rise to glycoform populations (Figure 15).

## 4. Oligosaccharide Technology

All glycoforms show the same amino acid sequence but differ with respect to the number, location, or sequence of attached glycan. One of the simplest examples of glycoforms is found in bovine pancreatic RNase[52] which is a globular protein composed of a single domain. It has one variably occupied glycosylation site and occurs naturally as a mixture of the unglycosylated protein (RNase A) and five glycoforms (RNase B Man 5−9) in which from 5−9 mannose residues are attached to the di-*N*-acetylchitobiose core.

Fractionation of the individual glycoforms can be achieved using capillary electrophoresis (Figure 16). However, the technique is not yet generally applicable for two principal reasons. First, the number of individual glycoforms is generally too large to allow a practical separation. Second, the fractional differences in physicochemical properties between neutral glycoforms are often relatively small, rendering their resolution and separation virtually impossible.

Analysis of protein glycosylation is most commonly performed by analyzing the glycans following release from their conjugate peptide. A strategy for protein glycosylation analysis consists of four distinct steps (Figure 17).

### 4.1. Release of Glycans from Glycoprotein

In order to release glycans a *general* method is required that is independent of the protein to which the glycan is attached. For this reason a chemical method is used, which involves the use of hydrazine to cleave both N- and O-glycosidic bonds (Figure 18).[178] Hydrazinolysis releases glycans which are intact and with a free reducing terminus and is

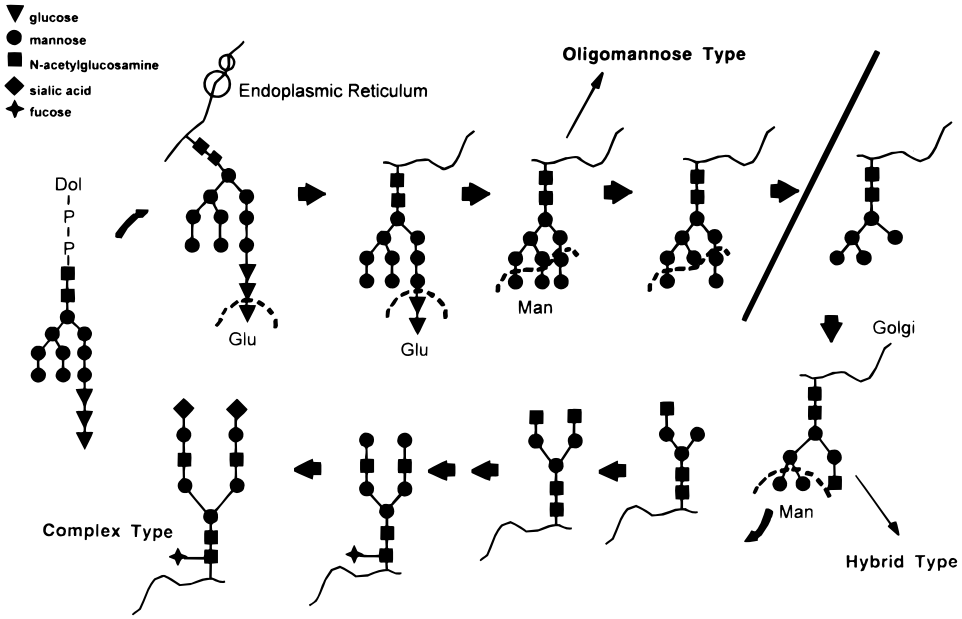



**Figure 14.** Representation of some of the steps in oligosaccharide biosynthesis.[34]

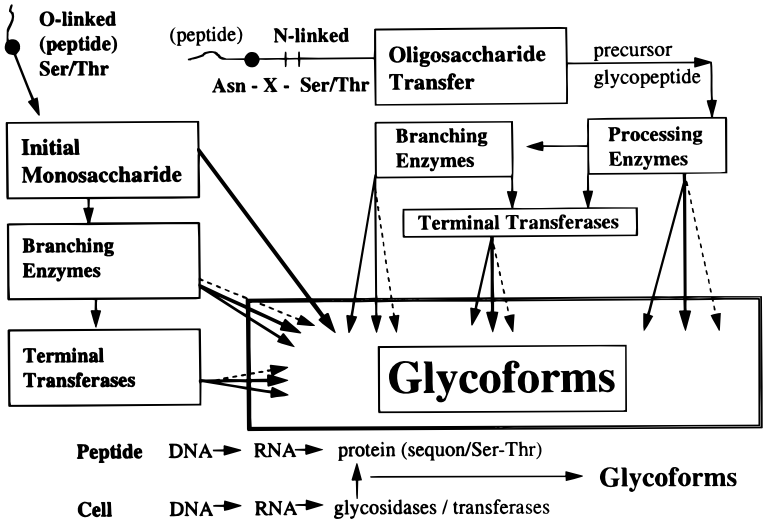



**Figure 15.** Schematic representation of N- and O-linked glycosylation pathways. The influence of the peptide is initially in providing an appropriate glycosylation site, but the 3D structure may influence subsequent processing. The repertoire of enzymes provided by the cell is a major factor in determining the glycosylation patterns of the peptide, as too, are the various competing reactions and pathways. Together these events result in the generation of multiple glycoforms, indicated by the arrows. (Note too the difference between the initial steps in the pathways. O-Linked addition of sugars is entirely a posttranslational event, while the precursor oligosaccharide in N-linked processing is added cotranslationally.)

nonselective (with respect to the glycan), and the process is amenable to automation. The GlycoPrep 1000 (Figure 19) produced by Oxford GlycoSystems automates the hydrazinolysis of glycans and has been optimized to release both N- and O-glycans or only O-glycans. Released glycans are separated from peptide fragments, leaving unreduced, intact glycans ready for labeling and analysis. Enzymatic methods are also much used to release glycans from peptides and denatured glycoproteins. For example, under denaturing conditions, peptide-*N*-glycosidase *F* (*F. meningosepticum*) will generally cleave the $C\gamma-N\delta$ bond of the glycosylated asparagine side chain releasing the intact N-linked glycans. However, care must be taken to ensure that the release is nonselective. Enzymatic release of O-glycans is much more difficult since most O-glycanases require an unsubstituted disaccharide core as a substrate.

### 4.2. Labeling of the Released Glycans To Enable Their Detection in Subsequent Procedures

This involves a reaction of the reducing terminus of individual glycans using a method which must be independent of the glycan sequence. Two methods which are commonly used to label glycans are reductive amination with a fluorescent compound, such as 2-aminobenzamide,[183] and reduction with alkaline sodium borotritide, to give the radiolabeled derivative.[182]

### 4.3. Profiling the Pool Glycans To Determine the Types of Glycans Present and Their Relative Molar Properties

Three types of glycan profiling are commonly used:

*Mass Profile.* The molecular weight of each glycan present can be quickly determined by mass spec-

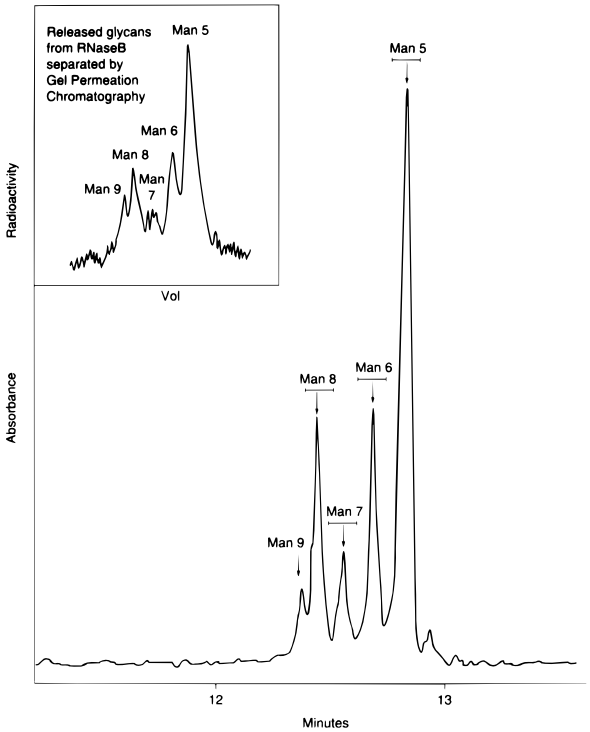



**Figure 16.** The glycoforms of bovine ribonuclease pancreatic B separated at the protein level by capillary electrophoresis. The inset shows the analysis of the released glycans by P4 gel permeation chromatography: the relative proportions of the oligomannose sugars also represents the distribution of glycoforms in the intact glycoprotein, since ribonuclease has only one glycosylation site.

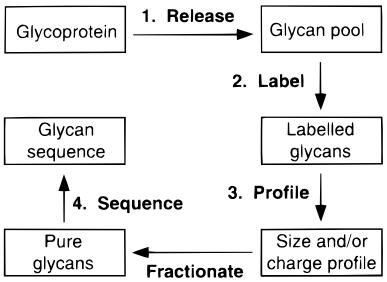



**Figure 17.** A complete strategy for protein glycosylation includes steps to release, label, profile/fractionate, and sequence glycans.

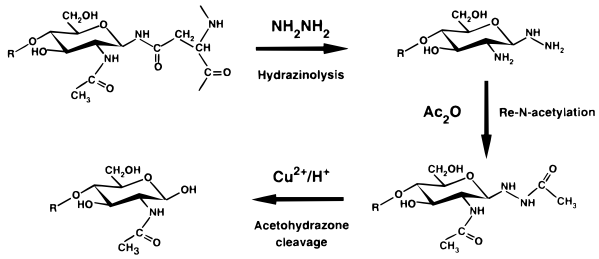



**Figure 18.** Method for release of oligosaccharides using hydrazinolysis.

trometry. Methods which give only the parent ion for each glycan include matrix-assisted laser desorption ionization (MALDI) (Figure 20) and produce spectra which are fairly easy to interpret on neutral glycans, although laser energy-induced desialylation can often be observed. Detection limits are in the region of 1 pmol for a single oligosaccharide. Fragmentation, giving information on sequence and branching, can be observed with time-of-flight instruments fitted with a reflectron.[53]

*Size Profile.* Gel permeation chromatography (GPC) is frequently used to determine the size profile (in glucose units GU) of a deacidified glycan pool by coinjection of dextran hydrolysate standard "ladder" with the sample.

*Charge Profile.* Anion exchange chromatography (AEC) is used to determine the charge profile of a glycan library.

## 4.4. Structural Analysis

There is no single technique that is able routinely to provide all the information required for structural analysis. This usually involves the combined use of several physical, chemical, and biochemical techniques including NMR, mass spectrometric, and enzymatic analysis. These techniques have been critically reviewed elsewhere.[54] The enzymatic is clearly the method of choice in biological systems for which often only very small amounts of material (picomoles or less) are available.

The basis of *enzymatic sequencing* for the elucidation of the structure of N- and O-linked glycans is to evaluate the susceptibility of the glycan to a series of sequence-grade exoglycosidases of defined specificity. There are still many linkages for which specific enzymes are not available. However, a sufficient number have been purified to allow the primary sequence analysis of N-linked oligosaccharides using exoglycosidases in multiple defined mixtures, with analysis performed in a single chromatographic step. This process called the reagent array analysis method (RAAM) is, in this case, summarized in Figure 21, parts a and b, and involves dividing a purified, labeled N-glycan sample into nine equal aliquots. Each aliquot is incubated with a precisely defined mixture of exoglycosidases called a reagent array. The products of each incubation are combined, and a single analysis is performed on the pool of products.

In essence, a mixture of exoglycosidases is used to digest the sample glycan until a linkage is reached which is resistant to all the exoglycosidases present in that mix. By omitting one or more different exoglycosidase(s) from each mixture, different "stop point" fragments of the oligosaccharide are generated. Full use is, therefore, made of all positive data (*the exoglycosidases hydrolyse linkages up to the "stop point"*) and negative data (*the exoglycosidases do not hydrolyse linkages beyond the "stop point"*). By labeling the original oligosaccharide at the single reducing terminus, fragments retaining the original reducing terminus are readily distinguished from released monosaccharides.

Chromatographic separation of the combined "stop point" fragments generates a pattern that is, in effect, a "signature" of that oligosaccharide treated by the enzyme array used. This signature is characterized by the size (GU) and relative signal intensity of each fragment. A RAAM computer program constructs the carbohydrate structure from the observed signature directly from the output of the RAAM Glycosequencer (Figure 19).

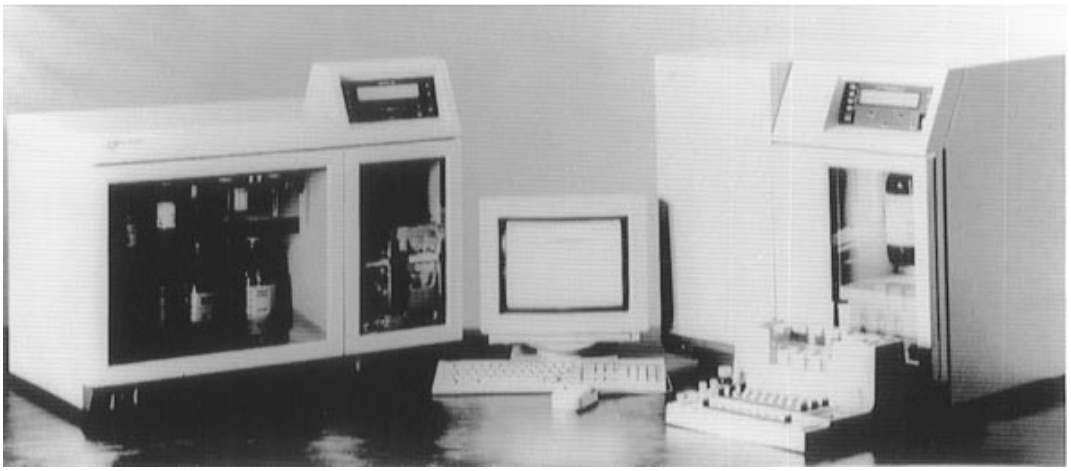


**Figure 19.** GlycoPrep 1000 and RAAM instruments from Oxford GlycoSystems.

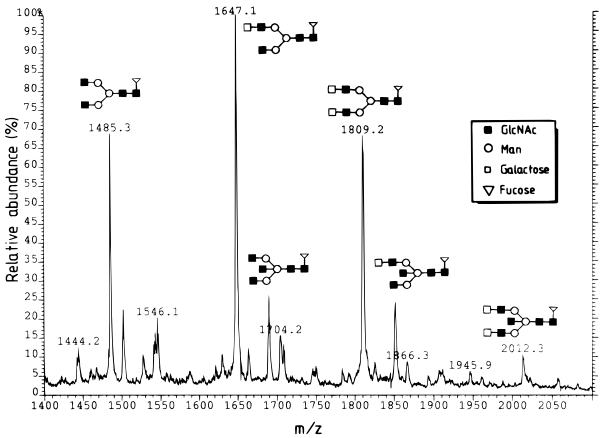



**Figure 20.** Matrix-assisted mass desorption spectrum of IgG glycans.

## *5. Characteristics of Protein Glycosylation*

There are three levels of understanding of protein glycosylation. These are (i) the influence of the general characteristics of the glycoprotein being processed; (ii) the local conformation at the individual glycosylation sites; and (iii) third, the available repertoire of glycosylation processing enzymes for the particular tissue or cell type.

The work on the three members of the immunoglobulin superfamily, CD4 and CD2,[55,56] and Thy-1,[50,51] well illustrate the main points of protein glycosylation. These molecules are shown schematically (Figure 22).

### 5.1. Importance of the Overall Protein Conformation in Determining Glycosylation

The chromatographic gel filtration profiles of the sugars released from soluble recombinant forms of human CD4, rat CD4, and rat CD2, expressed in Chinese hamster ovary (CHO) cells,[55,57] are shown in Figure 23. The glycosylation potential in CHO cells has been well characterized (see references in ref 58) and processing to multiantennary and poly-*N*-acetyllactosamine oligosaccharides. Rat soluble CD2 (sCD2) shows glycosylation typical of the CHO cell line, with bi-, tri-, and tetraantennary complex, glycans, some of which contain poly-*N*-acetyllactosamine extensions (Figures 23 and 24).[55] In contrast, despite the available repertoire of processing enzymes in this cell line, the N-linked glycans of rat and human soluble CD4 (sCD4) in this cell line had quite different glycosylation profiles. Most of the oligosaccharides were of the biantennary complex, hybrid, or oligomannose type.[56,59−61] These results indicate the importance of the protein contribution in determining glycosylation.

### 5.2. Effect of Local Protein Conformation—Glycosylation Shows Site Specificity

The overall degree of processing of CD4 is less than in CD2. As CD2 is structurally very similar to the first two domains of CD4[62] the general 3D conformation of these members of the immunoglobulin superfamily cannot be the only factor influencing their glycosylation. Clearly the local amino acid sequence and microenvironment of the glycosylation site must also be an important determinant.

To illustrate this, the site specificity of glycosylation in the rat CD4 was determined by isolating the glycopeptides containing the glycosylation sites at Asn-270 and Asn-159[56] from the wild type. The glycosylation patterns at each site were different (Figure 25). In particular, oligomannose and hybrid structures were restricted to Asn-159, the nonconserved site. The conserved site (Figure 22) contained exclusively biantennary complex oligosaccharides. This was identical to that reported for the equivalent site in human sCD4.[60] Therefore, overall differences in glycosylation between the rat and human glycoproteins can be accounted for by site-specific glycosylation at the nonconserved sites. The detailed analysis showed that there were three oligosaccharide structures associated with Asn-270 and 10 with Asn-159 giving an ensemble for CD4, in this case, of 30 glycoforms.

There is the further question of whether the processing at each site is independent of the processing at the other. One approach is to produce mutants with the appropriate glycosylation sites deleted. In this case the glycosylation patterns from the variants with either Asn-270 or Asn-159 mutated show strong similarity to that from the glycosylated peptides of the wild-type (Figure 25). It can be concluded that

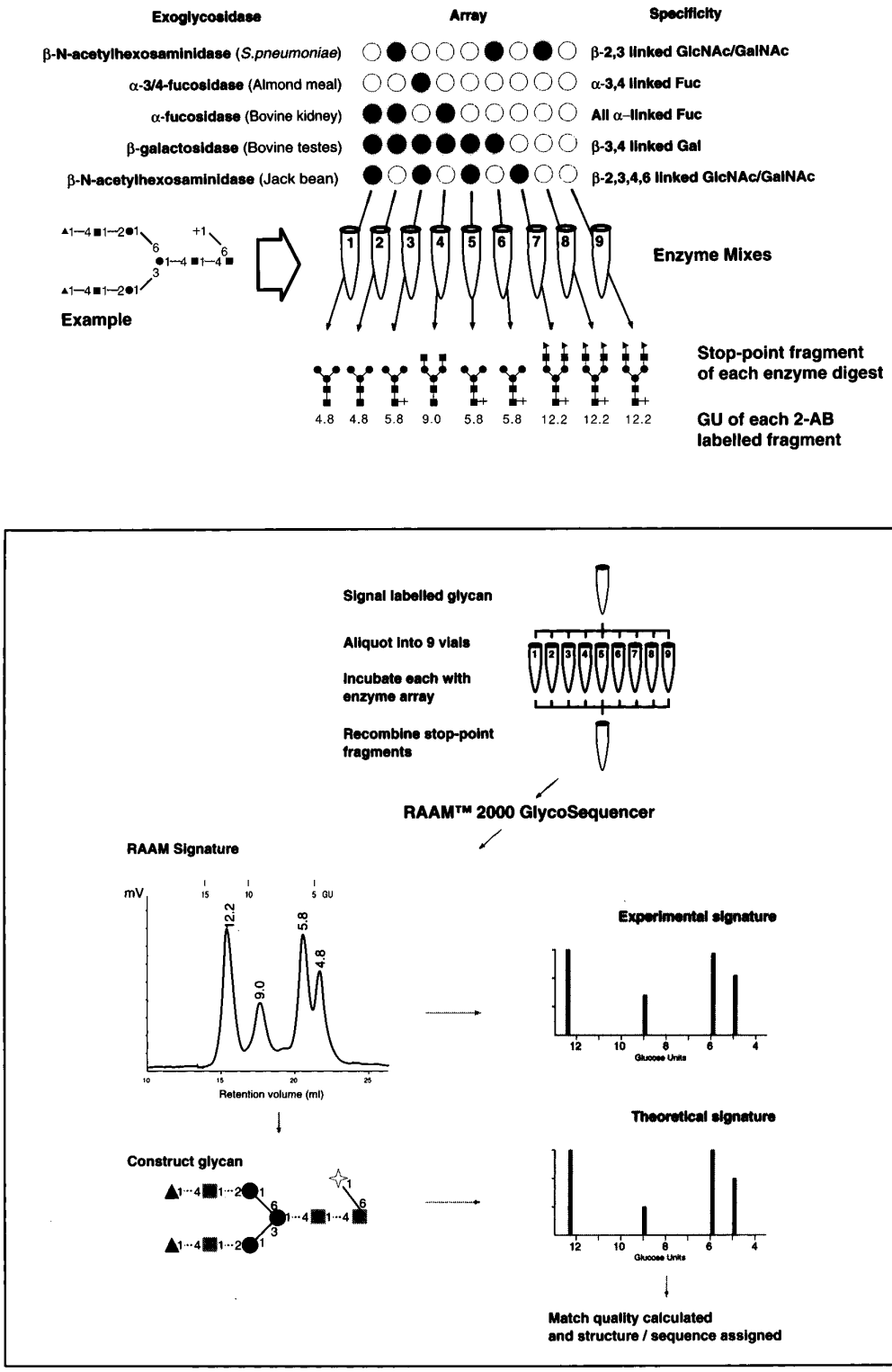



**Figure 21.** (top) The RAAM enzyme array consisting of eight different enzyme mixtures and an enzyme bank and (bottom) a summary of oligosaccharide sequencing by RAAM.

specific and independent processing occurs at each glycosylation site.

In an extension of this study the glycosylation profile from a truncated form of rat sCD4, consisting of only domains 1 and 2, were made. This variant contained the nonconserved glycosylation site at Asn-159.[57] There was now more processing at this glycosylation site than in either the glycopeptide or the full length glycosylation variant containing this site (Figure 25).[56] It seems therefore that the presence of domains 3 and 4 affect processing at this site in the intact molecule.

An unexpected finding[56] was the presence of terminal α-galactose residues on approximately 20% of the oligosaccharides from human sCD4. Although oligosaccharides terminating in α-galactose residues

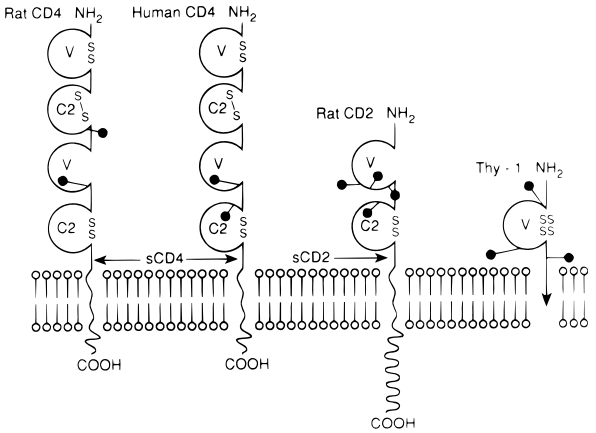



**Figure 22.** Schematic drawings of rat and human CD4, rat CD2, and Thy-1. The molecules are drawn with the circles representing immunoglobulin superfamily (IgSF) domains and the "lollipops" N-linked oligosaccharides. The glycosylphosphatidylinositol membrane anchor of Thy-1 is depicted as a vertical arrow. The IgSF domains are designated as V or C2 on the basis of sequence analysis.[166] The positions of the mutations introduced in the CD4 and CD2 molecules to produce the recombinant soluble forms are indicated by horizontal arrows. Adapted from ref 166.

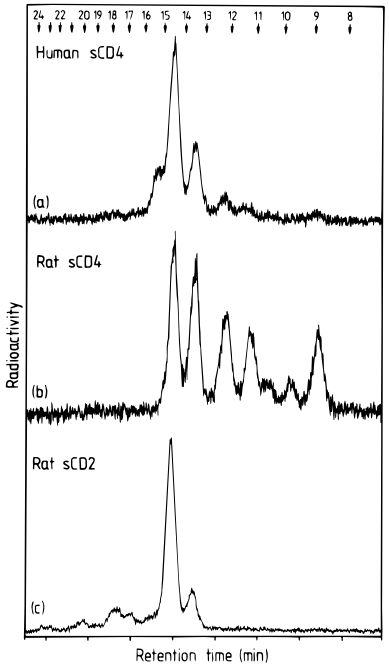



**Figure 23.** Bio-Gel P-4 gel filtration profiles of the desialylated, tritium-labeled oligosaccharides of recombinant soluble CD4 and CD2 expressed in CHO cells: (a) total oligosaccharides of human sCD4; (b) total oligosaccharides of rat sCD4; and (c) total oligosaccharides of rat sCD2. The vertical arrows indicate the elution positions of isomaltooligosaccharides containing the corresponding number of glucose units. The time axis is marked at 100 min intervals. Data taken from from refs 55 and 56.

are known to occur normally in glycoproteins from nonprimate mammals[63,64] examination of the glycosylation of other recombinant glycoproteins expressed in CHO cells has not revealed the presence of these moieties (see references in ref 58). The occurrence of terminal α-galactose residues in the oligosaccharides from human sCD4 is probably not associated specifically with this molecule but is more likely to be caused by the activation by the transfection process of a latent, endogenous α-galactosyltransferase.

## 5.3. Glycosylation Is Protein-Specific, Site-Specific, and Tissue/Cell-Specific

The characteristics of the natural expression of N-glycosylation at individual glycosylation sites have been probed within a single immunoglobulin domain, thereby eliminating interdomain effects. Thy-1, a GPI membrane-anchored molecule, has one (V-type) immunoglobulin-like domain and three N-glycosylation sites (Figure 22).[65]

A comparative study of rat Thy-1 from the thymocyte with that from the brain showed that there was tissue specificity of N-glycosylation[51] although the amino acid sequences are identical. Furthermore, the differential effects of tissue glycosylation were expressed at the level of individual glycosylation sites in the different tissue-derived Thy-1 molecules.[51] The site distribution of oligosaccharides was such that no Thy-1 molecules were found to be in common between the two tissues (Figure 26). Thus each tissue created unique sets of glycoforms.

An important consequence of this is that a glycoprotein must be viewed as an ensemble or collection of glycoforms. For instance, tissue plasminogen activator "tPA" from human colon fibroblast is characterized by a different set of glycoforms compared with cells from tPA.[66,67]

tPA is a protease glycoprotein consisting of five domains: a fibronectin "finger" domain, an epidermal-growth factor domain, two kringles, and the catalytic serine protease domain.[68,69] Binding sites for fibrin are present within the "finger" and kringle domains. There are two main classes of glycosylated variants of the tPA molecule, referred to as type I and type II; type I has three N-linked sugars, at Asn-117, Asn-184, and Asn-448 (Figure 27), whereas type II has only two, at Asn-117 and Asn-448. tPA shows site-specific glycosylation and each of the classes is characterized by a population of glycoforms. Variable occupancy of site 184 affects the fine structure of the glycan population at site 448 demonstrating, in this case, that glycosylation at one site can influence the processing at another. In Bowes melanoma type 1 tPA the major species at site 448 were found to be neutral glycans of the complex or hybrid structures. In type II, however 72% of the structures were sulfated complex glycans (Figure 27), and there were relatively few neutral sugars.[9] This is an interesting example of site-specific glycosylation at one site (448) being controlled or influenced by glycosylation at another site (184). Further detailed studies[70] have confirmed that some of the complexed glycans on tPA contain terminal galactose monosaccharide residues instead of, or as well as, terminal GalNAc residues. This suggests the presence of competing $\beta$1,4-galactosyl and *N*-acetylgalactosaminyl transferases in the Bowes cell line. Interestingly, when the galactose residue is present it is always found substituted on the Manα1−6 arm, suggesting that there may be steric restrictions by the growing glycoconjugate which result in arm-specific glycosylation. This is clearly an example of the protein glycoconjugate influencing its own glycosylation. Additionally, there

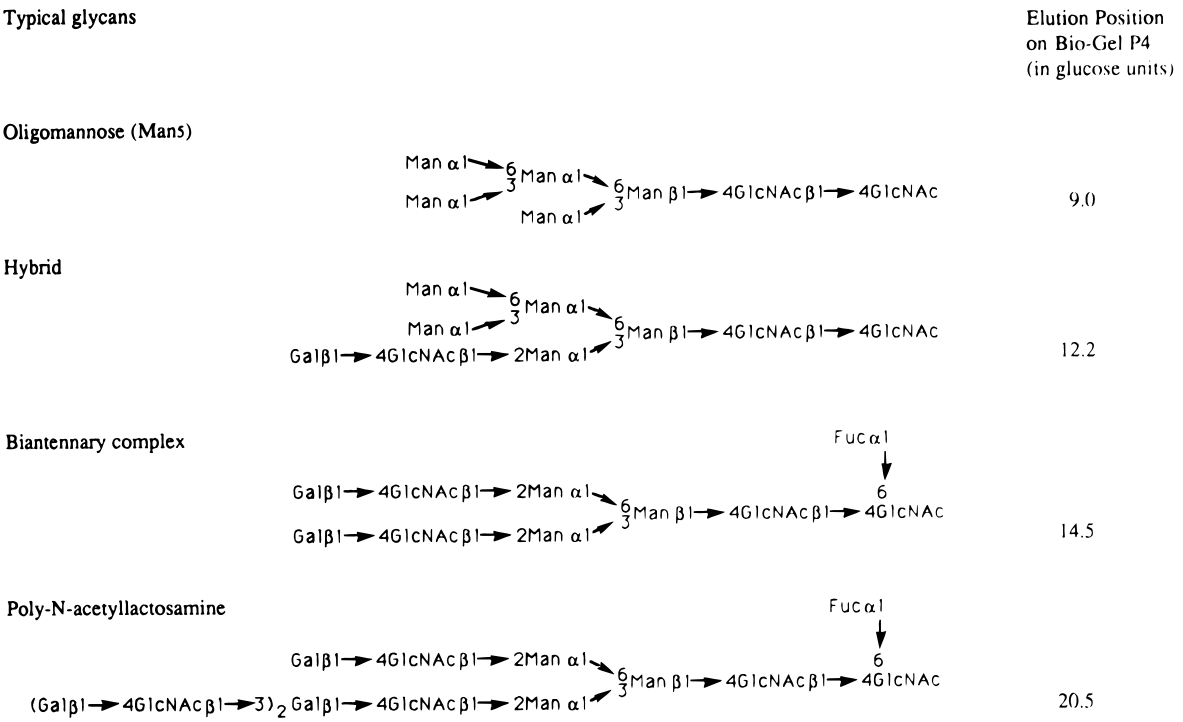



**Figure 24.** Examples of the structure of the different types of neutral desialylated N-linked oligosaccharides: Fuc, L-fucose; Gal, D-galactose; GlcNAc, D-*N*-acetylglucsamine; Man, D-mannose.

may well be "motifs" in the protein that are recognized by particular glycosylating enzymes, such as the Pro Leu Arg motif, which is recognized by the glycoprotein hormone-specific *N*-acetylgalactosamine-transferase.

## 6. Glycosylation Site Occupancy Can Modulate Enzyme Activities

The presence of natural glycosylation variants, as in tPA, makes it possible to probe the influence of oligosaccharides on protein interactions directly. In addition, to tissue plasminogen activator (tPA types I and II), ribonuclease (RNase A and B), and plasminogen (Plg types 1 and 2) (Figure 29) are also important examples of variable sequon occupancy. While RNase consists of a single domain, both plasminogen and its activator are multidomain proteins.

*Ribonuclease (RNase).* The size of this molecule makes it amenable to NMR studies. This has allowed the effects of glycosylation on the stability of the molecule to be probed[15,71] by comparing the hydrogen−deuterium solvent exchange rates for the NH protons of RNase A and RNase B. The presence of the sugar decreases the exchange rate in 30 of the 124 amino acid residues (Figure 28), suggesting that the overall dynamic stability of the molecule is enhanced such that it becomes more rigid. Some of the affected residues are close to the oligosaccharide attachment site, others are as far away as 30 Å, while several (including His-12, His-119, and Asp-121) are involved in the active site, which catalyzes the cleavage of the phospho diester linkages of RNA.

The functional variations associated with glycosylation of RNase have been probed in three ways: (1) by determining the relative abilities of RNase A and B to mediate the hydrolysis of double-stranded RNA,[52] (2) by examining the resistance of RNase A and B to proteases,[52] and (3) by the abilities of antibodies to distinguish between each form. Furthermore, in the case of RNase B, the enzyme activities of several glycoforms have been reported and may be ranked in terms of decreasing activity as: RNase A > RNase Man-0 = RNase Man-1 > RNase Man-5 = RNase B.[52] The enzyme's active site is located in a groove that bisects the protein.[72] Efficient hydrolysis of RNA necessitates correct alignment of the RNA and the enzyme's active site. This is achieved in part through an interaction between the 5′-terminal phosphate of RNA and a cluster of cationic residues on the proteins surface (Lys-31, Lys-37, Arg-10, and Arg-33).[73] Since Asn-34 is present on the surface of the protein near this binding site, it is tempting to speculate that the attenuated RNase activity of the glycoforms, relative to that of the nonglycosylated form, arises from steric hindrance between the oligosaccharide and the RNA.

*tPA.* The two variants of tPA differ by occupancy of site 184 (section 5.3). Various properties of tPA are affected by the occupancy of site 184 with an N-linked oligosaccharide (type I). These include the following:

*(a)* A slower rate for the plasmin-mediated conversion of single chain tPA to two-chain tPA. tPA is synthesized as a single-chain tPA molecule. Plasmin catalyzes the cleavage of the Arg275−Ile bond in native single-chain tPA to form a disulfide bond-linked two-chain species.[74] The second-order rate constant ($k_{cat}/K_m$) for type II tPA was found to be about twice that for type I tPA,[75] indicating that glycosylation at site 184 hinders the conversion of single- to two-chain tPA. Additionally, single-chain tPA has a lower activity and susceptibility to inhibition compared to the two-chain form[76,77] providing

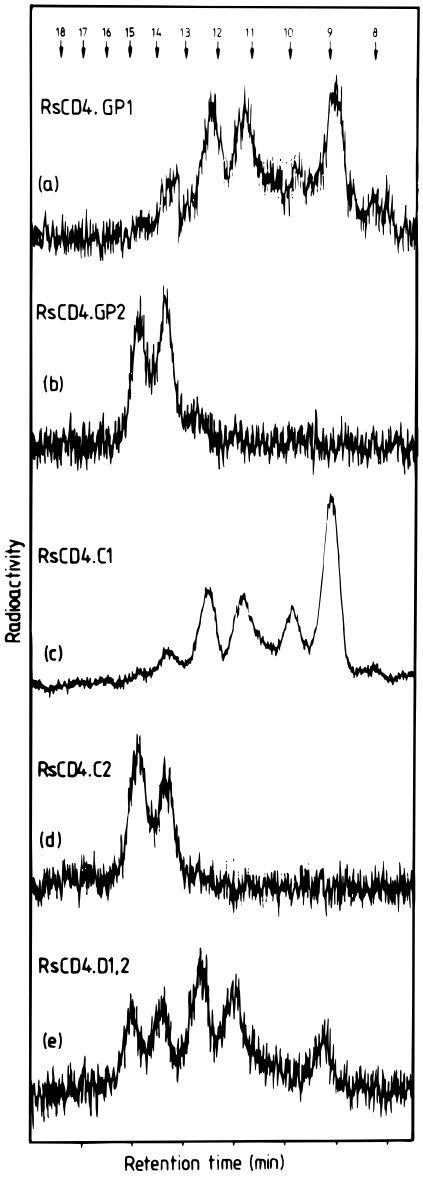



**Figure 25.** Bio-Gel P-4 gel filtration profiles of the desialylated, tritium-labeled oligosaccharides of the glycopeptides and glycosylation mutants of rat soluble CD4: (a) total oligosaccharides of rat sCD4-derived glycopeptide from the region of the first glycosylation site; (b) total oligosaccharides of rat sCD4-derived glycopeptide from the region of the second glycosylation site; (c) total oligosaccharides of rat sCD4 with the second glycosylation site removed; (d) total oligosaccharides of rat sCD4 with the first glycosylation site removed; (e) total oligosaccharides of rat sCD4. $NH_2$-terminal-two IgSF domain form. For details of the annotation see legend to Figure 2. (Reproduced from Figure 8 in ref 56. Copyright 1993 Journal of Biological Chemistry.)

further evidence that glycosylation of tPA serves to modulate activity.

*(b)* A decreased affinity of tPA for lysine and a lower fibrinolytic activity. The decreased affinity of type I tPA for lysine is well established, and affinity chromatography is commonly used to separate type I tPA from type II. Fibrin clot assays have shown that type II tPA binds more efficiently than type I tPA to fibrin,[67] probably through lysine sites exposed on degraded fibrin. Moreover, the fibrinolytic activity of type II tPA and plasminogen has consistently been found to exceed that of type I tPA and plasminogen, regardless of the cell line in which the tPA is produced. Variable occupancy of site 184 is therefore one of the factors which controls the rate at which plasmin is generated, and this may be related to the differences in the affinity of type I and type II tPA for lysine. The differential effects of glycosylation on the rate, particularly when combined with those for plasmin, may well be a mechanism for ensuring persistence of fibrinolytic activity.

*Plasminogen*. Plasminogen, the natural substrate of tPA, is a multidomain protein consisting of five kringle regions and a serine protease domain (Figure 30). It is a mixture of two major glycoforms[78,79] which have the same amino acid sequence[80] and contain one O-glycosylation site and a potential N-glycosylation site in kringle 3. Both type 1 and type 2 plasminogen contain an O-linked sugar chain at Thr-345 while the N-glycosylation site at Asn289ArgThr contains a biantennary sugar in type 1, but is unoccupied in type 2.[81] Molecular modeling (Figure 30) suggests that the presence of the N-linked oligosaccharide (type 1) may result in significant shielding of the kringle 3 domain of the protein. The extent of glycosylation may have physiological relevance.[82] Human recombinant plasminogen, expressed in *Escherichia coli* and consequently not glycosylated, is resistant to activation by tPA.[82] The importance of the O-linked glycoforms (in plasmingoen 2) has been demonstrated,[83] who isolated six glycoforms differing in sialic acid content. The O-linked sugar chain had previously been assumed to be a tetrasaccharide but the recent work[83] which suggests the presence of polysialic acid should prompt a detailed reexamination. However, they found that the kinetic activity of the different glycoforms was decreased as the sialylation increased.

Some of the properties of plasminogen which are altered by N-linked glycosylation include

(a) A slower rate of the $\beta$- to $\alpha$-conformational changes which is from an open conformation to a compact form. This change may be important in the formation and rearrangement of the ternary complex of plasminogen with fibrin and tPA.[84]

(b) A 10-fold weaker binding of plasminogen to U937 cells.[85] This suggests that the N-linked glycan may hinder sterically the recognition event.

(c) A lower affinity for lysine sepharose.[81] Again this may be explained sterically by the sugar at site 289 limiting access to the lysine binding site in kringle 4.

(d) A slower rate of activation of plasminogen by urokinase. A lower activation rate by tPA of human neonatal plasminogen.[86] This is also true of the activation of plasminogen by tPA on the surface of rat hepatocytes.[87]

### 6.1. The Multimolecular Interaction of tPA with Plasminogen and Fibrin Is Modulated by Glycosylation

The roles which both the cell and the individual protein play in determining glycosylation site occupancy and glycan processing may be an important means of control, enabling the cell to respond flexibly to its internal and external environment through the

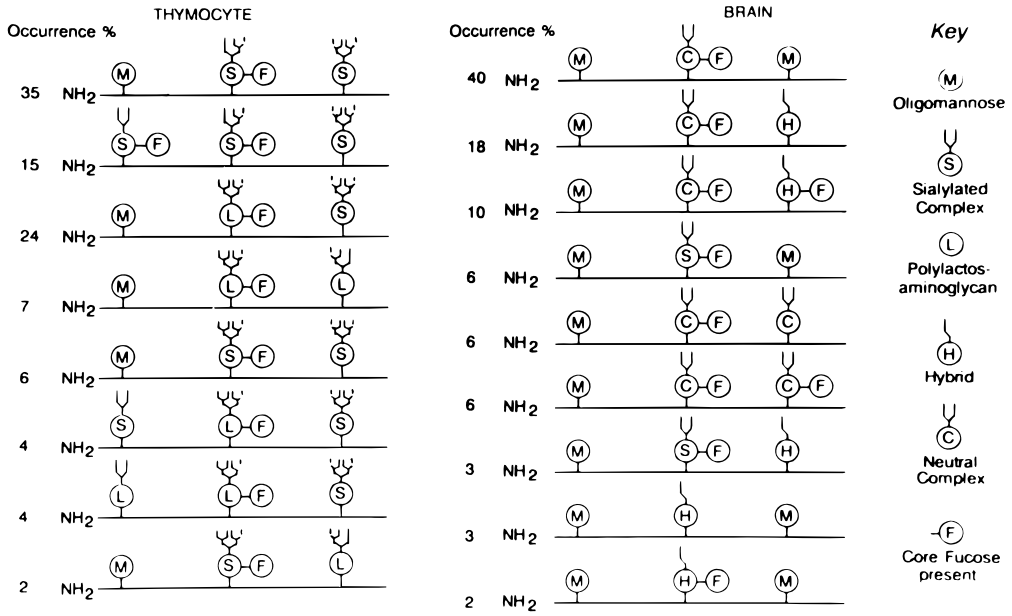



**Figure 26.** A comparison of the nature and percentage composition of the glycoforms of thymocyte and brain-derived rat Thy-1 glycoprotein. Glycoforms representing less than 2% abundance are not illustrated.

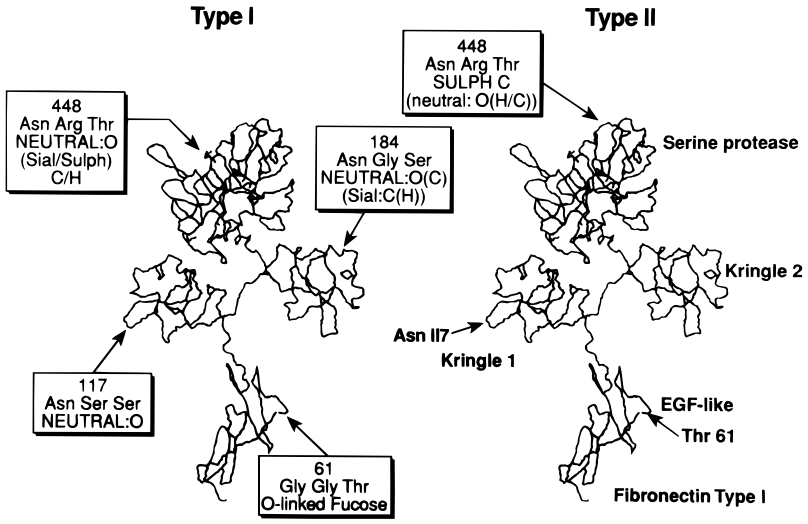



**Figure 27.** Site specificity of glycosylation in tPA. Note that the classes of glycan structures present at Asn-448 on tPA depend on the site occupancy of Asn-184, while those at site 117 do not.

production of a range of glycoforms, allowing the protein a range of dynamic and functional activities.

For instance, type 1 and type 2 plasminogen exhibit different kinetics[88] when interacting with the different types of tPA in the ternary complex with fibrinogen fragments (Figure 32). The variable glycosylation of both molecules affects the $k_{cat}/K_m$ values which range from 1.5 $\mu$M $M^{-1}$ $s^{-1}$ for interaction of the two fully glycosylated enzymes to 3.3 $\mu$M $M^{-1}$ $s^{-1}$ for the least glycosylated ones. The reaction was analyzed using the different glycosylated variants of tPA and plasminogen. It was shown that glycosylation at Asn-289 on plasminogen type 2 affects the association of the ternary complex, while glycosylation at Asn-184 as tPA type II affects kcat for the reaction. The exact nature of the glycosylation of multidomain proteins may be important in their control, particularly for those proteins that are involved in both inter- and intramolecular rearrangements.

## 6.2. Variable Glycosylation Site Occupancy on Carbohydrate-Deficient Glycoprotein Syndrome (CDGS)

The carbohydrate-deficient disorders are a group of multisystemic diseases characterized by severe disorders of the nervous system, which include growth and psychomotor retardation, abnormal distribution of adipose tissue, "peau d'orange", squints and abnormal ocular movements, and infertility.[89] One type of the disease results from a deficiency in the attachment of the oligosaccharides to glycoproteins. A large number of serum glycoproteins have been reported to be abnormal with respect to their glycoform populations and these include transport proteins, glycoprotein hormones, complement factors, enzymes, and enzyme inhibitors. Among these the transport protein, tansferrin, is an important diagnostic indicator of the disease. Transferrin is an iron

**Figure 28.** Effects of glycosylation on the rates of amide H/D exchange for selected residues of RNase. A schematic representation of ribonuclease highlighting those residues whose amide protons (shown by the space-filled balls) show modified H/D exchange as a result of glycosylation of the enzyme at Asn-34. The structure of $Man_5GlcNAc$ (drawn to the same scale as the protein) is shown attached to Asn 34 and extending away from the protein into solution. (Reprinted from ref 167. Copyright 1994 American Chemical Society.)

transport protein containing bi- and triantennary complex glycans attached to Asn413LysSer and Asn611ValThr, both of which sites are normally occupied. Normal transferrin contains glycoforms which differ only in the number of sialic acid residues present (usually between 2 and 6). Transferrin form CDGS patients may contain additional glycoforms where only one of the two glycosylation sites is occupied, and those in which neither site has been glycosylated.[90] These differences in glycoform populations may be an indication of disease severity, while also providing a rapid diagnostic test.[91] Further studies of this syndrome may reveal important new functional roles for glycosylation, particularly in development.

## *7. Some Structural Roles for Oligosaccharides*

We have seen in the previous section that the major effects of an oligosaccharide in modulating enzyme activity arise from the large size of the oligosaccharides which affect inter- or intramolecular interactions. We have also noted (with RNase) that the presence of a sugar can affect the dynamic stability of the protein.[71] More conventional structural roles for oligosaccharides are in anchoring a protein to a membrane, and in helping to maintain the quaternary structure of a protein such as the Fc portion of the immunoglobulin G molecule.

### 7.1. Glycosyl-Phosphatidylinositol (GPI) Anchors

Since 1985 over 100 examples of GPI anchored proteins have been described.[41] The GPI anchor is an alternative anchoring mechanism to the transmembrane polypeptide domain of type I membrane proteins. The first detailed structural studies on GPI anchors, of *T. brucei* variant surface glycoprotein (VSG) and of rat brain Thy-1, were carried out in Oxford.[92,93] The structures were determined by a combination of glycosyl linkage and compositional analyses by GC−MS, specific chemical cleavages,

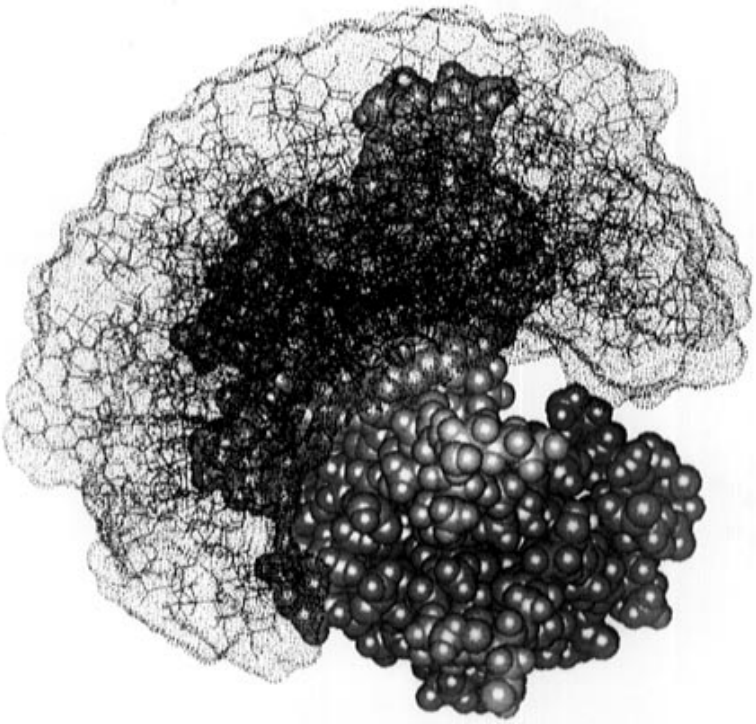


**Figure 29.** The effect of flexibility of the Asn-34 side chain on the orientation of the oligosaccharides attached to ribonuclease B. The Man-9 glycoform of RNase B based on the 2.5 Å X-ray crystal structure with an overlay of 10 oligosaccharide conformations from a 500 psec MD trajectory of Man-9. The total Van der Waals surfaces of the oligosaccharides are illustrated by dots.

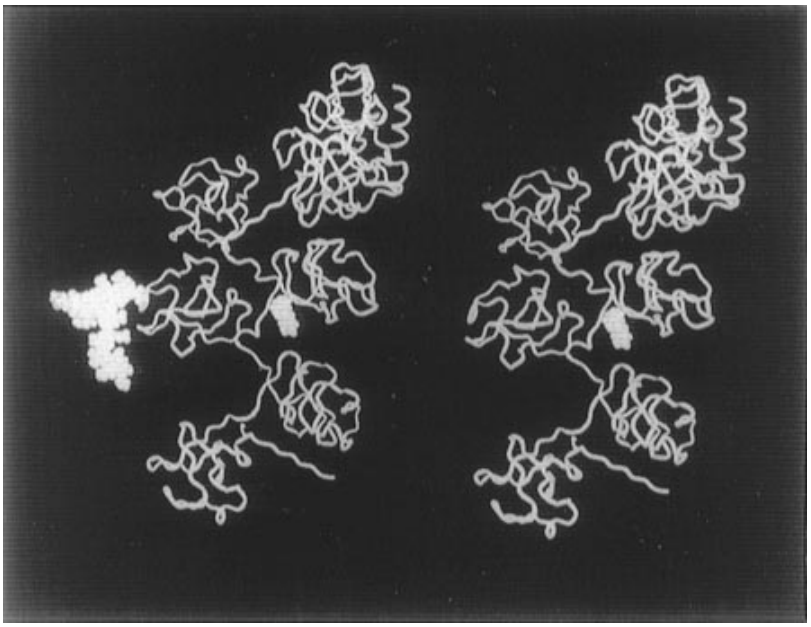


**Figure 30.** A schematic molecular model of plasminogen type 1 and type 2. Plasminogen consists of five kringle regions and a serine protease domain. This model was constructed using the coordinates from the corresponding domains on tPA and is intended to convey the relative size of the oligosaccharides and the protein. Type 1 plasminogen (left) has two occupied glycosylation sites, at Asn289ArgThr in kringle 3 and at Thr 345 in kringle 4. Type 2 (right) lacks the N-linked sugar at Asn 289.

sequential enzymatic digestion with exoglycosidases combined with Bio-Gel P-4 column chromatography, and 1D and 2D $^1$H NMR spectroscopy. In each case, C-terminus of the protein is linked by ethanolamine phosphate to a glycan with the conserved backbone sequence Manα1−2Manα1−6Manα1−4GlcNH$_2$, which in turn is linked to the sixth position of the *myo*-inositol ring of phosphatidylinositol.

The conserved sugar sequence of the anchors may be a consequence of their biosynthesis in which a preassembled core is added to a protein. The general principle is much the same for the biosynthesis of N-linked oligosaccharides on proteins. In both cases variations on the respective cores may be protein and cell specific. In VSG this tetrasaccharide backbone is substituted with branched side chains of α-galactose. It has been speculated that the arrangement of the GPI anchor of the VSG protein may allow for the close packing of these molecules on the parasite surface, with the heterogeneity of the glycan part of the GPI anchor allowing close packing in 2D space in order to create a surface. The solution structure[94] showed that the glycan exists in an extended configuration along the plane of the membrane spanning an area of 600 Å$^2$, which is similar to the cross-sectional area of the monomeric N-terminal VSG domain (Figure 33).

By contrast, computer modeling of the structure of Thy-1 and its GPI anchor, suggests that the lipid part of the GPI may not be the sole membrane anchor.[95] The model suggests that the glycan part of the GPI may lie within the Thy-1 protein moiety so that the Thy-1 protein moiety sits directly on the membrane with most of the GPI actually within the protein (Figure 34a,b). The smaller anchor could therefore allow access to Thy-1 by other "accessory" molecules within the membrane.

## 7.2. Structure/Function Relationships in IgG

X-ray crystallography has shown that each region of homology in the IgG molecule corresponds to a compact, independently folded unit and that these are linked together by short sections of polypeptide chain. Each domain consists of two $\beta$-pleated sheets with antiparallel strands connected by loop regions (the immunoglobulin fold).[96] Crystallographic studies of IgG Fc fragments[97,98] have shown that, unlike other immunoglobulin domains, the two C$_H$2 domains do not form extensive lateral associations. The resulting interstitial region accommodates the complex oligosaccharides, which are attached to Asn-297 on each heavy chain (Figure 35). In the Fc fragment for rabbit IgG one of the α-1,3 arms interacts with the trimannose core of the opposing oligosaccharide, while the other α-1−3 antenna extends toward the interface with the C$_H$3 domain. In the human Fc only the first two monosaccharide residues (Man,GlcNAc) of the 1−3 antenna are defined, but if the next galactose residue is built into the structure it would appear that this residue may form interactions with Thr-335, Ile-336, and Ser-337, on the protein surface. However, in both the human and rabbit Fc the α(1−6) antenna of each oligosaccharide is well defined and interacts with the hydrophobic and polar residues Phe-243 (Man-5 and GlcNAc-6), Pro-244/245 (Gal-7), and Thr-260 (GlcNAc-6 and Gal-7) on the domain surface.

The effects of glycosylation may be subtle and highly specific. Both the C1q and protein A binding sites on IgG are located in the C$_H$2 domain between Phe-319 and Ile-332 (distal or at the carboxyl-terminal side from the N-linked glycosylation site). Neither of these functions is markedly affected by the absence of carbohydrate.[99,100] The binding of monocytes, which involves sites 234−237 (proximal to the amino terminus) on the lower hinge region, is elimi-

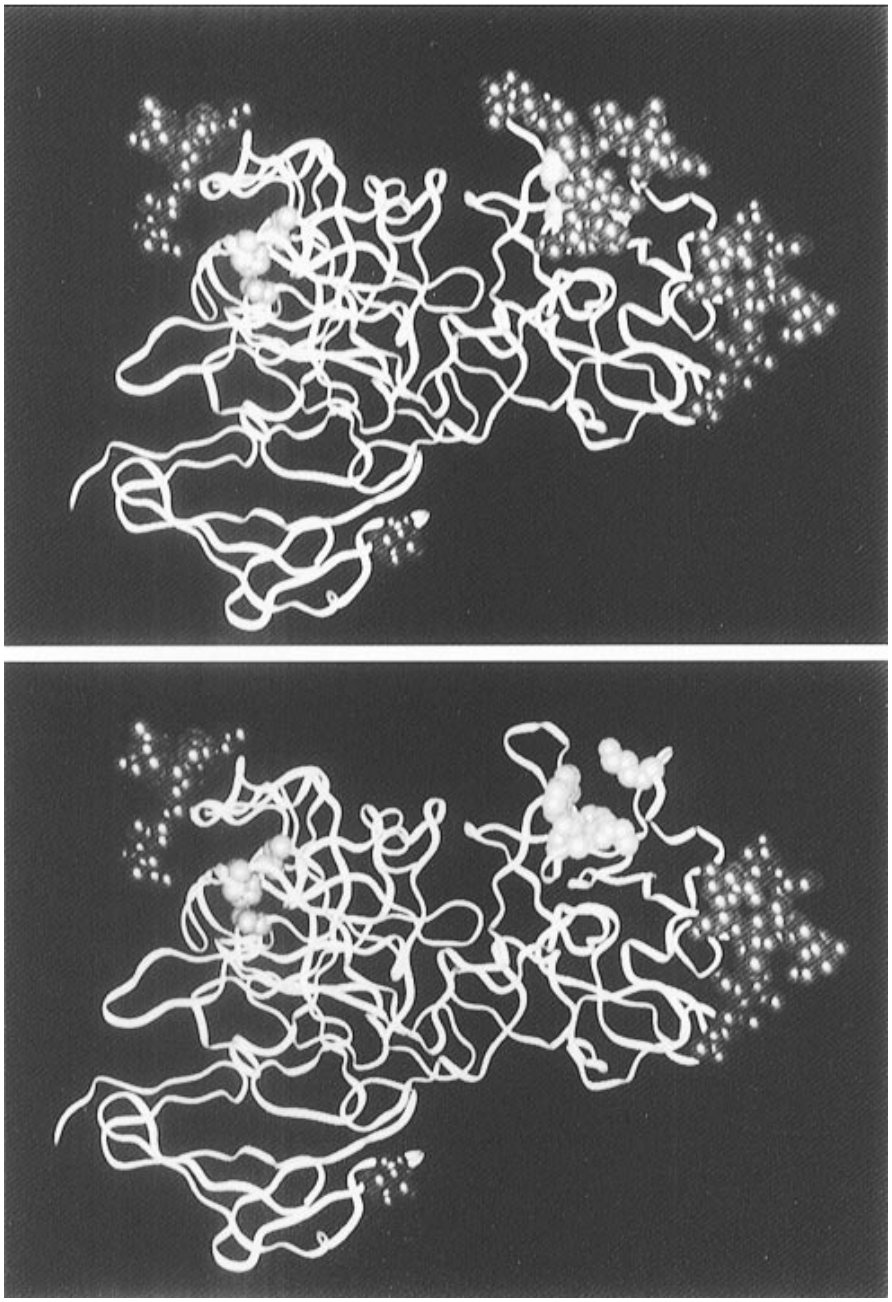


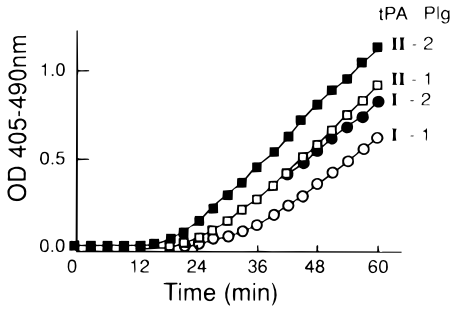
**Figure 31.** Schematic model of tissue-type plasminogen activator types I and II. tPA is composed of five domains: a fibronectin type 1 finger module, and EGF-like module, two kringles, and a serine protease domain. This model was constructed using the coordinates of the finger growth factor pair (Smith, B. O.; Downing, A. K.; Campbell, I. D.) and kringle 2[168] from human tPA. Kringle 1 and the serine protease domains were modeled by homology. The high-mannose carbohydrate at position 117 and the complex sugars at sites 184 and 448 are shown in green. The lysine binding site and the catalytic triad are highlighted in red and yellow, respectively. Comparison of the glycosylation of type I and type II tPA suggests that glycosylation at site 184 may sterically interfere with lysine binding.

nated in aglycosylated monoclonal murine IgG. This is consistent with a proposal that the absence of sugars results in a lateral movement of domains in the hinge region relative to the normally glycosylated antibody. Moreover, in aglycosylated IgG a protein structural change has been detected by $^1$H NMR at His-268,[101] which is also in the vicinity of the lower hinge. These data suggest that the oligosaccharide may stabilize a particular hinge conformation essential for monocyte binding. The spatial relationship between the $C_H2$ and $C_Hf$ domains, on the other hand, does not seem to depend on presence of the oligosaccharides, because in their absence protein A and C1q binding are unaffected.

## *8. Oligosaccharide Recognition*

### 8.1. Specific Interactions with Animal Lectins

To decode the information present in oligosaccharide structures, these must be recognized by other molecules. Carbohydrate-binding proteins are termed lectins. Following recognition of the oligosaccharides with the appropriate geometry (see later), these




Kinetic constants of activation of different glycoforms of plasminogen by tPA variants

| tPA | Plasminogen | Km (nM) | SD | kcat ($sec^{-1}$) | SD | kcat/Km ($\mu M^{-1} \cdot sec^{-1}$) |
|---|---|---|---|---|---|---|
| I | 1 | 81.8 | 4.7 | 0.12 | 0.01 | 1.5 |
| | 2 | 47.5 | 2.9 | 0.11 | 0.01 | 2.2 |
| II | 1 | 85.6 | 7.0 | 0.19 | 0.02 | 2.2 |
| | 2 | 51.1 | 2.3 | 0.17 | 0.03 | 3.3 |

*(n=4)*

**Figure 32.** The kinetics of the activation of type 1 and 2 plasminogen by tPA variants I and II in the presence of fibrinogen fragment and their kinetic contents.

lectins can mediate many specific biological functions. These include immune defense (e.g. the mannose-binding protein and the macrophage mannose receptor), clearance of glycoproteins (e.g. the asialoglycoprotein receptor), and cell–cell adhesion (e.g. the selectins). This important and major functional aspect of the glycobiology of animal lectins derives mainly from the pioneering work of Kurt Drickamer.[102] No review of glycobiology could be complete without acknowledgment of the fundamental importance of this field, and the readers attention is drawn to the prime reviews in this area.[5,103–105] Approaching 100 lectins have been isolated and many of these classified on the basis of sequence homology. By far the largest and most diverse class is that of the C-type lectins. These lectins bind carbohydrates in a $Ca^{2+}$ -dependent manner and are characterized by the presence of carbohydrate-recognition domains (CRDs). These domains contain a common sequence motif of approximately 120 amino acid residues and are characterized by 31 invariant or highly conserved amino acids.[106] Some of the C-type CRDs which have been reported are shown in (Figure 36).

One C-type lectin which we will refer to in section 9 is the mannose-binding protein (MBP) which mediates antibody-independent binding of pathogens which contain a high concentration of mannose or *N*-acetylglucosamine residues on their surface. The recognition event can lead to complement fixation or opsonization.[107] Recent X-ray data of the CRD from this protein illustrates the principles involved in the recognition of oligosaccharides. The CRD shows specificity for terminal residues and involves the sugar chelating to the $Ca^{2+}$ ion via the 3- and 4-hydroxyl groups[108] (Figure 37). There are also interactions between the CRD and the monosaccharide residues in the oligosaccharide that are mediated by water molecules, effectively increasing the surface area of the oligosaccharide in contact with the CRD. C-type lectins and their constituent CRDs display weak affinity for monosaccharides. In general the triggering of biological events requires multivalent receptors interacting with multivated targets (oligosaccharides). Multivalency in the receptor can be achieved by clustering CRDs. In the mannose-binding protein for example this occurs by the formation of oligomers of polypeptide chains, each of which contains a single CRD. In contrast the macrophage mannose receptor contains multiple CRDs in a single polypeptide. Clustering of CRDs could also arise from having multiple copies of the lectin in close proximity. (This may be the case for the selectins.) Similarly, the oligosaccharide may be multivalent by virtue of its branching if there is the "correct" geometry to bind to the receptor CRDs, or again there may be multiple copies appropriately presented on a surface.

The recent X-ray structures of the MBP trimer[3,4] can be used to illustrate the above points. The CRDs in the trimers are separated by some 44–53 Å. Matching multivalency by an oligomannose oligosaccharide to the trimer array can only be obtained if the terminal residues have the correct geometry in

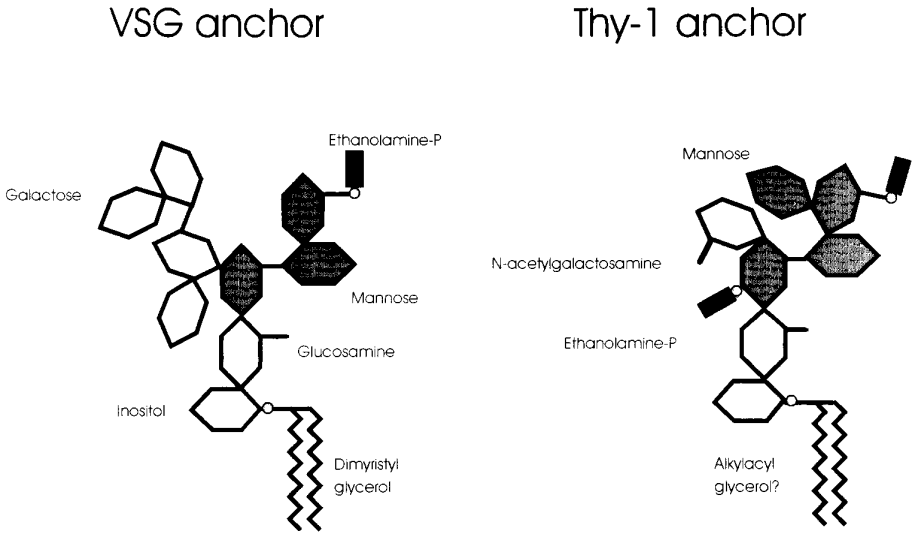



**Figure 33.** Comparison of the structures of the VSG and Thy-1 glycan anchors.

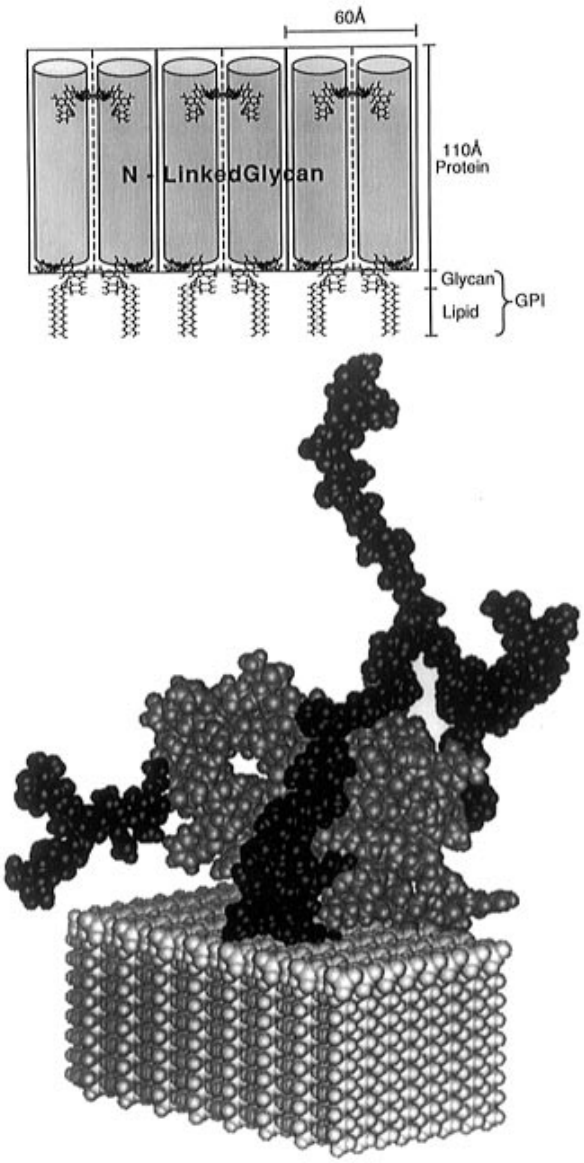



**Figure 34.** Molecular models of VSG (a) and Thy-1 (b) illustrating the potential interactions of the anchors and their proteins. In the case of VSG, the anchor is proposed to serve a space-filling role and spans an area along the membrane which is comparable to the cross-sectional area of the proteins N-terminal domains. In contrast the glycan part of the GPI anchor of Thy-1 may fit into a pocket in the protein with only the terminal part of the lipid extending, thus allowing significant direct interactions between the Thy-1 peptide and the membrane surface. Adapted from refs 95 and 169.

that they must also be separated by similar spacings of 44−53 Å. For instance, the Man-9 structure has three terminal monosaccharides each potentially able to bind to a CRD in the receptor. NMR studies, together with molecular dynamic calculations on Man-9,[109] have shown that in Man-9, the maximum distances between them is approximately 21 Å. The consequence is that the Man-9 structure can be bound via only one terminal residue to MBP, but this monovalent binding will not result in the biological triggering of complement. (This would only result from multiple presentation of Man-9 structures—as on a pathogen—with the appropriate spacings between the individual oligosaccharides to bind to the

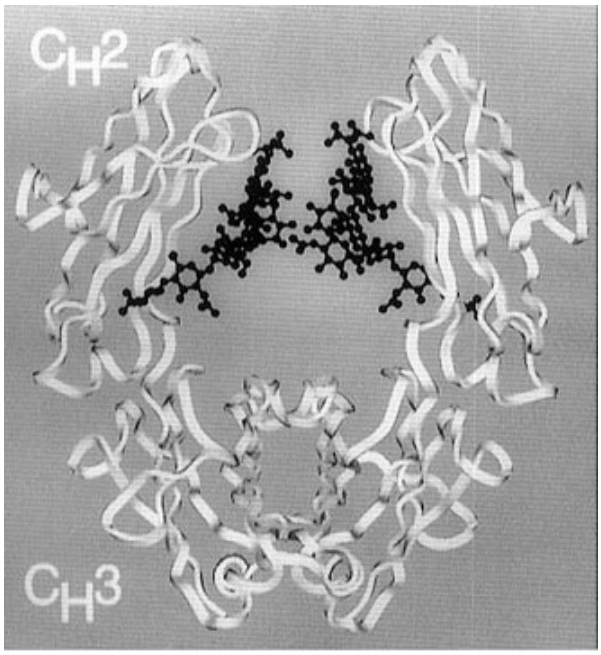



**Figure 35.** The Fc portion data from ref 97 of the IgG molecule.

different CRDs.) In this way the MBP molecule can "inspect" a range of molecules in the serum. Those with a single oligomannose structure will be bound and released, while these multiply presented as on a pathogen will lead to the biological effector functions being triggered. This type of mechanism has all the hallmarks of the discrimination in the immune system between self and nonself. Another receptor which is believed to form part of a basic defense mechanism against pathogens, is the macrophage mannose receptor.[5,110,111] This receptor can also mediate the clearance of endogenous proteins such as tissue plasminogen activator (via the oligomannose structures at site Asn-117) and lysosomal enzymes which carry high mannose-type oligosaccharides. Such proteins are often released into the circulation in response to pathological events. The macrophage mannose receptor internalizes the bound ligands and targets these to lysosomes for destruction.

## 8.2. Neural Glycosylation and Recognition

Glycosylation is a common feature to molecules implicated in cell−cell and cell−matrix interactions in the development and maintenance of the nervous system, as is involved, for example, in axon growth, guidance, and targeting. There is now increasing evidence to suggest that oligosaccharides play key recognition roles in these processes. A knowledge of oligosaccharide structures involved is essential for an understanding of carbohydrate-mediated interactions at the molecular level.

Technology developments which will help this field include obtaining milligram quantities of oligosaccharides from the glycoproteins of whole brain by hydrazinolysis. These sugar libraries provide a means of profiling brain oligosaccharides and determining the relative abundance of different classes. Identification of unusual or novel structures would

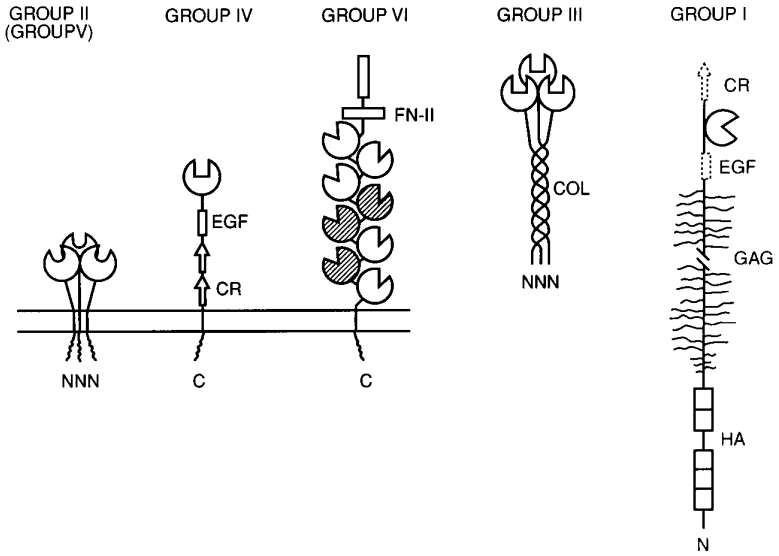



**Figure 36.** Summary of the structures of several groups of C-type animal lectins. Representative structures for three groups of membrane-associated lectins are shown: group II, the chicken hepatic lectin (homologue of the mammalian asialoglycoprotein receptor); group IV, the selectin cell adhesion molecules; and group VI, the macrophage mannose receptor. Group III lectins (collectins), such as mannose-binding protein, are found in extracellular fluids. Group I CRD-containing proteins are proteoglycans of the extracellular matrix. Other domains present in these molecules include EGF, epidermal growth factor-like repeats; CR, complement regulatory domains; FN-II, fibronectin type II repeats; COL, collagen-like sequences, GAG, glycosamino-glycan attachment sites; and HA, hyaluronic acid-binding domains. (Reprinted from ref 170. Copyright Cold Spring Harbor Laboratory.)

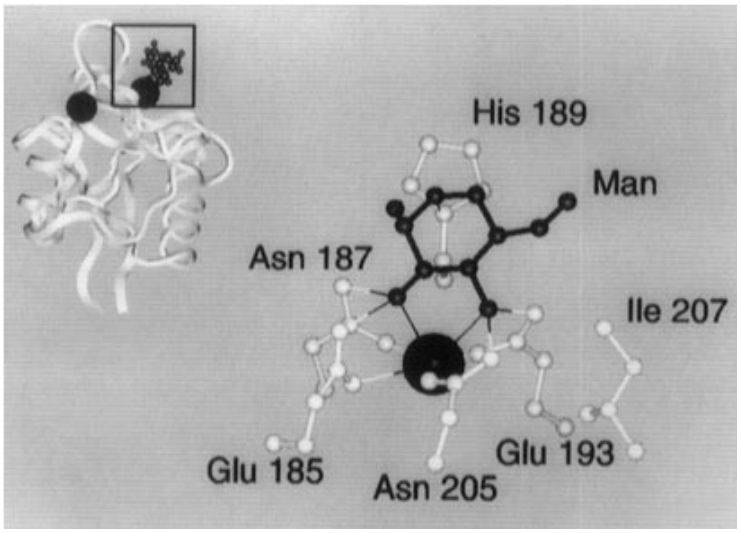



**Figure 37.** Structure of the C-type CRD from mannose-binding protein.[108] The boxed area is enlarged to illustrate the details of the interaction with mannose.

suggest different biosynthetic pathways from glycoproteins in the systemic side. Further development and integration of technologies for the sensitive detection and chromatographic separation of oligosaccharides are being pursued.

One way to probe for functional effects of glycosylation is to test glycans from the brain libraries as specific competitive inhibitors in neurite outgrowth assays *in vitro*. Preliminary work (Chen, Y. J.; Doherty, P.; Walsh, F.; Wing, D.; Dwek, R. A., unpublished work) has shown that crude sugar fractions from a rat brain library (at a concentration of 500 $\mu$M) can inhibit neurite outgrowth over astrocytes—a process involving recognition between cell adhesion molecules.

It is well known that carbohydrates are enriched at the synapse, and as recognition roles for the associated glycoproteins are becoming established, the exciting possibility exists for a glycoconjugate role in synaptic efficacy, a field with far-reaching implications in such areas as memory formation.

### 8.3. Major Histocompatibility Complex (MHC) Restricted Recognition of Glycopeptides by T-Cells

The major histocompatibility complex (MHC) class 1 molecules display peptides from self or foreign cellular proteins, on the antigen presenting cell surface. A critical event in the immune response is the recognition of these peptide—MHC complexes by

cytolytic T-cells. These identify and subsequently eliminate cells that present "foreign" peptides.

An interesting and exciting finding of carbohydrate-specific recognition has been the demonstration that T-cells recognize glycopeptides bound to MHC class I molecules. Harum et al.[112] have reported the efficient binding to class I MHC of a synthetic glycopeptide carrying the naturally occurring O-$\beta$-linked *N*-acetylglucosamine (GlcNAc). The peptide used in the investigation was an analogue of the epitope FAPGNYPAL from Sendai virus nuclear protein, which produces a specific cytotoxic T-cell (CTL) response. The glycosylated analogue (an O-GlcNAc on the serine residue) is highly immunogenic and elicits a specific, MHC-restricted, anti-glycopeptide CTL response.

The relevance of studying T-cell responses to modifications such as glycosylation finds support from the preliminary data obtained in which the presence of O-GlcNAc modified peptides has been detected among the pool of peptides eluted from human spleen class I MHC.[112]

Interestingly, a study by Michaelsson et al.[113] also suggested that T-cell recognition of glycosylated peptides from type II collagen may be important in the development of rheumatoid arthritis. Recognition of glycosylated peptides could therefore be of significant importance in immunity toward malignant diseases and in viral infections. It is also possible that acquisition or loss of glycosylation might provide a novel strategy for viruses or malignantly transformed cells to escape from selection pressure by CTL which does not depend on a mutation at the gene level.

### 8.4. Recognition of Oligosaccharides by Stimulated T-Cells

A quite different type of T-cell recognition involving oligosaccharides may exist in diseases such as rheumatoid arthritis, Behcet's syndrome, and IgA nephropathy. In these diseases up to 9% of the T-cells are bound by cytophilic IgA1 compared with none in resting T-cells from healthy subjects. Detailed studies have shown that the binding to this subset of T-cells can be inhibited by the "O"-linked sugars associated with the hinge region of IgA1.[114] While the functional significance of the receptor has yet to be established, the observations suggest that the oligosaccharides indicated recognition may have functional implications in these diseases.

Although the general area of T-cell recognition of saccharides is a difficult and challenging one, it is an important one which is well worth pursuing and which could provide yet more insights into the roles of oligosaccharides in the cellular immune response.

## *9. Glycosylation in Disease*

The glycosylation of some glycoproteins has been shown to alter in disease. For example, concanavlin A lectin chromatography suggests that the relative amounts of complex biantennary glycans on α-1 acid glycoprotein decreases during the acute phase response.[184] Another well-documented example which is discussed below is the IgG molecule, the glycosylation of which is altered in rheumatoid arthritis.

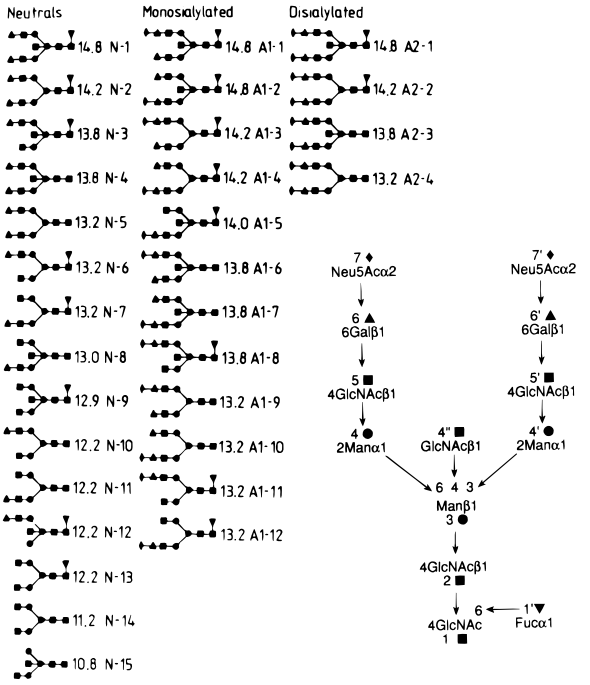



**Figure 38.** Primary sequences of the N-linked oligosaccharides associated with IgG. The hydrodynamic volume of each structure, measured in glucose units, is indicated.

### 9.1. The IgG Molecule

The IgG molecule contains 2.5 oligosaccharide chains per molecule. Two of these represent the conserved glycosylation sites in the Fc portion of all IgGs at Asn-297. The remainder occur in the hypervariable regions of the Fab fragment with a frequency and position dependent on the occurrence of an N-glycosylation site (Asn/Xaa/Ser(Thr). The glycosylation is of the complex biantennary class and about 30 variants occur in the IgG molecule (Figure 38). This glycosylation of IgG results in many different glycoforms.

#### *9.1.1. Site-Specific Glycosylation of IgG*

There are clear glycosylation site-specific differences between the Fab and Fc, with regard to sialylation and the presence of bisecting GlcNAc residues (both mainly in the Fab fragment). Characteristics of Fc N-glycosylation include a low incidence (10%) of monosialylated structures, the absence of disialylated structures, a low incidence of cores carrying a "bisecting" GlcNAc, and heterogeneity in the galactose residues. In general, the Fc oligosaccharides are mainly restricted to biantennary oligosaccharides terminating in 2, 1, and 0 galactose residues (G2, G1, and G0). The unique 3D environment in the Fc may limit the accessibility of sugar processing enzymes so resulting in these glycoforms. Fab N-glycosylation is characterized by a high incidence of di- and monosialylated structures, and of cores with the "bisecting" GlcNAc residue.[115] It should be stressed that the large number of different structures associated with IgG is not the result of studying a polyclonal population, since a similar heterogeneity is found upon analysis of myeloma and hybridoma-derived IgG.

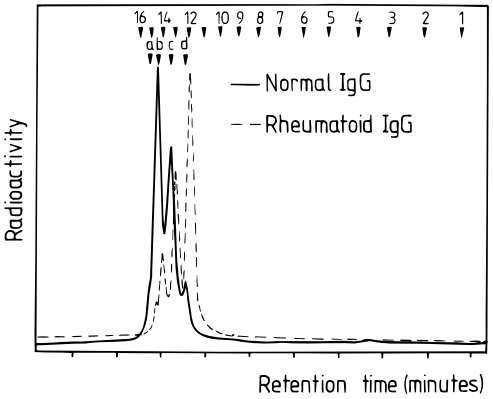



**Figure 39.** Representative Bio-Gel new P-4 (−400 mesh) gel permeation chromatogram of the asialooligosaccharides of total serum IgG from a healthy individual and a patient with rheumatoid arthritis.

### *9.1.2. IgG Glycoforms Associated with Rheumatoid Arthritis*

Serum IgG from patients with rheumatoid arthritis contains the same set of biantennary oligosaccharides found in normal individuals, but in very different proportions (Figure 39). The incidence of structures with outer arm galactose is dramatically decreased, and the incidence of those structures terminating in outer arm *N*-acetylglucosamine correspondingly increased.[8] A comparison of the N-glycosylation of Fab and Fc fragments, derived from total serum IgG of patients with rheumatoid arthritis or from a control group, shows that the decreased galactosylation found upon analysis of total serum IgG is largely due to changes in the N-linked oligosaccharides of the Fc. There are also quantitatively minor, but potentially significant differences in Fab glycosylation (Scragg, I.; Chen Chang, S., unpublished work), in which there is an *increase* in the light chain in the proportion of fully galactosylated (G2), bisected, and core fucosylated oligosaccharides.

The cause of the altered glycosylation of IgG in RA is not clear at present. It may have a biosynthetic origin[116] in which there is decreased levels of the galactotransferase (GalTase) activity. Indeed, in peripheral blood B and T lymphocytes isolated from RA patients, there is a decrease in this activity[117,118] which correlates with the overall decreased galactosylation of serum IgG. In general, such a correlation is expected to be insensitive to the Fab glycosylation since it only contributes a small amount to the overall value. However, the detailed site-specific changes of ref 119 in which the Fab and Fc glycosylations are quite differently affected cannot be explained by an overall decrease in the GalTase activity. Neither does it explain the normal levels of galactose detected on rheumatoid serum $IgA_1$ when compared with $IgA_1$.[120] Analysis of oligosaccharides released from IgG and its fragments does not give any information on the combination of the sugars present on the molecule. Thus, it is entirely possible that RA IgG which has increased L chain galactosylation is synthesized by a different population of B cells than those molecules which have decreased Fc fragment glycosylation. Since B cells have differing levels of galactosyltransferase activities[121] this provides a straightforward explanation for site-specific galactosylation. Changes in the population of B cells may be inferred from the finding that there is an increase in the number of B lymphocytes which are $CD5^+$ [122,123] and the discovery that IgG subclass distribution is altered in rheumatoid arthritis.[124] Alternatively, there is the possibility that multiple galactosyltransferases of different protein (and hence potentially IgG fragment) specificity are present in B cells. An IgG-specific galactosyltransferase of reduced activity in RA has been reported[118,125] but as yet no other evidence has confirmed this interesting finding. Destruction of galactose caused by reactive oxygen species has also been proposed to contribute to lower levels of galactose in rheumatoid arthritis[126] but it is difficult to envisage how this mechanism could lead to the site-specific changes between the Fab and Fc fragments. Clearly, more work is necessary to understand the mechanisms which regulate the glycosylation of human IgG.

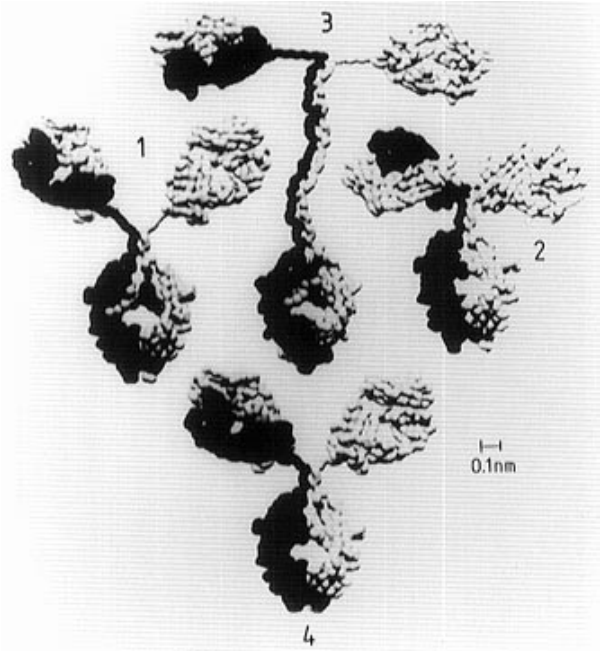



**Figure 40.** Representation of the four human immunoglobulin subclasses. (Reprinted from ref 171. Copyright 1990 Pergamon Press.)

There are four human subclasses of IgG (Figure 40). The most striking difference between them lies in the hinge regions. Thus the hinge region of IgG1 consists of 23 amino acids, that for IgG2−19, that for IgG3−70 and that for IgG4−20 (see Table 1). In serum the mean concentration (in mg/mL) is −9 for IgG1, 3 for IgG2, 1 for IgG3, and 0.5 for IgG4. Table 2 shows that in RA the changes in glycoforms of IgG are seen across all four subclasses.

### *9.1.3. Glycosylation Changes on the IgG Molecule Are "Disease Restricted" and Are an Important Factor in Rheumatoid Arthritis*

The change in galactosylation of the serum IgG of patients with rheumatoid arthritis is not common to all other autoimmune or inflammatory disorders. On the other hand, agalactosyl IgG has been consistently found in patients with juvenile rheumatoid arthritis, Crohn's disease, and tuberculosis.[127,128]

The Fc glycoform distribution changes with age (Figure 39), with disease severity of rheumatoid

**Table 1. Amino Acid Sequences of Human IgG Hinge Regions[a]**

| | upper hinge 216 | middle hinge | lower hinge 231 |
|---|---|---|---|
| IgG1 | EPKSCDKTHT | CPPCP | APELLGGP |
| IgG2 | ERK | CCVECPPCP | APPVA GP |
| IgG3 | ELKTPLGDTTHT | CPRCP (EPKSCDTPPP-CPRCP)$_{x3}$ | APELLGGP |
| IgG4 | ESKYGPP | CPSCP | APEFLGGP |

[a] The upper and middle hinge are encoded by one exon in the case of IgG1, 2, and 4 and four exons for IgG3. The lower hinge is encoded by the $C_H2$ exon. The numbering system refers to the amino acid sequence position of the beginning of the upper and lower hinges.[165]

**Table 2. Galactosylation of Total Human Serum IgG and Its Component Subclasses (%) of G2, G1, and G0 Structures**

| sample[a] | galactosylation[b] | serum IgG | subclasses IgG1 | IgG2 | IgG3 | IgG4 |
|---|---|---|---|---|---|---|
| H1 | G2 | 45.0 | 50.9 | 38.1 | 61.7 | 40.8 |
| | G1 | 34.0 | 33.7 | 38.6 | 23.6 | 34.4 |
| | G0 | 21.0 | 15.4 | 23.4 | 14.7 | 24.8 |
| H2 | G2 | 47.6 | 52.5 | 37.8 | 57.5 | 46.4 |
| | G1 | 38.2 | 41.1 | 41.5 | 30.1 | 36.7 |
| | G0 | 14.2 | 6.4 | 20.6 | 12.4 | 16.9 |
| H3 | G2 | 39.2 | 54.2 | 37.1 | 61.2 | 38.3 |
| | G1 | 41.7 | 34.4 | 38.2 | 27.6 | 26.2 |
| | G0 | 19.1 | 11.4 | 24.7 | 11.2 | 25.5 |
| R1 | G2 | 19.4 | 26.4 | 18.6 | 23.6 | 19.3 |
| | G1 | 35.5 | 35.0 | 31.3 | 29.6 | 32.4 |
| | G0 | 45.0 | 38.6 | 50.3 | 46.8 | 48.3 |
| R2 | G2 | 18.7 | 19.1 | 13.8 | 30.7 | 16.4 |
| | G1 | 30.2 | 27.4 | 25.5 | 23.2 | 25.5 |
| | G0 | 51.1 | 53.5 | 60.7 | 46.1 | 58.1 |
| R3 | G2 | 14.2 | 13.9 | 10.8 | 20.6 | 15.3 |
| | G1 | 30.8 | 26.5 | 20.4 | 28.4 | 20.7 |
| | G0 | 55.0 | 59.6 | 68.8 | 51.0 | 64.0 |

[a] H1−H3 refer to three nonarthritic individuals while R1−R3 are three patients with rheumatoid arthritis. [b] Oligoaccharides were assessed to be di- (G2), mono- (G1), or non- (G0) galactosylated. Scragg, I.; Youings, A.; Dwek, R. A. Unpublished data.

arthritis, and with pregnancy. The increase and decrease may reflect the control of galactosyl transferase activity under different physiological conditions. Clearly, rheumatoid arthritis is associated with changes in G0 glycoforms so that this parameter is a good biochemical disease marker of diagnostic value (Figure 41). In an arthritic woman with pathologically elevated levels of G0 glycoforms, changes in G0 correlated with remission of arthritis during gestation and post-partum recurrence.[129]

That the G0 glycoforms may be an important factor in rheumatoid arthritis can be demonstrated in the collagen-induced arthritis (CIA)[174] mouse model. Passive transfer of an acute synovitus in T-cell primed mice can be achieved using IgG containing autoantibodies to type II collagen only when these autoantibodies are present in the G0 glycoforms. The basic protocol is given in Figure 42. IgG was purified from pooled sera of mice with CIA at days 17 and 38 when peak levels of type II collagen autoantibodies were produced. The IgG fractions were isolated and divided. One fraction was treated with $\beta$-galactosidase from *Streptococcus* strain 6646K as to generate

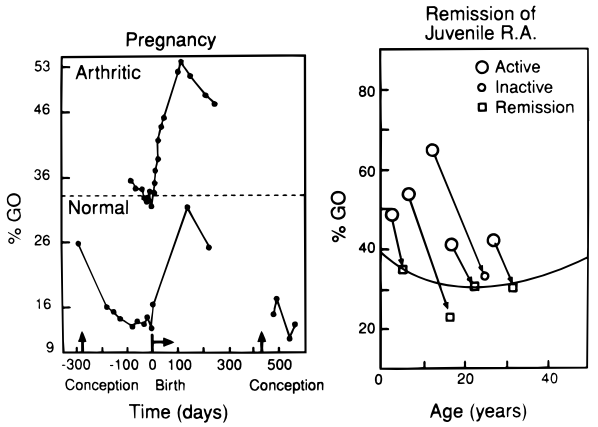



**Figure 41.** Variations in the galactosylation of IgG. There are changes with age and disease activity as illustrated for juvenile arthritis. A comparison between the changes in percentage G0 during the course of pregnancy for a normal and rheumatoid arthritis patient, shows that both patients tend to increase their galactose levels during pregnancy. The dashed line indicates the aged matched expected value of a normal healthy individual. It is seen that the pregnant patient with rheumatoid arthritis achieves this level and this correlates with a remission of arthritis in the patient.

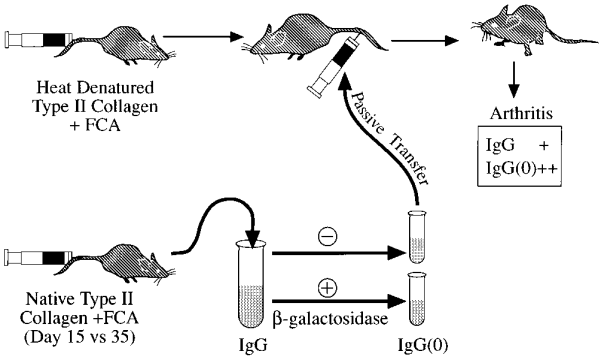



**Figure 42.** Protocol for passive transfer of IgG in the collagen-induced arthritis model in mouse. Purified IgG is transferred to a suitably primed mouse or is first treated with $\beta$-galactosidase to enrich the IgG0 glycoforms.

the IgG (G0) glycoforms exclusively. Other IgG samples contain a distribution of glycoforms only some of which are the G0 glycoform. Thus by increasing the level of agalactosyl IgG glycoform of the anti-type II collagen antibodies present in the preparation, the IgG becomes more effective in causing arthritis in passive transfer. This indicates that the glycosylation status of an autoantibody is one factor in determining if an antibody is pathogenic.

### *9.1.4. Structural Changes in IgG Fc on Loss of Galactose*

NMR relaxation data from a control (26% G0) and RA (55% G0) sample show a very slowly relaxing (and hence a more mobile) subset of proton resonances between 3.5 and 4.25 ppm (Figure 43) which can be assigned to oligosaccharide protons. This subset is characterized by T2 relaxation times of ~0.15 s. This compares to the values of 0.005 s for the protein backbone and static aromatic side chains, 0.07 s for ring flipping aromatic side chains, and ~0.2−0.5 s for released oligosaccharides of a similar size. There are approximately 18 and 41 protons in this subset in the control and RA samples respectively. These

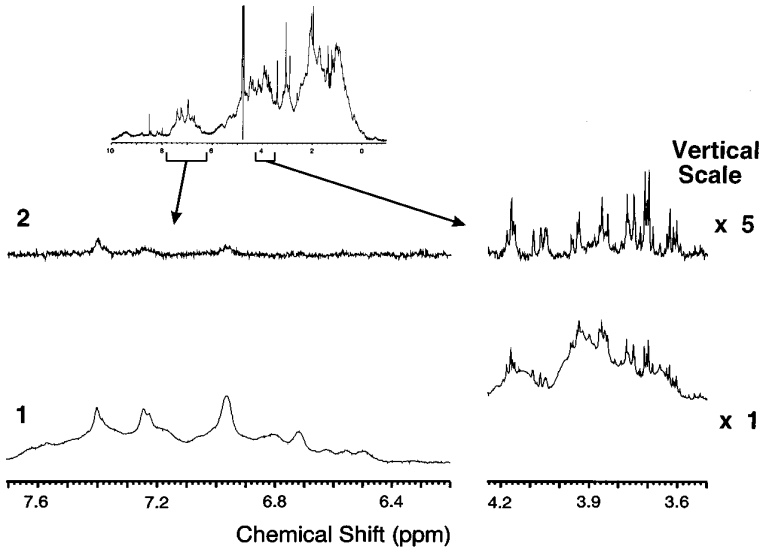



**Figure 43.** NMR relaxation weighted spectra of IgG Fc from a rheumatoid arthritic patient. The top insert shows the 500 MHz 1D $^1$H NMR spectrum of IgG Fc isolated from a rheumatoid arthritic patient. Selected regions of the CPMG spectra are shown at total spin echo times of 0.001 s (1) and 0.128 s (2). The first spectrum is essentially the 1D spectrum. The second spectrum shows only the subset of more slowly relaxing proton resonances.

are not enough to account for all the oligosaccharide protons in either sample and the remaining ones must have relaxation properties very similar to the peptide backbone. Since the 2-fold increase in the number of more mobile protons mirrors the two-fold increase in G0 values between the two samples, this allows assignment of these mobile protons to the G0 oligosaccharides. By taking into account the percentage occurrence of G0 in the two samples (and that there are two oligosaccharides per Fc), each G0 oligosaccharide must contribute on average 36 protons to the subset. This suggests that the protein only has a very small effect on the mobility of the G0 oligosaccharides. This analysis leads to a model of G1 and G2 oligosaccharides having restricted motion relative to the peptide surface (as seen in the crystal structure), while the G0 oligosaccharides are free to move (Figure 44). X-ray data for normal IgG Fc[97] and calculations of solvent accessibility (Figure 45) show that one of the main interactions between the protein surface and the oligosaccharide is mediated by interaction of the (1−6) arm terminal galactose residue with Thr-260. Loss of this galactose therefore results in a decreased interaction between the oligosaccharide and the protein surface. On this basis, it would be expected that G0 oligosaccharides and those of G1 with a galactose residue on the (1−3) arm would have extra mobility relative to the protein. The NMR analysis, however, indicates that no G1 structures have extra mobility. Further inspection of the X-ray data of the Fc shows a potential interaction site between the (1−3) galactose and residues Thr-335, Ile-336, and Ser-337, on the protein surface. Thus increased mobility of the oligosaccharide results from loss of both galactoses (Figure 45) and leads to exposure of the nonreducing terminal GlcNAc residues.

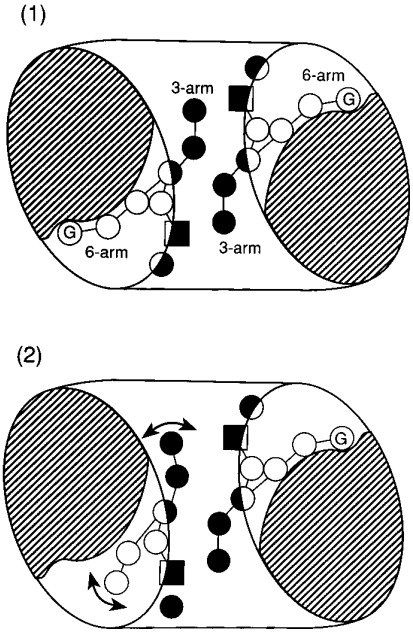



**Figure 44.** Schematic representation of the oligosaccharides of IgG Fc looking along the pseudo-$C_2$ axis. The squares show the positions of the GlcNAc 1 residues, the circles the positions of the other monosaccharide residues, G indicates a nonreducing terminal galactose residue on the 6-arm of the oligosaccharide. (1) Two G1 oligosaccharides (as seen in the crystal structure). Binding of the 6-arm to the protein surface holds the rest of the oligosaccharide rigid with respect to the protein. (2) A G0 (left) and G1 (right) oligosaccharide. Loss of interaction between the G0 oligosaccharide 6-arm and the protein surface leaves the whole of the oligosaccharide free to move.

### *9.1.5. Functional Implications of IgG Glycoforms*

The studies of the Fc fragment of X-ray crystallography indicate that galactose residues present on the α (1−6) arms can interact with the protein. An increase in the level of IgG glycoforms lacking terminal galactose and thereby terminating in *N*-acetylglucosamine could lead to the exposure of certain Fc determinants. This may elicit an immune response, or raise a preexisting subclinical response to a pathological one which may be relevant to rheumatoid arthritis. In addition, the now vacant galactose sites on the protein may create a lectin-

**Figure 45.** Structure of the IgG Fc $C_H2$ domain (the full IgG Fc structure is shown in the top left insert) showing part of the solvent accessible surface: (left) crystal structure[97] with a G1 oligosaccharide showing the protein/oligosaccharide interactions; (right) model structure with a G0 oligosaccharide after allowing the oligosaccharide to move within the pocket between the two $C_H2$ domains.

like activity in the IgG resulting in the formation of complexes or autoaggregates typical of the disease[130] without an actual autoimmune response. The contribution of such factors to the pathology of the disease has still to be assessed.

The Fc region of antibody molecules mediate interaction with many of the effector functions of the immune system following antigen binding. Of these, the complement system is a major immune defense mechanism. Inappropriate or chronic localized complement activation can cause severe damage to host tissue, and this is an important factor in the pathogenesis of several diseases.[131] The first step in the classical complement cascade is the binding of C1q to the $C_H2$ domains in the Fc region of the antibody, following antigen recognition. It involves surface "matching" of charged amino acid residues between C1q and Fc.[132,133] Activation of the classical complement pathway by a second route, which does not involve C1q, is mediated by serum lectin mannose-binding protein (MBP), and until now has been reported as antibody independent.[107,134] MBP binds terminal fucose, glucose, mannose, or *N*-acetylglucosamine (GlcNAc), but not galactose, residues through its carbohydrate recognition domains (CRDs).[108,135] These are connected to collagen-like domains, giving MBP an overall ultrastructure similar to C1q.[134,135] Serum MBP mediates antibody-independent recognition of pathogens which have a high concentration of mannose or GlcNAc residues on their surface. This recognition is followed by either direct opsonization or complement fixation. The NMR studies show that (GlcNAc) becomes exposed in those molecules in which the Fc oligosaccharides lack galactose and is a candidate for MBP binding. Further those glycoforms of IgG in which Fc carbohydrate terminates in GlcNAc and, when multiply presented, may activate complement via MBP. If so, this lectin route provides a mechanism for inducing inflammation above the normal background of recognition and clearance.

### *9.1.6. Modeling of the Possible Interaction between Agalactosyl IgG Fc and MBP*

The X-ray data[97] show that in galactosylated IgG Fc none of the monosaccharide residues is accessible for recognition by lectins. On loss of galactose, the 4-OH group of GlcNAc becomes available, allowing it potentially to interact with the CRDs from MBP.[108,136] Molecular modeling shows, however, that such interactions still cannot occur (due to protein–protein steric interactions) without displacement of the oligosaccharide from the position observed in the X-ray structure.[97] The NMR data show that such displacements occur spontaneously on loss of galactose, and molecular modeling indicates that these make either the 3-arm or 6-arm terminal GlcNAc residues available for binding to the CRD (Figure 46).

A CRD recognizes a single-terminal monosaccharide weakly.[108] Each CRD of MBP is connected to a collagen-like domain. The individual polypeptides of the protein associate into groups of three, forming a collagen triple helix attached to a cluster of three CRDs in a globular head. These trimeric building blocks associate, via the collagenous region, to form larger structures with up to six globular heads (i.e. with 18 CRDs, Figure 47). MBP recognizes pathogens containing high concentrations of mannose or GlcNAc residues on their surfaces.[108,134,185] Multiple interactions with multivalent targets are required for physiologically relevant binding. It is interesting to note from the recent X-ray data on the MBP CRD trimers from human[4] and rat[108] that the binding sites are separated by 4.4 and 5.3 nm, respectively. This is much too far apart for CRDs in the single trimer to bind multiple GlcNAc residues in a single IgG molecule (there are four terminal GlcNAc residues per agalactosyl IgG Fc). Therefore, multivalency can only arise from presentation of multiple IgG molecules. Activation of complement by MBP and IgG–G0 is likely to have similar spatial requirements as those seen with C1q–IgG interactions, i.e. at least

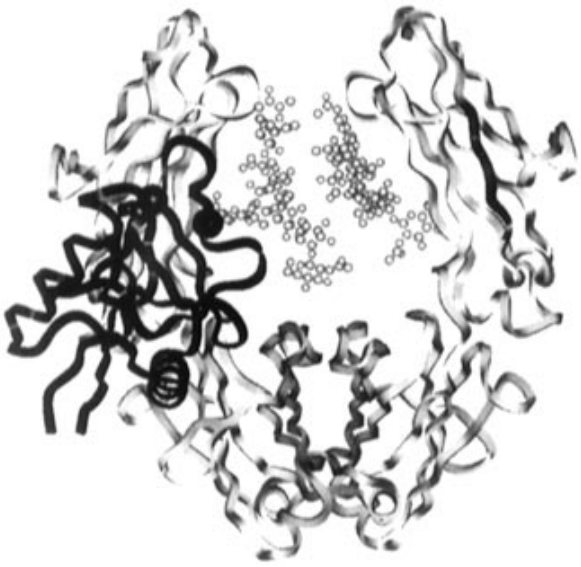


**Figure 46.** Theoretical model of the interaction between Fc[97] (light gray) and a single CRD from MBP[108] (black). The MBP interacts with the (1−6) arm terminal GlcNAc residue on a G0 oligosaccharide. [Note: The CRD can also interact with the 3-arm nonreducing terminal GlcNAc residue by an alternative displacement of the Fc oligosaccharide chain.] The C1q binding site on the right-hand of the Fc $C_H2$ domain[132,133] is also indicated (dark gray).

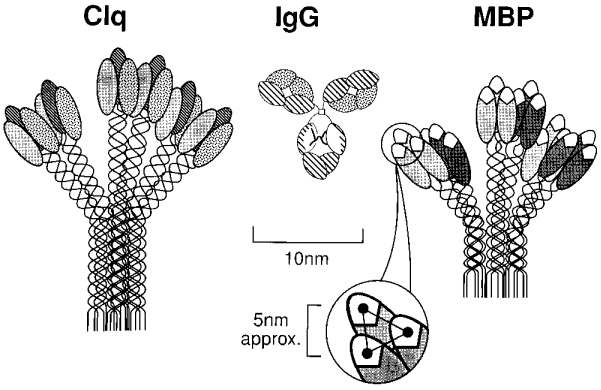



**Figure 47.** The relative sizes of C1q, MBP, and IgG based on refs 135, 172, and 173. C1q has six globular regions each consisting of three similar, but non-identical independently folding domains. In contrast the CRDs of MBP are identical. The spacing of carbohydrate binding sites (shown on insert) on MBP is between 4.5 and 5.3 nm.[3,4]

two globular heads of C1q (or MBP) must bind to immobilized IgG. In RA patients, autoantibodies carrying G0 structures may be localized in the joints bound to collagen and may thus present a multivalent surface to the MBP, similar to that presented by pathogens.

### *9.1.7. $Ca^{2+}$-Dependent Binding of MBP to IgG Is Mediated by the Agalactosyl Fc Glycoforms*

The results in Figure 48 show that there is an increase in specific MBP binding when normal IgG (G0 = 20%) is converted enzymatically to 100% IgG−G0. Comparison of the binding data for IgG−G0 and the corresponding Fab and Fc fragments with those of normal IgG and its fragments indicates that the increased binding arises from interactions between MBP and the Fc-associated oligosaccharides. Although normal IgG, its Fab and Fc fragments, and Fab−G0 also contain G0 oligosaccharide structures, the data suggest that their density and presentation to the multivalent MBP is insufficient to give rise to strong binding. In IgG it is, therefore, the alteration of the levels of the Fc glycoforms containing G0 structures that could modulate MBP binding to IgG. Figure 48b shows a representative IgG preparation from a rheumatoid patient (G0 = 36%) which exhibits increased binding of MBP.

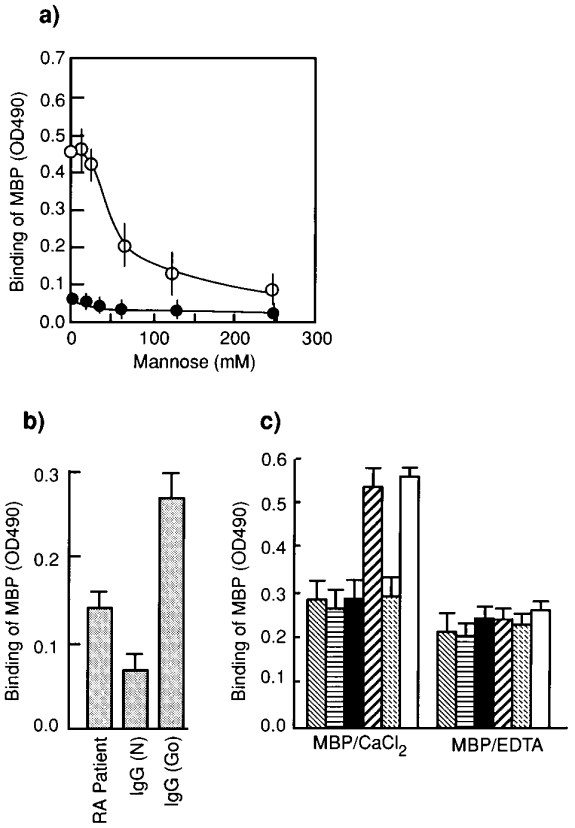



**Figure 48.** Interaction of MBP with IgG, IgG-G0, and their fragments. (a) The $Ca^{2+}$-dependent binding of MBP to IgG−G0 (○) or IgG (●) is inhibited in a concentration-dependent manner by mannose. The binding of MBP in the presence of EDTA (OD value of 0.3) was subtracted from the data to take account of nonspecific binding. The error bars represent the range of three different determinations. All experiments were done in triplicate. (b) The binding of MBP in the presence of $Ca^{2+}$ to normal IgG (20% G0), IgG from an RA patient (36% G0), and IgG−G0 (100% G0). The background value of OD = 0.06 was subtracted from the data to take account of nonspecific binding (in EDTA). (c) Glycoforms of normal IgG and Fc containing only G0 type sugars show a higher $Ca^{2+}$-dependent binding of MBP than the unmodified normal populations of IgG and Fc. The bars are from left to right Fc, Fab, IgG, Fc−G0, Fab−G0, IgG−G0.

It is important to stress that the assays involve binding IgG or its fragments to microtiter plates. IgG fixed on these plates is multiply presented so that the data indicate that multiple presentation of agalactosyl IgG leads to interaction with MBP.

### *9.1.8. MBP Activation of Complement by Agalactosyl IgG Glycoforms*

Figure 49 shows the activation of the complement system by MBP following binding to IgG or IgG−G0

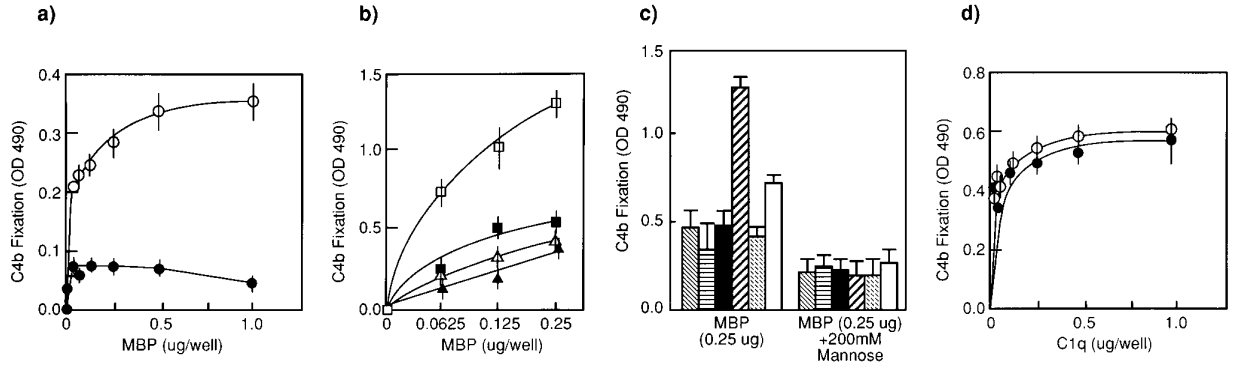



**Figure 49.** IgG−G0-induced activation of the complement system. (a) IgG−G0 (○) induces 5 times more deposition of C4b than IgG (●). Deposition of C4b is dependent on MBP concentration. The nonspecific binding was reflected in an OD of 0.18 which was subtracted from the data points. (b) Increase in C4b deposition occurs when Fc (■) is converted to Fc−G0 (□). There is no change in C4b deposition when Fab (▲) is converted to Fab−G0 (△). The background value of OD = 0.19 was subtracted from each data point. (c) Binding of MBP is through the oligosaccharide. A complement activation assay was performed as in Figures 6 and 7 except that MBP was incubated in the protein-coated wells in TBS−$Ca^{2+}$ or TBS−$Ca^{2+}$ containing mannose. The incubation of MBP with (from left to right) Fc or Fab or IgG or Fc−G0 or Fab−G0 or IgG−G0 in the presence of mannose decreased the deposition of C4b. A background value of OD = 0.19 was subtracted from each data point. (d) Conversion of IgG (●) to IgG−G0 (○) does not markedly alter the C1q-mediated deposition of C4b. Serial dilutions of C1q (100 $\mu$L; 10 $\mu$g/mL) were incubated in IgG−G0 or IgG coated wells and C1q-mediated deposition of C4b was measured as above. The background value of OD = 0.18 was subtracted from each data point in the IgG curve, while a value of 0.32 was subtracted from the IgG−G0 points to account for background due to the residual MBP present in the depleted serum.

immobilized on microtiter plates. The results in Figure 49a show that the amount of C4 activated and deposited was dependent on the concentration of MBP. At the saturation point the amount of C4 fixed was ca. 5-fold higher when MBP was complexed with IgG−G0 compared with normal pooled IgG. The results in Figure 49b show that the MBP-mediated activation of complement is induced predominantly by the interaction between MBP and the Fc/G0 and not the Fab/G0 fragment. Fixation of C4 was reduced considerably when MBP was incubated with IgG, IgG−G0, or their fragments in the presence of mannose, an inhibitor of binding of MBP to oligosaccharides (Figure 49c). This demonstrates that the activation by MBP is mediated through the IgG−G0 sugars. In contrast, the activation by C1q is independent of the glycosylation state of the IgG (Figure 49d), confirming that treatment with glycosidases has not altered the conformation of the IgG.

### *9.1.9. MBP and Agalactosyl IgG Are Present in Synovial Fluid*

The levels of G0 are elevated in the synovium compared with serum.[137,138] The presence of MBP in synovial fluid has been shown to be very similar to that in serum.[139] The occurrence of MBP in synovial fluid coupled with the presence of high levels of IgG−G0 structures suggests that the activation of complement by MBP could contribute to the chronic inflammation of the synovial membrane of affected joints. This would provide a link between the elevated levels of G0 structures on IgG found in RA and the onset of tissue damage and a role for G0 structures in the pathogenesis of the disease.

## *10. Glycosylation Inhibitors as Antiviral Agents*

The biosynthesis of N-linked oligosaccharides involves the cotranslational transfer of a $Glc_3Man_9$-$GlcNAc_2$ precursor from a dolichol carrier onto the asparagine residue of an Asn-X-Ser/Thr glycosylation sequon of the protein. Terminal glucose residues are rapidly cleaved by endoplasmic reticulum α-glucosidases of which two distinct activities have been identified:[140] glucosidase I, which removes the terminal α1,2-linked glucose residue, and glucosidase II hydrolyzing the remaining two α1,3-linked glucose residues. Further trimming of the oligosaccharides by ER- and Golgi-located mannosidases permits subsequent processing to complex- and hybrid-type structures through the action of Golgi resident glycosyl transferases.[34,141]

Several compounds that inhibit purified glucosidases have been identified.[142,143] In the presence of these inhibitors complex-type oligosaccharide synthesis would be expected to be blocked. However, it has been shown in a number of systems that when cells are treated with these compounds some complex-type oligosaccharide formation still occurs. There are several possible reasons for this. These include not achieving a high enough inhibitor concentration within the endoplasmic reticulum and the presence of endomannosidase activity which would provide a bypass mechanism to circumvent glucosidase inhibition (see ref 144 and reference therein). It has also been found that the effects of glucosidase inhibition on cellular glycoproteins are selective. Some glycoproteins require correct oligosaccharide processing for secretion or cell surface expression, while for others complete processing of their oligosaccharides is less critical.

When a range of sugar analogues were screened for anti-HIV activity *in vitro*,[145,146] *N*-butyldeoxygalactonojirimycin (NB-DNJ) was found to be a potent inhibitor of infection and exhibited minimal cytotoxicity. This compound inhibited purified α-glucosidase I with a $K_i$ of 0.22 $\mu$m.

The envelope glycoproteins of the HIV virus are heavily N-glycosylated. HIV-1 gp120 has 20−25 potential sites for N-linked glycosylation with the carbohydrate contributing 50% of its apparent mo-

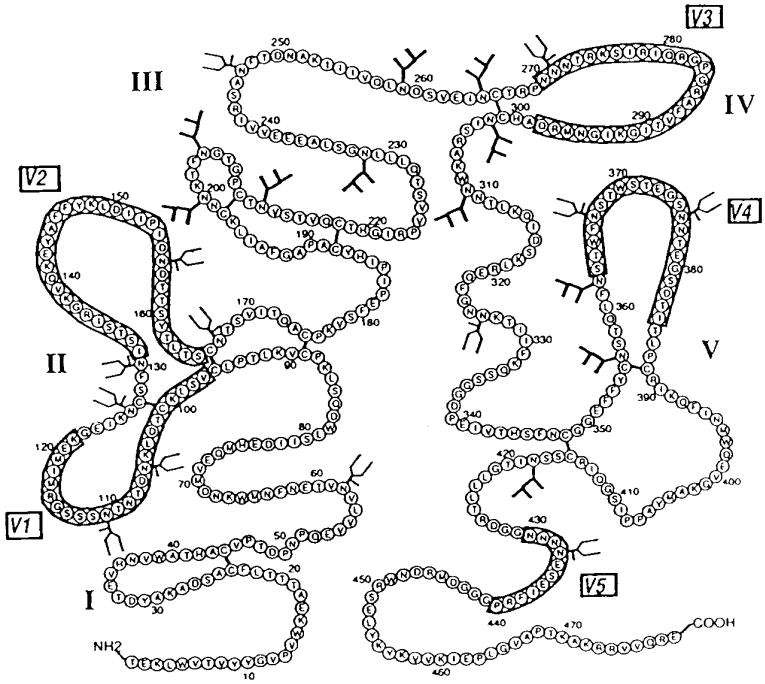



**Figure 50.** Schematic diagram of the HIV-1 envelope glycoprotein gp120 (adapted from ref 149). Glycosylation sites containing high mannose and/or hybrid oligosaccharides and those containing complex-type oligosaccharides are indicated. The disulfide-bonded domains are labeled with roman numbers and the hypervariable regions are enclosed in boxes and labeled V1−V5.

lecular weight.[147] The positions of the glycosylation sites within the primary amino acid sequence of gp120 are relatively consistent between different isolates of HIV-1[148−150] (Figure 50). At least 13 of the glycosylation sites are conserved, and the remaining sites usually are not located more than approximately 10 residues from the sites in the reference strain, HIV-$1_{IIIB}$.[149] Studies have shown that a diverse range of high mannose, hybrid, and bi-, tri-, and tetraantennary structures are present both on recombinant CHO expressed and virally derived gp120 and that the proportions of the different structures are similar between the two systems.[151−154] Analysis of gp120 mutants suggests that N-glycosylation of either gp120 or gp41 is necessary for a post-CD4 binding event, such as the fusion of the viral and cellular membranes.[155,156]

It has been shown that NB-DNJ inhibits glycoprotein processing in intact cells at concentrations equivalent to those which inhibit HIV replication *in vitro*.[157] Using metabolic labeling with ($^3$H) mannose, endo H release of labeled oligosaccharides, and gel filtration chromatography, we have demonstrated that treatment with antiviral concentrations of NB-DNJ results in the terminal sequence Glcα1,2Glcα,1,-3Glcα1,3Man being present on gp120 oligosaccharides (Figure 51).

This has also been demonstrated for gp120 derived from H9 cells, acutely infected with the HIV-1 IIIb strain.[158] Two consequences of treatment with NB-DNJ are the inhibition of synctia formation in cells infected with HIV-1 and the reduction in release of infectious virus. This reduction is not caused by a decrease in the release of virus particles but from a reduced infectivity of the virus released (Figure 52).

The alterations in infectivity of the virus particles released in the presence of NB-DNJ has been analyzed by using Cocal (HIV) pseudotype viruses in an "entry" assay. The production and assay of these viruses is illustrated schematically in Figure 53. The assay is based on the entry of Cocal (HIV) pseudotypes carrying an HIV envelope and a Cocal virus core into Hela T4 cells. After entry the Cocal core is released into the cells leading to plaque formation from cytolitic replication of Cocal virus. Plaques formed by wild type Cocal virions can be neutralized by anti-Cocal antiserum. Wild-type HIV and HIV (Cocal) pseudotypes do not form plaques in these walls.

The assay thus enables the study of the result of altering the oligosaccharides of the viral envelope by using NB-DNJ. It was found that entry of Cocal (HIV) pseudotype viruses produced in the presence of NB-DNJ was greatly impaired. This demonstrates that altering oligosaccharides in HIV envelope leads to reduced viral infectivity. Further studies detailed in ref 159 suggest that the altered oligosaccharides affect viral entry post binding to CD4 but prior to fusion.

In contrast to the HIV envelope glycoproteins, which contains about 30 glycosylation sites, the hepatitis B virus envelope proteins contain only one or two glycosylation sites. *In vitro* treatment of this virus with NB-DNJ results in a high proportion of virus particles being retained inside the cell. Preliminary data show that these viruses contain a large proportion of endo-H-sensitive oligosaccharides.[160] This suggests that correct glycosylation is necessary for the processes involving transport of the hepatitis B virus out of the cell. Comparison of the effect of NB-DNJ on these two viruses emphasize that oligosaccharides attached to proteins can have very different functions.

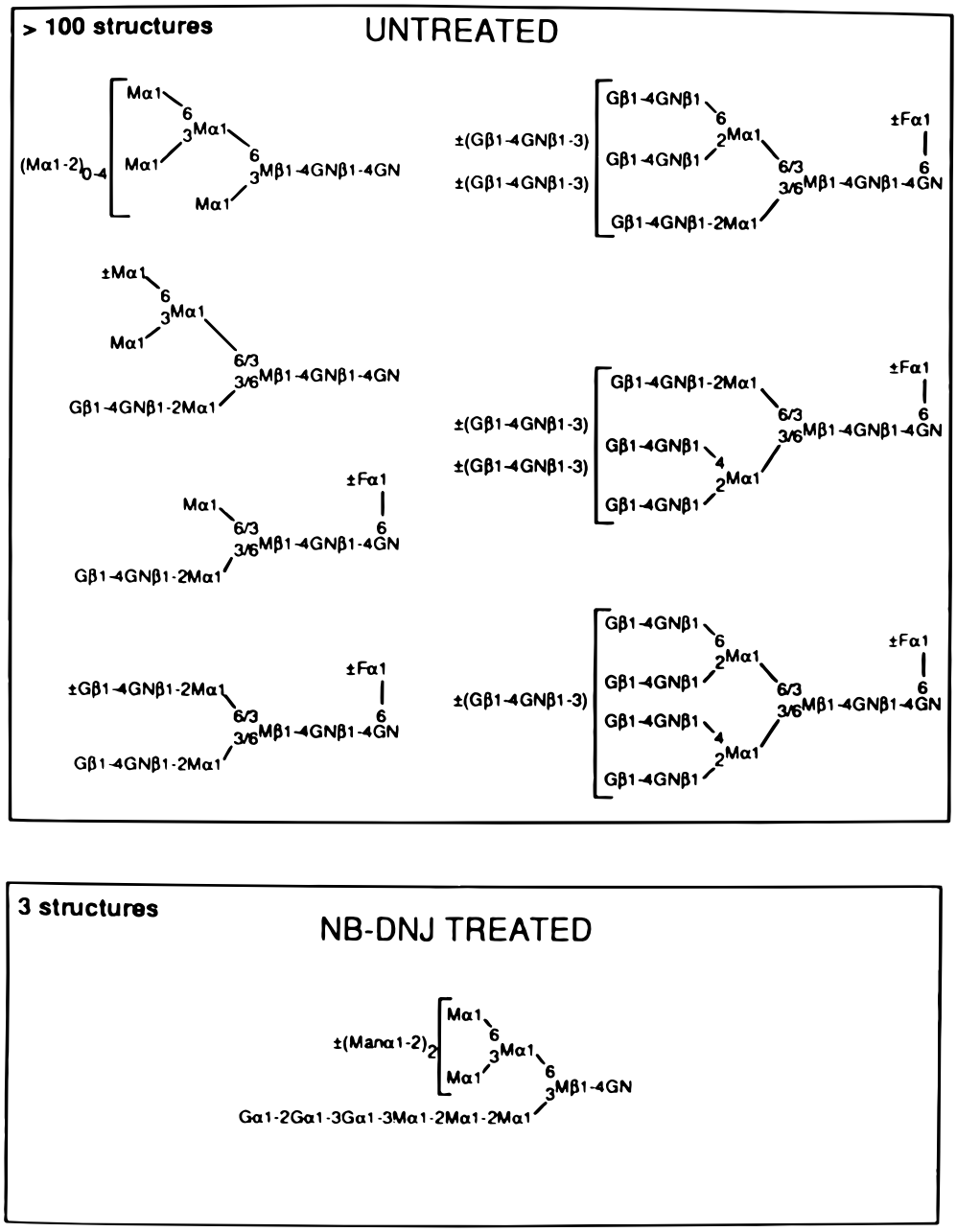
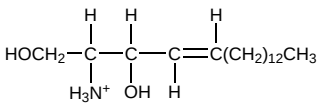



**Figure 51.** Summary of N-glycosylation structures found on untreated and NB-DNJ treated recombinant gp120 expressed in CHO cells. The upper panel shows the structures from untreated gp120[151,155] and the lower panel shows the structures present following NB-DNJ treatment.[158]

## 10.1. Glycosphingolipids

Glycosphingolipids, as their name implies, are sugar-containing lipids and are a major class of glycoconjugates synthesized by eukaryotic cells. They consist of three characteristic building block components: one of sphingosine (or one of its derivatives) and one molecule of a fatty acid with carbohydrate residues of varying complexity as their polar head groups. Sphingosine itself is a non-glycosylated long-chain amino alcohol.[181]




The amino group of the sphingoid base is attached to a long-chain fatty acyl group (that can be saturated or monounsaturated depending on different factors, such as, cell type, diet, or age) to form ceramide. While the ceramide moiety is located in the membrane lipid bilayer, the saccharide head group extends out from the plasma membrane surface to the extracellular space. The sugar residue in glycosidic linkage (always $\beta$) to the ceramide is either glucose or galactose, with glucose the more prevalent. The most complex sphingolipids are the glycosphingolipids, which include neutral lipids containing from one (cerebrosides) to 20 or more glucose units and acidic glycosphingolipids containing one or more sialic acid residues (gangliosides) or sulfate esters (sulfatides).

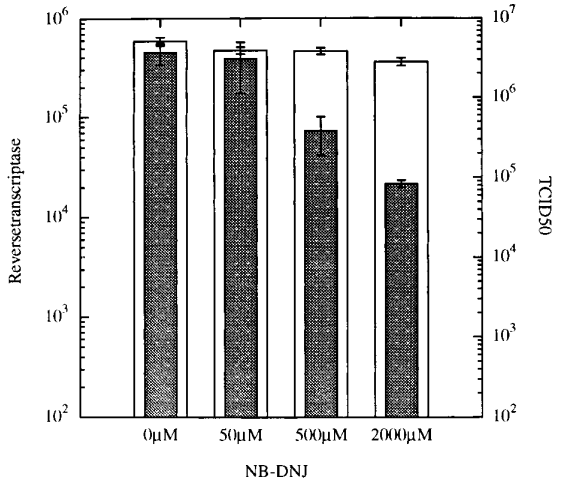



**Figure 52.** Effect of NB-DNJ on virus output and infectivity in H9 cells: (□) reverse transcriptase activity released; (■) infectivity of released virus.

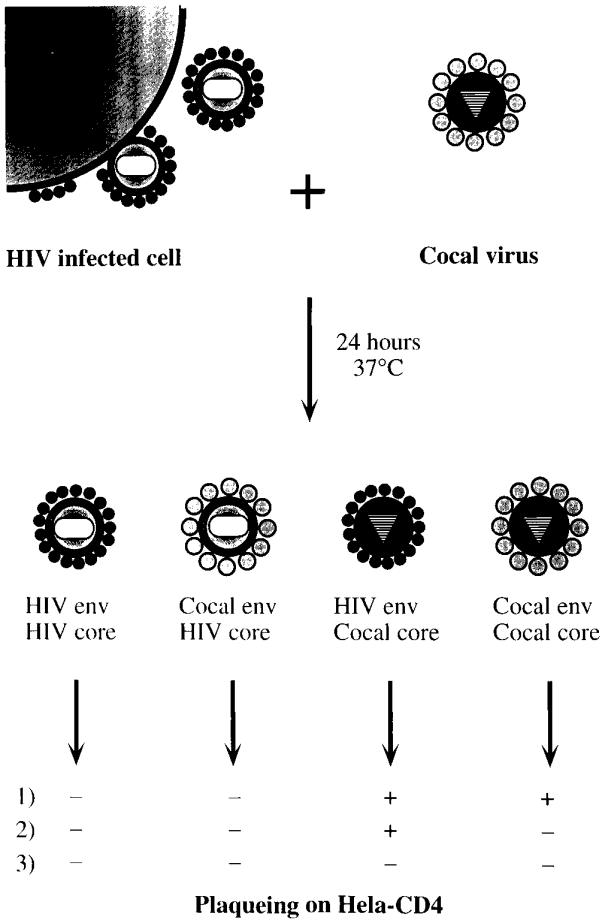



**Figure 53.** Production and assay of HIV-cocal pseudotypes: (1) no neutralization; (2) anti-Cocal; (3) anti-Cocal + anti-gp120.

Glycosphingolipids are classified according to their carbohydrate components. These complex lipids can be divided into two major groups: neutral and acidic. Cerebrosides are the simplest glycosphingolipids and consist of sphingosine, a fatty acid residue, and a single monosaccharide polar head group (Figure 54).[181] More complex neutral glycosphingolipids have unbranched or branched oligosaccharide head groups

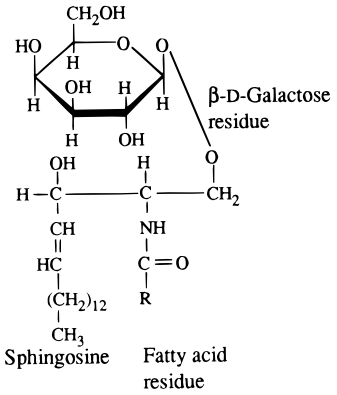



**Figure 54.** A galactocerebroside. (Reprinted from ref 181. Copyright 1992 International Thompson Publishing Services.)

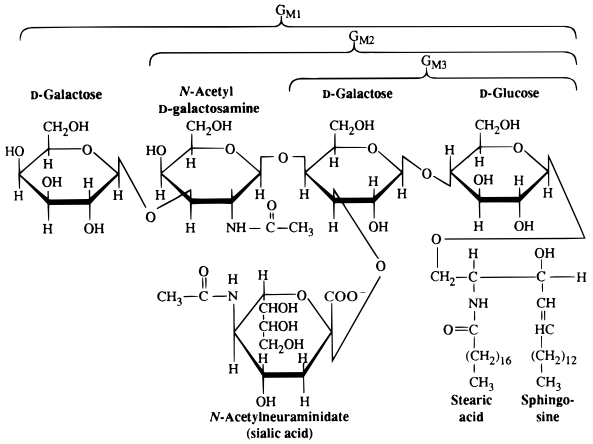



**Figure 55.** Gangliosides $G_{M1}$, $G_{M2}$, and $G_{M3}$. (Reprinted from ref 181. Copyright 1992 International Thompson Publishing Services.)

containing up to 20 sugar residues. Acidic glycosphingolipids are sulfatides or gangliosides. Gangliosides have a similar basic structure to the neutral glycosphingolipids but are differentiated from each other by the fatty acid residue they contain and by the sialylated oligosaccharides which comprise the polar head group. Some examples are shown in Figure 55.[181]

Over 100 gangliosides have been identified. Gangliosides are found in all vertebrate cells but are highly enriched in nervous tissues. They have considerable physiological and medical significance. The expression of glycosphingolipids on the cell surface changes as the cells divide and differentiate. There is considerable evidence that glycosphingolipids are specific determinants of cell–cell recognition, so that they probably have an important role in growth and differentiation of tissues. Specific glycosphingolipids of host cells have been shown to interact with the proteins of viral and bacterial parasites. For instance, ganglioside $G_{M1}$ acts as the cell surface receptor for the bacterial toxin that causes the debilitating diarrhea of cholera.

The glycans on glycosphingolipids are in very close proximity to the plasma membrane, and it is likely that the spatial proximity of lipid-linked glycans to the cell surface is biologically important in mediating interactions which require intimate cell–cell contact.

The common step in the biosynthesis of the vast majority of glycosphingolipids involves the transfer of glucose to ceramide. Partially inhibiting this step will result in a reduction of the glycosphingolipids on the cell surface. This approach should allow some of their functional roles to be investigated. Additionally, such inhibitors of glycolipid biosynthesis may also have therapeutic potential in treating the human glycosphingolipid storage disorders.[179,180]

### 10.2. Glycosphingolipid Storage Disorders

These result from the inheritance of defects in the genes encoding catabolic enzymes (which are then impaired) required for the sequential breakdown of the glycolipids within lysosomes. The therapeutic options for the treatment of these disorders are very limited. Currently, only the nonneuronopathic form of Gaucher's disease (type 1), a condition characterized by glucocerebrosidase deficiency, which occurs at high frequency in Ashkenazi Jews, is being successfully treated using enzyme-replacement therapy.[161,162] However, the use of specific inhibitors of glycosphingolipid biosynthesis to reduce their levels in cells, is one approach which may be generally applicable to treating storage disorders. This might then allow the impaired enzymes to catabolize fully the reduced levels of glycosphingolipids thus preventing their accumulation.

One example of this has recently been shown using NB-DNJ, which in addition to being an α-glucosidase I α II inhibitor, is also a potent inhibitor of glycosphingolipid biosynthesis.[163] This compound inhibits the initial monosaccharide attachment and so this approach may be useful in all glycosphingolipid storage disorders. Interestingly, the galactose analogue *N*-butyldeoxygalactonojirimycin (NB-DGJ) was found to inhibit glycosphingolipid biosynthesis.[164] This is a more selective compound in that unlike NB-DNJ it does not inhibit the α-galactosidases.

Clearly the ability to manipulate glycoprotein and glycolipid biosynthetic pathways using specific glycosyltransferase and glycosidase inhibitor has enormous potential. In this way the roles for glycolipids in such areas as neuronal and embryonic development can also be investigated, as well as their roles in adhesion processes and tumor metasis.

## *11. Concluding Remarks*

The chemistry of simple sugars was worked out in the late nineteenth century by Emil Fischer, and the ring structures determined in the interwar years by Haworth and colleagues. Simple polysaccharides such as starch, glycogen, and cellulose, as well as more complex molecules such as chitin and hyaluronic acid had also received attention and their component sugars identified by classical means. By the 1960s, especially through work on blood-group determinants by Morgan, Watkins, and their associates at the Lister Institute, it had become clear that besides simple mono- and polysaccharides, naturally occurring carbohydrates were commonly conjugated to proteins and lipids (as glycoproteins and glycolipids).

Little progress could be made to determine the structure or function of these complex molecules until sensitive and sophisticated techniques became available to analyze the component sugars and the order and structural details of their attachment to protein. Today automatic techniques are available for analysis of glycoproteins (in picomole amounts) and the progress in technology has advanced considerably our understanding of carbohydrate structures attached to proteins.

Protein glycosylation is influenced by three main factors: the overall protein conformation, the effect of local conformation, and the available repertoire of glycosylation-processing enzymes for the particular cell type. In general, the pattern of glycoforms is protein specific, site specific, and tissue or cell specific.

Glycobiology deals with the nature and role of carbohydrates in biological events. Glycoproteins are now known to be fundamental to many important biological processes including fertilization, immune defense, viral replication, parasitic infection, cell growth, cell–cell adhesion, degradation of blood clots, and inflammation. They are major components of the outer surface of mammalian cells. Over half the biologically important proteins are glycosylated. Oligosaccharide structures change dramatically during development and it has been shown that *specific sets* (i.e. specific sequences) of oligosaccharides are expressed at distinct stages of differentiation. Further, alterations in cell surface oligosaccharides are associated with various pathological conditions including malignant transformation.

The finding that glycosylation may vary with disease also leads to the concept that its manipulation might alter the properties of glycoproteins and result in beneficial therapeutic results. The ability to manipulate and modify sugar structures also provides an important approach in understanding the different functions of oligosaccharides.

The elegant biosynthetic glycan-processing pathway in the cell allows, in principle, the same oligosaccharide to be attached to quite different proteins without having to code the information into the DNA of the individual proteins. However, the orientation of the attached oligosaccharide with respect to the polypeptide may markedly affect the properties of the glycoproteins. Further, different glycoforms of a protein may display quite different orientations of the oligosaccharides with respect to the protein, thus conferring different properties. A striking example is the structure of the Fc fragment of IgG. The conserved N-linked complex oligosaccharides at Asn-297 on each heavy chain of the $C_H2$ domain occupy the interstitial space between the domains and also interact with the domain surface. Loss of the two terminal galactoses from the oligosaccharide as in the Fc fragment from patients with rheumatoid arthritis, results in a loss of interaction between the domain surface and the oligosaccharide. This permits displacement and consequent exposure of the oligosaccharides, giving them the potential to be recognized by endogenous receptors lectins.

The recognition of oligosaccharides (lectins) is influenced by their accessibility, the number of copies of the oligosaccharides, and their precise geometry of presentation. These factors introduce a high degree of specificity and control as to whether the

recognition is physiologically relevant or not.

That one set of structures on different proteins can result in quite dramatic variations in properties of glycoproteins or that different glycoproteins may have different properties emphasizes that there is no single unifying function for oligosaccharides.

## 12. Acknowledgments

It is a privilege and honor to acknowledge my debt to the late Rodney Porter whose support led me into the field of immunology and then glycobiology. He encouraged me to form a relationship with Monsanto Company to develop and automate the technology of microsequencing of oligosaccharides. The Monsanto/Oxford University relationship was the first major industrial partnership for the University. My colleagues at Monsanto were enlightened and concerned with developing the science. Their approach also led, with the University of Oxford, to the foundation of Oxford GlycoSystems, the first University spin-off company which was to develop technology arising mainly from within the Glycobiology Institute.

I have benefited enormously from the help, advice, and skills of my colleagues and students in developing the Glycobiology Institute and equipping it so as to be a resource for glycobiology. It is particularly appropriate to acknowledge fundamental contributions made to the field by Professors T. Rademacher, M. Ferguson, and S. W. Homans. Dr. Parekh, who is now Chief Scientist at Oxford GlycoSystems, made an enormous contribution in the early years.

More recently, the work of Drs. F. Platt, T. Butters, G. B. Karlsson, and P. Fischer helped to establish the antiviral program on a firm footing and also laid the foundations for the study of glycolipid storage disorders using amino sugars. Dr. David Wing has been a resource for the whole Institute. Dr. Pauline Rudd has also been a constant source of support, working with me to develop many different projects and has been of significant help with this manuscript.

I would also like to acknowledge contributions made by Drs. D. Ashford, C. Edge, D. Fernandes, G. Guile, D. Harvey, B. Matthews, N. Karlsson, I. Scragg, J. Thomas, P. Williams, R. Woods, S. Wong, M. Wormald, and S. Zamze much of whose data I have used.

My thanks are also due to Drs. Maureen Taylor and Kurt Drickamer for their help in understanding C-type lectins.

We thank the Biochemical Society for permission to reproduce figures from the 7th Wellcome Trust Award for Research in Biochemistry Related to Medicine, 1994 and published in the *Biochemical Society Transaction*, Volume 23, Wellcome Trust Award Lecture.

## 13. References

(1) (1) Metzler, M.; Gertz, A.; Sarkar, M.; Schachter, H.; Schrader, J. W.; Marth, J. D. *EMBO* **1994**, *13*, 2056−2065.
(2) Ioffe, E.; Stanley, P. *Proc. Natl. Acad. Sci.* **1994**, *91*, 728−732.
(3) Weis, W. I.; Drickamer, K. *Curr. Biol.* **1994**, *2*, 1227−1240.
(4) Sherrif, S.; Chang, C. Y.; Ezekowitz, R. A. B. *Nat. Struct. Biol.* **1994**, *11*, 789−794.
(5) Drickamer, K.; Taylor, M. E. *Annu. Rev. Cell Biol.* **1993**, 237−264.
(6) Rademacher, T. W.; Parekh, R. B.; Dwek, R. A. *Annu. Rev. Biochem.* **1988**, *57*, 785−838.
(7) Varki, A. *Glycobiology* **1993**, *3*, 97−130.
(8) Parekh, R. B.; Dwek, R. A.; Sutton, B. J.; Fernandes, D. L.; Leung, A.; Stanworth, D.; Rademacher, T. W.; Mizuochi, T.; Taniguchi, K.; Matsuta, K.; Takeuchi, Y.; Nagano, T.; Miyamoto, T.; Kobata, A. *Nature* **1985**, *316*, 452−457.
(9) Parekh, R. B.; Dwek, R. A.; Thomas, J. R.; Rademacher, T. W.; Opdenakker, G.; Wittwer, A. J.; Howard, S. C.; Nelson, R.; Siegel, N. R.; Jennings, M. G.; Harakas, N. K.; Feder, J. *Biochemistry* **1989**, *28*, 7644−7662.
(10) Williams, R. L.; Greene, S. M.; McPhearson, A. *J. Biol. Chem.* **1987**, *262*, 16020−16031.
(11) Robertson, A. D.; Purisima, E. O.; Eastman, M. A.; Scheraga, H. A. *Biochemistry* **1989**, *28*, 5930−5938.
(12) Rico, M.; et al. *J. Biomol. NMR* **1991**, *1*, 283−298.
(13) Rico, M.; et al. *Annu. Eur. J. Biochem.* **1989**, *183*, 623−638.
(14) Santoro, J.; et al. *J. Mol. Biol.* **1993**, *229*, 722−734.
(15) Joao, H. C.; Scragg, I. G.; Dwek, R. A. *FEBS Lett.* **1992**, *307*, 343−346.
(16) Wooten, E. W.; Edge, C. J.; Bazzo, R.; Dwek, R. A.; Rademacher, T. W. *Carbohyd. Res.* **1990**, *203*, 13−17.
(17) Rao, U.; Teeter, M. M. *Protein Eng.* **1993**, *6*, 837−847.
(18) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230−252.
(19) Wooten, E. W.; Bazzo, R.; Edge, C. J.; Zamze, S.; Dwek, R. A.; Rademacher, T. W. *Eur. Biophys. J.* **1990**, *18*, 139−148.
(20) Homans, S. W.; Pastore, A.; Dwek, R. A.; Rademacher, T. W. *Biochemistry* **1987**, *26*, 6649−6655.
(21) Homans, S. W.; Dwek, R. A.; Boyd, J.; Mahmoudian, M.; Richards, W. G.; Rademacher, T. W. *Biochemistry* **1986**, *25*, 6342−6350.
(22) Brisson, J.-R.; Carver, J. P. *Biochemistry* **1983**, *22*, 3680−3686.
(23) Brisson, J.-R.; Carver, J. P. *Biochemistry* **1983**, *22*, 3671−3680.
(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Phys. Chem.* **1983**, *79*, 926−935.
(25) Shaanan, B.; Lis, H.; Sharon, N. *Science* **1991**, *254*, 862−866.
(26) Bode, W.; Meyer, E.; Powers, J. C. *Biochemistry* **1989**, *28*, 1951−1963.
(27) Marquart, M.; Deisenhofer, J.; Huber, R.; Palm, W. *J. Mol. Biol.* **1980**, 369−391.
(28) Freymann, D.; Down, J.; Carrington, M.; Roditit, I.; Turner, M.; Wilery, D. *J. Mol. Biol.* **1990**, *216*, 141−160.
(29) Woods, R. J.; Edge, C. J.; Wormald, M. R.; Dwek, R. A. In *Complex Carbohydrates in Drug Research. Alfred Benzon Symposium 36.*; Clausen, K. B. a. H., Ed.; Munksgaard: Copenhagen, 1994.
(30) Woods, R. J.; Fraser-Reid, B.; Dwek, R. A.; Edge, C. J. In *Modelling the Hydrogen Bond*, 1993; Vol. 569.
(31) Bush, C. A.; Blumberg, K.; Brown, J. N. *Biopolymers* **1982**, *21*, 1971−1977.
(32) Delbaere, L. T. J. *Biochem. J.* **1974**, *143*, 197−205.
(33) Wormald, M. R.; Wooten, E. W.; Bazzo, R.; Edge, C. J.; Feinstein, A.; Rademacher, T. W.; Dwek, R. A. *Eur. J. Biochem.* **1991**, *198*, 131−139.
(34) Kornfeld, R.; Kornfeld, S. *Annu. Rev. Biochem.* **1985**, *54*, 631−634.
(35) Harris, R. J. *Biochemistry* **1993**, *32*, 6539−6547.
(36) Harris, R. J.; Leonard, C. K.; Guzetta, A. W.; Spellman, M. W. *Biochemistry* **1991**, *30*, 2311−2314.
(37) Nishimura, H.; Takao, T.; Hase, S.; Shimonishi, Y.; Iwanaga, S. *J. Biol. Chem.* **1992**, *276*, 17520−17525.
(38) Haltiwanger, R. S.; Kelly, W. G.; Roquemore, E. P.; Blomberg, M. A.; Dong, L. D.; Kreppel, L.; Chou, T.; Hart, G. W. *Biochem. Soc. Trans.* **1992**, *20*, 264−269.
(39) Hart, G. W. *Curr. Opin. Cell Biol.* **1992**, *4*, 1017−1023.
(40) Ferguson, M. A. J.; Williams, A. F. *Annu. Rev. Biochem.* **1988**, *1*, 522−529.
(41) Ferguson, M. A. J. *Curr. Opin. Struct. Biol.* **1991**, *1*, 522−529.
(42) Yamashita, K.; Kamerling, J. P.; A., K. *J. Biol. Chem.* **1982**, *257*, 12809−12814.
(43) Fukuda, M. *Mol. Glycobiol.* **1994**, 1−52.
(44) Schachter, H.; Brockhausen, I. *Glycoconjugates: Composition Structure, and Function*; Dekker: New York, 1992; pp 263−332.
(45) Goochee, C. F.; Monica, T. *Bio/Technology* **1990**, *8*, 421−427.
(46) Schachter, H. *Mol. Glycobiol.* **1994**, 88−162.
(47) Schachter, H. *Biochem. Cell Biol.* **1986**, *64*, 163−181.
(48) Mononen, I.; Karjalainen, E. *Biochem. Biophys. Acta* **1987**, *788*, 364−367.
(49) Gavel, Y.; Von Heijne, G. *Protein Eng.* **1990**, *3*, 433−442.
(50) Williams, A. F.; Parekh, R. B.; Wing, D. R.; Willis, A. C.; Barclay, A. N.; Dalchau, R.; Fabre, J. W.; Dwek, R. A.; Rademacher, T. W. *Glycobiology* **1993**, *3*, 339−348.
(51) Parekh, R. B.; Tse, A. G. D.; Dwek, R. A.; Williams, A. F.; Rademacher, T. W. *Eur. Mol. Biol. Organ. J.* **1987**, *6*, 1233−1244.
(52) Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. *Biochemistry* **1994**, *33*, 17−22.
(53) Harvey, D. J.; Rudd, P. M.; Bateman, R. H.; Bordoli, R. S.; Howes, K.; Hoyes, J. B.; Vickers, R. G. *Org. Mass Spectrom.* **1994**, *29*, 753−766.
(54) Dwek, R. A.; Edge, C. J.; Harvey, D. J.; Wormald, M. R.; Parekh, R. B. *Annu. Rev. Biochem.* **1993**, *62*, 65−100.

(55) Davis, S. J.; Puklavec, M. J.; Ashfor, D. A.; Harlos, K.; Jones, E. Y.; Stuart, D. I.; Williams, A. F. *Protein Eng.* **1993**, *6*, 229−232.
(56) Ashford, D. A.; Alafi, C. D.; Gamble, V. M.; MacKay, D. J. G.; Rademacher, T. W.; Williams, P. J.; Dwek, R. A.; Barclay, A. N.; Davis, S. J.; Somoza, C.; Ward, H. A.; Williams, A. F. *J. Biol. Chem.* **1993**, *268*, 3260−3267.
(57) Davis, S. J.; Ward, H. A.; Puklavec, M. J.; Willis, A. C.; Williams, A. F.; Barclay, A. N. *J. Biol. Chem.* **1990**, *265*, 10410−10418.
(58) Dwek, R. A.; Ashford, D. A.; Edge, C. J.; Parekh, R. B.; Rademacher, T. W.; Wing, D. R.; Barclay, A. N.; Davis, S. J.; Williams, A. F. *Phil. Trans. R. Soc. London Biochem.* **1993**, *342*, 43−50.
(59) Carr, S. A.; Hemling, M. E.; Folena-Wasserman, G.; Sweet, R. W.; Anumula, K.; Barr, J. R.; Huddleston, M. J.; Taylor, P. *J. Biol. Chem.* **1989**, *264*, 21286−21295.
(60) Spellman, M. W.; Leonard, C. K.; Baa, L. J.; Gelineo, I.; Van Halbeek, H. *Biochemistry* **1991**, *30*, 2395−2406.
(61) Harris, R. J.; Chamow, S. M.; Gregory, T. J.; Spellman, M. W. *Eur. J. Biochem.* **1990**, *188*, 291−300.
(62) Jones, E. Y.; Davis, S. J.; Williams, A. F.; Harlos, K.; Stuart, D. I. *Nature* **1992**, *360*, 232−239.
(63) Spiro, R. G.; Bhoyroo, V. D. *J. Biol. Chem.* **1984**, *259*, 9858−9866.
(64) Galili, U.; Shohet, S. B.; Kobrin, E.; Stults, C. L. M.; Macher, B. A. *J. Biol. Chem.* **1988**, *263*, 17755−17762.
(65) Williams, A. F.; Gagnon, J. *Science* **1982**, *216*, 696−703.
(66) Feder, J.; Tolbert, W. R.; Rademacher, T. W.; Parekh, R. B.; Dwek, R. A. Tissue plasminogen activator from normal human colon cells, US Patent No. 4,751,084, 1988.
(67) Wittwer, A. J.; Howard, S. C.; Carr, L. S.; Harakas, N. K.; Feder, J.; Parekh, R. B.; Rudd, P. M.; Dwek, R. A.; Rademacher, T. W. *Biochemistry* **1989**, *28*, 7662−7669.
(68) Patthy, L. *Cell* **1985**, *41*, 657−663.
(69) Ny, T.; Elgh, F.; Lund, B. *Proc. Natl. Acad. Sci.* **1984**, *81*, 5355−5359.
(70) Zamze, S.; Jacques, A.; Rademacher, T. W.; Opdenakker, G.; Dwek, R. A. Manuscript in preparation, 1995.
(71) Joao, H. C.; Dwek, R. A. *Eur. J. Biochem.* **1993**, *218*, 239−244.
(72) Shall, S.; Barnard, E. A. *J. Mol. Biol.* **1969**, *41*, 237−251.
(73) McPherson, A.; Brayer, G.; Cascio, D.; Williams, R. *Science* **1986**, *232*, 765−8.
(74) Rijken, D. C.; Collen, D. *J. Biol. Chem.* **1981**, *256*, 7035−7041.
(75) Wittwer, A. J.; Howard, S. C. *Biochemistry* **1990**, *29*, 4175−4180.
(76) Wallen, P.; Pohl, G.; Bergsdorf, N.; Ranby, M.; Ny, T.; Jornvall, H. *Eur. J. Biochem.* **1983**, *132*, 681−686.
(77) Ranby, M.; Bergsdorf, N.; Pohl, G.; Wallen, P. *FEBS Lett.* **1982**, *146*, 289−292.
(78) Spellman, M. W.; Basa, L. J.; Leonard, C. K.; Chakel, J. A.; O'Conner, J. V.; Wilson, S.; van Halbeek, H. *J. Biochem.* **1989**, *264*, 14100−14111.
(79) Brockway, W.; Castellino, F. J. *Arch. Biochem. Biophys.* **1972**, *151*, 194−199.
(80) Sottrup-Jensen, L.; Claeys, H.; Zajdel, M.; Petersen, T. E.; Magnussen, S. Raven Press: New York, 1978; pp 191−209.
(81) Hayes, M. L.; Castellino, F. J. *J. Biol. Chem.* **1979**, *254*, 8768−8780.
(82) Gonzales-Gronow, M.; Grennet, H. E.; Fuller, G. M.; Pizzo, S. V. *Biochim. Biophys. Acta* **1990**, *1039*, 269−276.
(83) Pirie-Shepherd, S.; Jett, E. A.; Andon, N. L.; Pizzo, S. V. *J. Biol. Chem.* **1995**, *270*, 5877−5881.
(84) Ponting, C. P.; Marshall, J. M.; Cederholm-Williams, S. A. *Blood Coagulation Fibrinolysis* **1992**, *3*, 605−614.
(85) Gonzalez-Gronow, M.; Edelberg, J. M.; Pizzo, S. V. *Biochemistry* **1989**, *28*, 2374−2377.
(86) Edelberg, J. M.; Enghild, J. J.; Pizzo, S. V.; Gonzalez-Gronow, M. *J. Clin. Invest.* **1990**, *86*, 107−112.
(87) Hall, S. W.; VandenBerg, S. R.; Gonias, S. L. *J. Cell Biochem.* **1990**, *43*, 213−227.
(88) Mori, K.; Dwek, R. A.; Downing, A. K.; Opdenakker, G.; Rudd, P. M. *J. Biol. Chem.* **1995**, *270*, 3261−3267.
(89) Jaeken, J.; van Eijk, H. G.; van der Heul, C.; Corbeel, L.; Eeckels, R.; Eggermont, E. *Clin. Chim. Acta* **1984**, *144*, 245−247.
(90) Yamashita, K.; Ideo, H.; Ohkura, T.; Fukushima, K.; Yusa, I.; Ohno, K.; Takeshita, K. *J. Biol. Chem.* **1993**, *268*, 5783−5789.
(91) Iurin, O.; Mattu, T.; Kier, G.; Mian, N.; Rudd, P. M.; Winchester, B.; Dwek, R. A. Submitted for publication, 1995.
(92) Ferguson, M. A. J.; Homans, S. W.; Dwek, R. A.; Rademacher, T. W. *Science* **1988**, *239*, 753−759.
(93) Homans, S. W.; Ferguson, M. A. J.; Dwek, R. A.; Rademacher, T. W.; Anand, R.; Williams, A. F. *Nature* **1988**, *333*, 269−272.
(94) Homans, S. W.; Edge, C. J.; Ferguson, M. A. J.; Dwek, R. A.; Rademacher, T. W. *Biochemistry* **1989**, *28*, 2881−2887.
(95) Rademacher, T. W.; Edge, C. J.; Dwek, R. A. *Curr. Biol.* **1991**, *1*, 41−42.
(96) Amzel, L. M.; Poljak, R. A. *Annu. Rev. Biochem.* **1979**, *48*, 961−997.
(97) Deisenhofer, J. *Biochemistry* **1981**, *20*, 2361−2370.
(98) Sutton, B. J.; Phillips, D. C. *Biochem. Soc. Trans.* **1983**, *11*, 130−132.
(99) Leatherbarrow, R. J.; Rademacher, T. W.; Dwek, R. A.; Woof, J. M.; Clar, A.; Burton, D. R.; Richardson, N.; Feinstein, A. *Mol. Immunol.* **1985**, *22*, 407−415.
(100) Leatherbarrow, R. J.; Dwek, R. A. *FEBS Lett.* **1983**, *164*, 227−230.
(101) Lund, J.; Tanaka, T.; Takahashi, N.; Sarmay, G.; Arata, Y.; Jefferis, R. *Mol. Immunol.* **1990**, *27*, 1145.
(102) Drickamer, K. *J. Biol. Chem.* **1988**, *263*, 9557−9560.
(103) Taylor, M. E. *Blood Cell Biochem.* **1993**, *5*, 347−370.
(104) Drickamer, K. *Curr. Opin. Struct. Biol.* **1993**, *3*, 393−400.
(105) Drickamer, K. *Curr. Opin. Struct. Biol.* **1995**, in press.
(106) Weis, W.; Kahn, R.; Fourme, R.; Drickamer, K.; Hendrickson, W. A. *Science* **1991**, *254*, 1608−1615.
(107) Ikeda, K.; Sannoh, T.; Kawasaki, N.; Kawasaki, T.; Yamashina, I. *J. Biol. Chem.* **1987**, *262*, 451−454.
(108) Weis, W. I.; Drickamer, K.; Hendrickson, W. A. *Nature* **1992**, *360*, 127−134.
(109) Woods, R. J.; Edge, C. J.; Dwek, R. A. *Nature Struct. Biol.* **1994**, *1*, 499−501.
(110) Mullin, N. P.; Hall, K. T.; Taylor, M. E. *J. Biol. Chem.* **1994**, *269*, 28405−28413.
(111) Taylor, M. E. *Biochem. Soc. Trans.* **1993**, *21*, 467−474.
(112) Haurum, J. S.; Arsequell, G.; Lellouch, A. C.; Wong, S. Y. C.; Dwek, R. A.; McMichael, A. J.; Elliot, T. *J. Exp. Med.* **1994**, *180*, 739−744.
(113) Michaelsson, E.; Malmstrom, V.; Reis, S.; Engstrom, A.; Burkhardt, H.; Holmdahl, R. *J. Exp. Med.* **1994**, *180*, 745−749.
(114) Rudd, P. M.; Fortune, F.; Patel, R. B.; Dwek, R. A.; Lehner, T. *Immunology* **1994**, *83*, 99−106.
(115) Rademacher, T. W.; Homans, S. W.; Parekh, R. B.; Dwek, R. A. *Biochemical Soc. Symp.* **1986**, *51*, 131−148.
(116) Bodman, K. B.; Sumar, N.; Mackenzie, L. E.; Isenberg, D. A.; Hay, F. C.; Roitt, I. M.; Lydyard, P. M. *Clin. Exp. Immunol.* **1992**, *88*, 420−42335.
(117) Axford, J. S.; Mackenzie, L.; Lydyard, P. M.; Hay, F. C.; Isenberg, D. A.; Roitt, I. M. *Lancet* **1987**, *ii*, 1486−1488.
(118) Furukawa, K.; Matsuta, K.; Takeuchi, F.; Kosuge, E.; Miyamoto, T.; Kobata, A. *Int. Immunol.* **1990**, *2*, 105−112.
(119) Youings, A.; Chang, S.-C.; Dwek, R. A.; Scragg, I. G. *J. Biol. Chem.* Manuscript in preparation, 1993.
(120) Field, M. C.; Amatayakul-Chantler, S.; Rademacher, T. W.; Rudd, P. M.; Dwek, R. A. *Biochem. J.* **1994**, *299*, 261−275.
(121) Jefferis, R.; Lund, J.; Mitzutani, H.; Nakagawa, H.; Kawazoe, Y.; 'Arata, Y.; Takahashi, N. *Biochem. J.* **1990**, *268*, 529−537.
(122) Burastero, S. W.; Casali, P.; Wilder, R. L.; Notkins, A. L. *J. Exp. Med.* **1988**, *168*, 1979−1992.
(123) Hardy, R. R.; Hayakawa, K.; Shimizu, M.; Yamasaki, K.; Kishimoto, T. *Science* **1987**, *236*, 81−83.
(124) Shakib, F.; Stanworth, D. R. *Ann. Rheum. Dis.* **1976**, *35*, 263−266.
(125) Kobata, A. *Glycobiology* **1990**, *1*, 5−8.
(126) Griffiths, H. R.; Dowlin, E. J.; Sahinoglu, T.; Blake, D. R.; Parnham, M.; Lunec, J. *Agents Actions* **1992**, *36*, 107−111.
(127) Parekh, R. B.; Isenberg, D. A.; Ansell, B. M.; Roitt, I. M.; Dwek, R. A.; Rademacher, T. W. *Lancet* **1988**, *1 (8592)*, 966−969.
(128) Parekh, R. B.; Isenberg, D. A.; Rook, G.; Roitt, I. M.; Dwek, R. A.; Rademacher, T. W. *J. Autoimmun.* **1989**, *2*, 101−114.
(129) Rook, G. A. W.; Steele, J.; Brealey, R.; Whyte, A.; Isenberg, D.; Sumar, N.; Nelson, J. L.; Bodman, K. B.; Young, A.; Roitt, I. M.; Williams, P.; Scragg, I.; Edge, C. J.; Arkwright, P. D.; Ashford, D.; Wormald, M.; Rudd, P. M.; Redman, C. W. G.; Dwek, R. A.; Rademacher, T. W. *J. Autoimmun.* **1991**, *4*, 779−794.
(130) Roitt, I. M.; Dwek, R. A.; Parekh, R. B.; Rademacher, T. W.; Alavi, A.; Axford, J. S.; Bodman, K. B.; Bond, A.; Cooke, A.; Hay, F. C.; Isenberg, D. A.; Lydyard, P. M.; Mackenzie, L.; Rook, G.; Smith, M.; Sumar, N. *J. Autoimmun.* **1988**, *1*, 499−506.
(131) Morgan, B. P. *Complement: Clinical aspects and relevance to disease*; Academic Press: London, 1990.
(132) Burton, D. R.; Boyd, J.; Brampton, A. D.; Easterbrook-Smith, S. B.; Emanuel, E. J.; Novotny, J.; Rademacher, T. W.; van Schravendijk, M. R.; Sternburg, M. J. E.; Dwek, R. A. *Nature* **1980**, *288*, 338−344.
(133) Duncan, A. R.; Winter, G. *Nature* **1988**, *332*, 738−740.
(134) Malhotra, R.; Lu, J.; Holmskov, U. S.; Sim, R. B. *Clin. Exp. Immunol.* **1994**, *97* (supp. 2), 4−9.
(135) Drickamer, K.; Dordal, M. S.; Reynolds, L. *J. Biol. Chem.* **1986**, *261*, 6878−6887.
(136) Iobst, S. T.; Wormald, M. R.; Weis, W. I.; Dwek, R. A.; Drickamer, K. *J. Biol. Chem.* **1994**, *269*, 15505−15511.
(137) Rademacher, T. W.; Parekh, R. B.; Dwek, R. A.; Isenberg, D.; Rook, G.; Axford, J.; Roitt, I. *Springer Semin. Immunopathol.* **1988**, *10*, 231−249.
(138) Tsuchiya, N.; Endo, T.; Matsuta, K.; Yoshinoya, S.; Takenchi, F.; Nagano, Y.; Shiota, M.; Furutawa, K.; Kochibe, N.; Ho, K.; Kobata, A. *J. Immunol.* **1993**, *151*, 1137−1146.
(139) Malhotra, R.; Wormald, M. R.; Rudd, P. M.; Fischer, P. B.; Dwek, R. A.; Sim, R. B. *Nature Med.* **1995**, *1*, 237−243.
(140) Grinna, L. S.; Robbins, P. W. *J. Biol. Chem.* **1979**, *254*, 8814−8818.
(141) Hubbard, S. C.; Ivatt, R. J. *Annu. Rev. Biochem.* **1981**, *50*, 555−583.

(142) Winchester, B.; Fleet, G. W. J. *Glycobiology* **1992**, *2*, 199−210.
(143) Datema, R.; Olofosson, S.; Romero, P. A. *Pharmacol. Ther.* **1987**, *33*, 221−286.
(144) Lubas, W. A.; Spiro, R. G. *J. Biol. Chem.* **1987**, *262*, 3775−3781.
(145) Fleet, G. W. J.; Karpas, A.; Dwek, R. A.; Fellows, L. E.; Tyms, A. S.; Petrusson, S.; Namgoong, S. K.; Ramsden, N. G.; Smith, P. W.; Son, J. C.; Wilson, F.; Witty, D. R.; Jacob, G. S.; Rademacher, T. W. *FEBS Lett.* **1988**, *237*, 128−132.
(146) Karpas, A.; Fleet, G. W. J.; Dwek, R. A.; Petrusson, S.; Namgoong, S. K.; Ramsden, N. G.; Jacob, G. S.; Rademacher, T. W. *Proc. Natl. Acad. Sci.* **1988**, *85*, 9229−9233.
(147) Lasky, L. A.; Groopman, J. E.; Fennie, C. W.; Benz, P. M.; Capon, D. J.; Dowbenko, D. J.; Nakamuira, G. R.; Nunes, W. M.; Renz, M. E. B., P. W. *Science* **1986**, *233*, 209−212.
(148) Willey, R. L.; Rutledge, R. A.; Dias, S.; Folks, T.; Theodore, T.; Buckler, C. E.; Martin, M. A. *Proc. Natl. Acad. Sci.* **1986**, *83*, 5038−5042.
(149) Leonard, C. K.; Spellman, M. W.; Riddle, L.; Harris, R. J.; Thomas, J. N.; Gregory, T. J. *J. Biol. Chem.* **1990**, *265*, 10373−10382.
(150) Alizon, M.; Wain-Hobson, S.; Montagnier, L.; Sonigo, P. *Cell* **1986**, *46*, 63−74.
(151) Mizuochi, T.; Spellman, M. W.; Larkin, M.; Solomon, J.; Basa, L. J.; Feizi, T. *Biomed. Chromatogr.* **1988**, *2*, 260−270.
(152) Mizuochi, T.; Spellman, M. W.; Larkin, M.; Solomon, J.; Basa, L. J.; Feizi, T. *Biochem. J.* **1988**, *254*, 599−603.
(153) Mizuochi, T.; Matthews, T. J.; Kato, M.; Hamako, J.; Titani, K.; Solomon, J.; Feizi, T. *J. Biol. Chem.* **1990**, *265*, 8519−8524.
(154) Geyer, H.; Holschbach, C.; Hunsmann, G.; Schneider, J. *J. Biol. Chem.* **1988**, *263*, 11760−11767.
(155) Willey, R. L.; Smith, D. H.; Lasky, L. A.; Theodore, T. S.; Earl, P. L.; Moss B.; Capon, D. J.; Martin, M. A. *J. Virol.* **1988**, *62*, 139−147.
(156) Lee, W. R.; Yu, X.-F.; Syu, W.-J.; Essex, M.; Lee, T.-H. *J. Virol.* **1992**, *66*, 1799−1803.
(157) Karlsson, G. B.; Butters, T. D.; Dwek, R. A.; Platt, F. M. *J. Biol. Chem.* **1993**, *268*, 570−576.
(158) Karlsson, G. B. University of Oxford: Oxford, 1993.
(159) Fischer, P. B.; Collin, M.; Karlsson, G. B.; James, W.; Buters, T. D.; Gordon, S.; Davis, S. J.; Dwek, R. A.; Platt, F. M. *J. Virol.* **1995**, in press.
(160) Block, T. M.; Lu, X. L.; Platt, F. M.; Foster, G. R.; Gerlich, W. H.; Blumberg, B. S.; Dwek, R. A. *Proc. Natl. Acad. Sci.* **1994**, *91*, 2235−2239.
(161) Barton, N. W.; Brady, R. O.; Dambrosia, J. M.; Di Biscelgie, A. M.; Doppelt, S. H.; Hill, S. C.; Mankin, H. J.; Murray, G. J.; Parker, R. I.; Argoff, C. E.; Grewel, R. P.; Yu, K.-T.; Collaborators *N. Engl. J. Med.* **1991**, *324*, 1464−1470.
(162) Beutler, R. *Proc. Natl. Acad. Sci.* **1993**, *90*, 5384−5390.
(163) Platt, F. M.; Neises, G. R.; Dwek, R. A.; Butters, T. D. *J. Biol. Chem.* **1994**, *269*, 8362−8365.
(164) Platt, F. M.; Neises, G. R.; Karlsson, G. B.; Dwek, R. A.; Butters, T. D. *J. Biol. Chem.* **1994**, *269*, 27108−27114.
(165) Brekke, O. H.; Michaelsen, T. E.; Sandlie, I. *Immunol. Today* **1995**, *16*, 85−90.
(166) Williams, A. F.; Davis, S. J.; He, Q.; Barclay, A. N. *Cold Spring Harbor Symp. Quant. Biol.* **1989**, *LV1*, 637−647.
(167) Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. *Biochemistry* **1994**, *33*, 17−22.
(168) de Vos, A. M.; Ultsch, M. H.; Kelley, R. F.; Padmanabahn, K.; Tulinsky, A.; Westbrook, M. L.; Kossiakoff, A. A. *Biochemistry* **1992**, *31*, 270−279.
(169) Thomas, J. R.; Dwek, R. A.; Rademacher, T. W. *Biochemistry* **1990**, *29*, 5413−5422.
(170) Weis, W. I.; Quesenberry, M. S.; Taylor, M. E.; Bezouska, K.; Hendrickson, W. A.; Drickamer, K. *Cold Spring Harbor Symp. Quant. Biol.* **1992**, *57*, 281−289.
(171) Pumphrey, R. S. H. *Molecular analysis of structure function and regulation*; Shakib, F., Ed.; Pergamon Press: Elmsford, NY, 1990.
(172) Dwek, R. A.; Sutton, B. J.; Perkins, S. J.; Rademacher, T. W. *Biochem. Soc. Symp.* **1984**, *49*, 123−136.
(173) Malhotra, R.; Thiel, S.; Reid, K. B. M.; Sim, R. B. *J. Exp. Med.* **1990**, *172*, 955−959.
(174) Rademacher, R. W.; Williams, P.; Dwek, R. A. *Proc. Natl. Acad. Sci.* **1994**, *91*, 6123−6127.
(175) Davidson, D. J.; Castellino, F. J. *Biochemistry* **1991**, *30*, 625−633.
(176) Davidson, D. J.; Castellino, F. J. *Biochemistry* **1991**, *30*, 6689−6696.
(177) Hayes, M. L.; Castellino, F. J. *J. Biochem.* **1979**, *254*, 8772−8776.
(178) Takasaki, S.; Mizuochi, T.; Kobata, A. *Methods Enzymol.* **1982**, *83*, 263−268.
(179) Sandhoff, K.; van Echten, G. *Adv. Lipid Res.* **1993**, *26*, 119−142.
(180) van Echten, G.; Sandhoff, K. *J. Biol. Chem.* **1993**, *268*, 5341−5344.
(181) Olivera, A.; Spiegel, S. *Glycoconjugate J.* **1992**, Number 3, 110−117.
(182) Yoshima, H.; Mizuochi, T.; Ishii, M.; Kobata, A. *Cancer Res.* **1980**, *40*, 4276−4281.
(183) Bigge, J. C.; Patel, T. P.; Bruce, J. A.; Goulding, P. N.; Charles, S. M.; Parekh, R. B. *Anal. Biochem.* **1995**, *230*, 229−238.
(184) Mackiewicz, A.; Dewey, M. J.; Berger, F. G.; Baumann, H. *Glycobiology* **1991**, *1*, 265−269.
(185) Ashwell, G.; Harford, J. *Annu. Rev. Biochem.* **1982**, *51*, 531−554.

CR940283B