Glycosylation and the Complement System

Gayle E. Ritchie,[†] Beryl E. Moffatt,[‡] Robert B. Sim,[‡] B. Paul Morgan,[§] Raymond A. Dwek,[†] and Pauline M. Rudd*[,†]

The Glycobiology Institute and MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and Department of Medical Biochemistry, University of Wales, Cardiff, U.K

Received September 12, 2001

Contents

* Corresponding author e-mail: pmr@glycob.ox.ac.uk; telephone: 01865 275340; fax: 01865 275216.
[†] The Glycobiology Institute, Department of Biochemistry, University of Oxford.
[‡] MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford.
[§] Department of Medical Biochemistry, University of Wales.

1. Introduction

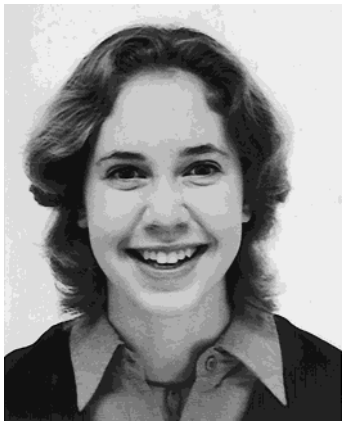
1.1. Overview of the Complement System

The human complement pathway is a highly efficient recognition and effector system that is designed to destroy infectious microbes and damaged host material. Over 30 plasma proteins or membrane-bound regulatory and receptor molecules are involved (Figure 1;[1] Tables 1 and 2),[2] and most of them are glycosylated. Together these proteins provide a highly regulated protease cascade and membrane pore-forming complex that targets foreign surfaces while host cells are protected by complement control proteins on their surfaces. This review focuses on the structures of the N- and O-linked oligosaccharides and on the contributions they make to the interactions and functions of the glycoproteins in the human complement system.

Initiation of the Complement Pathway. The cascade of reactions that follows complement activation can be initiated in at least three different ways. These are known as the alternative pathway (AP), the lectin pathway (LP), and the classical pathway (CP). The AP is activated by the continual "tickover" of C3[3] and involves the covalent deposition of thousands of copies of C3b onto a microbial target, a process called opsonization. Opsonization facilitates destruction of the microbe by allowing its recognition and engulfment by phagocytic cells via the C3b receptors on the phagocyte cell surface, but the deposited C3b can also associate with either AP or CP C3 convertases (C3bBb and C4bC2a respectively) to form the C5 convertases (C3bBbC3b and C4bC2aC3b) that leads to formation of the membrane attack complex (MAC) and subsequent microbial cell lysis. The LP allows activation of complement through the recognition of carbohydrate structures, on the surface of yeasts, bacteria, and viruses by mannose binding lectin (MBL). In the CP,[3,4] the recognition molecule, C1q, binds to antigen−antibody complexes, in particular to the Fc regions of clustered IgG and IgM[5,6] and to lipid A of Gram-negative bacteria. All three pathways can be activated in the presence or absence of antibodies.[3]

Complement Cascade Terminates with the Formation of the Membrane Attack Complex. Activation by all of these pathways results in the formation of multicomponent serine proteases called C3 convertases (C4b2a in the CP and LP and C3bBb in the AP) that can proteolytically cleave fluid-phase

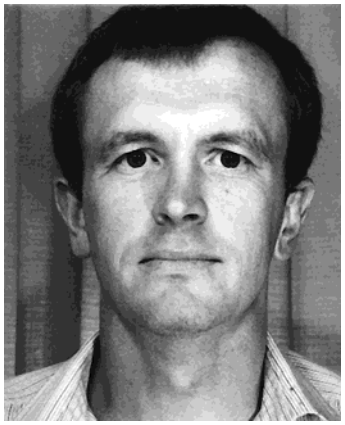
10.1021/cr990294a CCC: $39.75 © 2002 American Chemical Society
Published on Web 01/30/2002



Gayle E. Ritchie is currently studying for her Ph.D. at Linacre College, University of Oxford. She completed her M.Biochem. (Oxon) at Somerville College at the end of 2001, having done an undergraduate research project in the Oxford Glycobiology Institute under the supervision of Dr. Pauline Rudd, Dr. Raymond Dwek, and Dr. Bob Sim. She is continuing her studies at this department pursuing research into the role of the glycans attached to the Ebola virus glycoproteins.

Beryl E. Moffatt's (FIBMS, M.Sc.) scientific career began in the Health Service, with 19 years in the Oxford Regional Blood Transfusion Service and in Medical Pathology Laboratories in Oxford, London, and Northern Ireland. During this time, experience was gained in a broad spectrum of biomedical research, including early crossmatching procedures for kidney transplantation. Fellowship of the Institute of Biomedical Sciences was awarded with specialisation in hematology, blood transfusion, and bacteriology. She moved to a research environment in 1980 in the MRC Immunochemistry Unit, Department of Biochemistry, Oxford University. Her specialization is in the purification and functional assay of complement proteins, particularly C3, factor H, and factor I. She was awarded an M.Sc. in 1996, for studies on complement factor H.

Robert B. Sim was born 1951 in Crieff, Scotland. He graduated in 1973 from Edinburgh University (B.Sc.) and from Oxford University (Ph.D.) in 1976. He worked for INSERM, in Grenoble for 2 years, and then became a member of the U.K. Medical Research Council Scientific Staff in Oxford in 1978. His main research interests are in innate immunity, the recognition of microorganisms and altered host cells by innate immune components, and in investigation of mechanisms by which pathogens and parasites evade the human immune system.



B. Paul Morgan is a Professor of Medical Biochemistry at the University of Wales College of Medicine in Cardiff. He qualified in Medicine in 1980 and obtained a Ph.D. in 1984. After postdoctoral posts in Baltimore, MD, and Gainesville, FL, he returned to Cardiff as a lecturer in medical biochemistry. His research has focused on various aspects of the complement system and in particular its relevance to human disease. He was awarded a Personal Chair in 1995, elected a Fellow of the Royal College of Pathologists in 1997, and elected a Fellow of the Academy of Medical Sciences in 1998.

C3. The resulting C3b fragment undergoes a conformational change to expose a previously concealed thioester that enables the C3b molecule to attach covalently to the surface of targets bearing nucleophilic groups. This thioester is highly reactive so that the spatial range of covalent attachment is restricted to those surfaces in the immediate vicinity of C3 activation. Thus, C3b becomes predominantly deposited on the surface of potential pathogens rather than on host cells. C3b can interact with both types of C3 convertase and upon doing so changes the substrate specificity of the protease component of the complex (either C2a or Bb) to cleave C5 in preference to C3. The proteolytic cleavage of C5 generates C5a, a potent chemotactic factor and anaphylatoxin, and C5b, which is able to sequentially recruit the terminal components of the cascade (C6, C7, C8, and C9) to form the MAC. The MAC consists of a single C5b-8 complex and between 12 and 18 molecules of C9 that polymerize to tubular, transmembrane pores of about 100Å in diameter.[7,8]

1.2. Overview of N- and O-Glycosylation of Complement Components

Protein glycosylation is an important feature of most of the molecules that are involved in all three complement pathways (Table 2; Figure 2). Most of them are glycoproteins synthesized in the liver, macrophages, or lymphoid tissue (Table 2). They contain from one (C1q monomer) to eight N-glycosylation sites (C2, factor H), and some molecules such as DAF and CD59 that protect host cells from complement-mediated lysis contain O-glycans. All consist of a mixture of glycosylated variants (glycoforms). At one extreme, human erythrocyte CD59

Raymond A. Dwek is a Professor of Glycobiology, Director of the Glycobiology Institute, and Head of the Department of Biochemistry, Oxford University. He is a Professorial Fellow at Exeter College, Oxford. He obtained his B.Sc. (1963) and M.Sc. (1964) degrees at Manchester University and his Ph.D. (1966) degree at Oxford. He founded the Glycobiology Institute at Oxford University in 1991 and has received several awards for his work on Glycobiology, including the 7th Wellcome Trust Award for Research in Biochemistry Related to Medicine and the First Scientific Leadership Award, Hepatitis B Foundation, Philadelphia, PA. He is a member of the European Molecular Biology Organization and Fellow of the Royal Society. In 1996, he was awarded a Doctoris Honoris Causa by the Katholieke Universiteit, Leuven, Belgium, for his research contributions to NMR, antibodies, and glycobiology. In 2000, he was awarded the National Romanian Order for merit with rank of Commander for his major contribution to the Romanian–British cooperation in biochemistry and molecular biology. This year he has been awarded the Doctor Philosophiae Honoris Causa from the Ben-Gurion University of Negev, Israel, for his pioneering work in glycobiology.

Dr. Pauline M. Rudd is a Reader in Glycobiology in the University of Oxford. Her research group focuses on roles for glycosylation in the immune system and inflammation. In addition, over the past 5 years her group has pioneered the development and application of novel strategies for the rapid, sensitive identification of N- and O-linked sugars. The underlying technology involves HPLC analysis of exoglycosidase array digestions of total glycan pools released from glycoproteins and computer-assisted data analysis. Combined with protein structural data, glycan analysis has provided insight into the complete structure of a number of glycoproteins of current biological interest and their interactions with other molecules. Most recently these include molecules in the T-cell synapse, serum immunoglobulins, prion protein, CD59, and gelatinase B.

exists as over 150 glycoforms[9] while at the other end of the scale C9 has only three. Glycan processing is regulated by the levels of the glycosylation processing enzymes that vary with the cell type in which the glycoprotein is synthesized and, within a given cell, by the local 3D structure at each glycosylation site. These factors ensure that the surface of a single cell will display a diverse repertoire of oligosaccharides and that each glycoprotein displays the particular glycan structures needed for its range of essential functions. In general, these functions may include protein folding and assembly, quality control, ER-associated retrograde transport of misfolded proteins, and provision of recognition epitopes. By virtue of their size, glycans shield large regions of the protein surfaces, providing protease protection and limiting nonspecific, lateral protein−protein interactions. Oligosaccharides located close to the cell membrane, or in heavily glycosylated O-linked domains that provide molecular spacers, can determine the orientation and location of the binding faces of the proteins to which they are attached.

1.3. Classical Pathway

Activation of the CP can be antibody-dependent or antibody-independent and involves the binding of C1q via its globular head domains to target surfaces (Figure 3). The targets can be either foreign anionic molecules (antibody-independent activation) or the Fc regions of IgG (Figure 4), which are aggregated or IgM (antibody-dependent activation). Functional recognition results in the activation of the serine proteases associated with C1q, namely, C1r and C1s. In their proenzymic form, these proteases form a Ca^{2+}-dependent[10−12] tetrameric complex that associates with the collagen-like region of C1q. Activation of the proteases requires proteolytic cleavage at a conserved Arg−Ile bond; C1r is autoactivated following C1q binding and proteolytically cleaves and activates C1s. Activated C1s then cleaves both C4 and C2 with the resultant formation of the CP C3 convertase, C4bC2a. This convertase cleaves fluid-phase C3 generating reactive C3b molecules that deposit on the target surface and allow formation of the multicomponent CP C5 convertase, C4bC2aC3b. Cleavage of C5 by this convertase initiates formation of the MAC to bring about lysis of the potential pathogen.

1.4. Alternative Pathway

The AP is activated by deposited C3b molecules during "C3 tickover". The hydrolysis of the thioester bond in C3 by water in plasma results in formation of $C3(H_2O)$ (that has a conformation very similar to C3b) and allows association of $C3(H_2O)$ with the serine protease factor B; factor D proteolytically cleaves and activates factor B when factor B is present in a C3bB complex forming a fluid-phase AP C3 convertase. Deposited C3b is also able to interact with factor B to form surface-bound C3 convertases. The cleavage of more C3 generates the AP C5 convertase, C3bBbC3b, which then initiates MAC formation in the same way as the CP.

1.5. Lectin Pathway

The LP allows activation of complement via carbohydrate structures found on the surface of yeasts, bacteria, and viruses. It involves recognition of these structures via MBL, which is structurally similar to C1q. Like C1q, MBL (Figure 4) associates with modular serine proteases termed MASP1 and MASP2

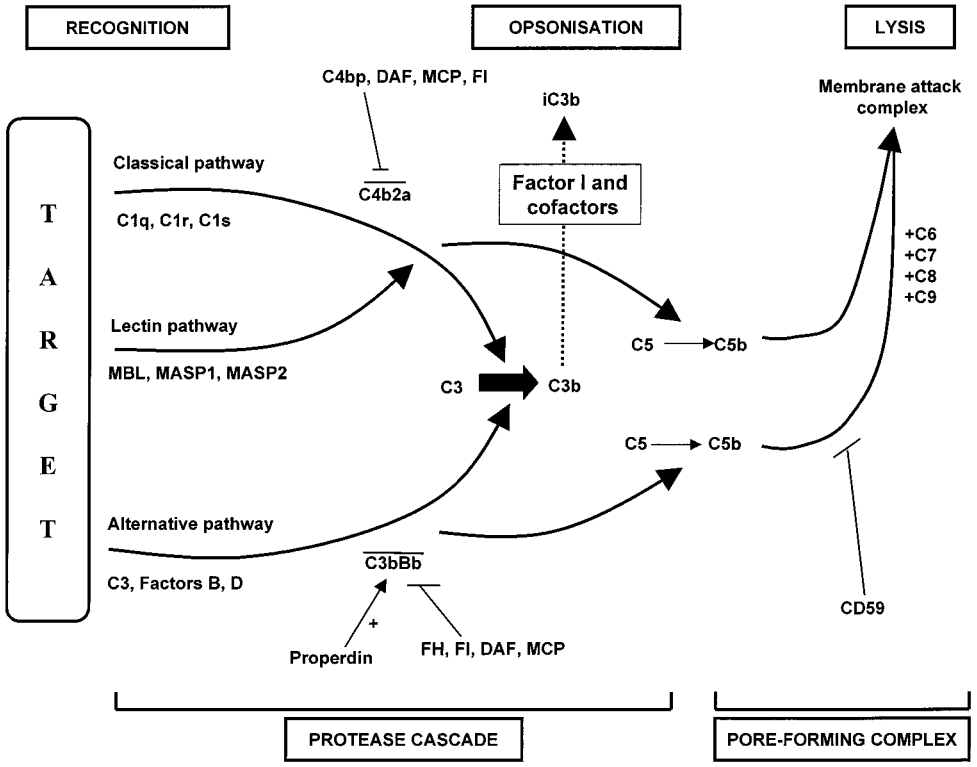


Figure 1. Complement cascade. The complement cascade can be activated by three distinct pathways. All pathways ultimately lead to the formation of multicomponent serine proteases (C3 convertases) that allows these three routes to converge to the activation of C3. Activated C3b can then covalently deposit on the surface of the target in the immediate vicinity of C3 activation; the target is said to be opsonized. Such opsonization facilitates phagocytosis of the invading microorganism. The deposited C3b also interacts with the C3 convertases to form the C5 convertase. This cleaves C5 into C5a, a potent anaphylatoxin, and C5b, which initiates formation of the membrane attack complex (MAC). Although C5 is highly homologous to both C3 and C4, C5b does not possess an internal thioester and is consequently not reactive toward covalent deposition on target surfaces. C5b interacts in a sequential manner with the terminal components of the pathway to form the MAC that forms transmembrane pores ~100 Å in diameter. These pores are formed by polymers of between 12 and 18 molecules of C9 in the target membrane and give rise to lysis and destruction of the potential pathogen. There are membrane-bound glycoproteins found on host cells that protect against the destructive action of complement. These include the GPI-anchored proteins CD59 and DAF. CD59 has a single extracellular protein domain that protects against complement-mediated lysis by binding to the C5b-8 complex during MAC formation. This prevents the recruitment, polymerization, and membrane insertion of C9, which prevents the formation of lytic pores. DAF is able to bind to both CP and AP C3 convertases (C4bC2a and C3bBb, respectively) and in so doing promotes their dissociation, preventing activation of complement on host surfaces. Adapted from ref 1.

(and a MASP3 has also been identified[13]) that are homologous in sequence to C1r and C1s.[14] MASP2 has shown proteolytic activity toward C4 and C2 and so allows formation of the C3 convertase in a manner similar to the CP but in response to a different target. Formation of the MAC then ensues in the same way as the CP.

1.6. Regulatory Factors

There are several serum regulatory proteins of the complement cascade. These include factor H, factor I, properdin, and C4b binding protein (C4bp) (Figure 1). Factors H and I and properdin all operate within the AP. Factors H and I are negative regulators, inhibiting amplification of the pathway, while properdin acts as a positive regulator, promoting activation of complement. Factor H binds to C3b molecules deposited on the surface of host cells using surface charge features to distinguish between self and nonself and acts as a cofactor for the serine protease factor I, which then cleaves C3b into an inactive form, iC3b. Properdin forms oligomeric structures that are able to bind to the AP C3 convertase (C3bBb) and stabilize the complex and inhibit its inactivation by factors H and I. C4bp works in an analogous fashion to factor H within the CP. It binds to C4b within the CP C3 convertase (C4bC2a) and acts to promote dissociation of these components as well as acting as a cofactor for factor I to allow cleavage and inactivation of C4b.

Complement can act independently of the adaptive immune response, but it can also interact with elements of the adaptive immune response. Proteins within the cascade are able to signal to lymphocytes; C3b deposition can occur on the Fc regions of aggregated IgG and IgM and so promote solubilization and clearance of antibody−antigen complexes (ICs);[15] C1q directly activates complement via interaction with the Fc regions of aggregated IgG and IgM, and IgG-containing ICs activate the AP.

Within the cascade, proteins of analogous function display a high degree of homology and are extensively glycosylated (Table 1; Figure 2). Although highly homologous, these proteins also show a high degree of specificity of interaction with their various partners; therefore, glycosylation of these proteins may

Table 1. Main Plasma Components of Complement Together with Their Molecular Weights, the Number of Potential N-Linked Glycosylation Sites, Their Primary Site of Synthesis, and Their Plasma Concentration

protein	no. of N-linked sites	M_r (kDa)	primary site of synthesis	plasma concn (μg/mL)
C1q	1 (only on A chain)	monomer: 75	macrophage	80
C1r	4	intact: 85 A-chain: 50 B-chain: 35	liver	50
C1s	2	intact: 80 A-chain: 50 B-chain: 30	liver	50
C2	8	intact: 100 C2a fragment: 70 C2b fragment: 30	liver	20
C3	3	intact: 185 α-chain: 115 β-chain: 75	liver	1500
C4	4	intact: 205 α-chain: 97 β-chain: 75 γ-chain: 32	liver	600
C5	4 (only 2 sites thought to be occupied)	intact: 188 α-chain: 115 β-chain: 75	liver	70
C6	2	104	liver	64
C7	2	110	lymphoid tissue	56
C8	2	150	liver	55
C9	2	71	liver	59
factor B	4	intact: 90 Ba fragment: 30 Bb fragment: 60	liver	210
factor H	9	155	liver	500
factor I	6	intact: 90 A-chain: 50 B-chain: 38	liver	35
properdin	1	monomer: 55	monocytes, T-cells, granulocytes	5
C4bp	3	monomer: 70 heptamer: 490	liver	250

Table 2. Main Domain Types Found in Proteins of the Complement System

abbreviation	name of module	occurs in
CCP/SCR	complement control protein/short consensus repeat	MASP1, MASP2, C1r, C1s, C2, FB, FH, DAF, MCP, CR1, CR2, C4bp
FIMAC	factor I/membrane attack complex	FI, C6 C7
TSR	thrombospondin repeat	properdin, C6, C7, C8, C9
vWF-A	Von Willebrand factor type A	FB, C2
MACPF	membrane attack complex/perforin-like	C6, C7, C8, C9
LDLr	LDL receptor-like	FI, C6 C7, C8, C9
EGF-like	epidermal growth factor-like	MASP1, MASP2, C1r, C1s, C6, C7, C8, C9
SP	serine protease	MASP1, MASP2, C1r, C1s, C2, FB, FI, FD
CD5 or SRCR	scavenger receptor cysteine-rich	FI
CUB	complement subcomponents C1r/C1s, Uegf, bone morphogenetic protein-1	C1r, C1s, MASP1, MASP2
STP	serine−threonine−proline-rich domain	DAF, MCP

be a significant factor in the regulation and specificity of their interactions.

2. Cell-Specific Glycosylation of Complement Components

See Figure 5 for common glycan structures and their nomenclature.

2.1. Complement Components Synthesized Mainly in the Liver Contain Predominantly Complex Biantennary Glycans

Different proteins made within the same organ are exposed to the same set and expression levels of glycosylating enzymes yet they can show different glycosylation profiles. This reflects the influence of the local 3D protein structure on the accessibility of the glycan sites to the glycosylating processing enzymes. Although most complement components synthesized predominantly in the liver (C1r, C1s, C2, C3, C4, C5, C6, C8, C9, FB, FH, FI, C4bp) contain complex biantennary glycans with varying degrees of sialylation (e.g., Figure 6 panels a−c; C3, C4, and C5α, respectively), C3 (Figure 6a) contains only oligomannose type glycans.[16] This suggests that the N-linked sites on C3 are relatively inaccessible to the glycan processing enzymes, in particular to mannosidases. Some of the components contain low levels of core fucosylated glycans. For example, 17% of the sugars attached to C1s contain core fucose, consistent

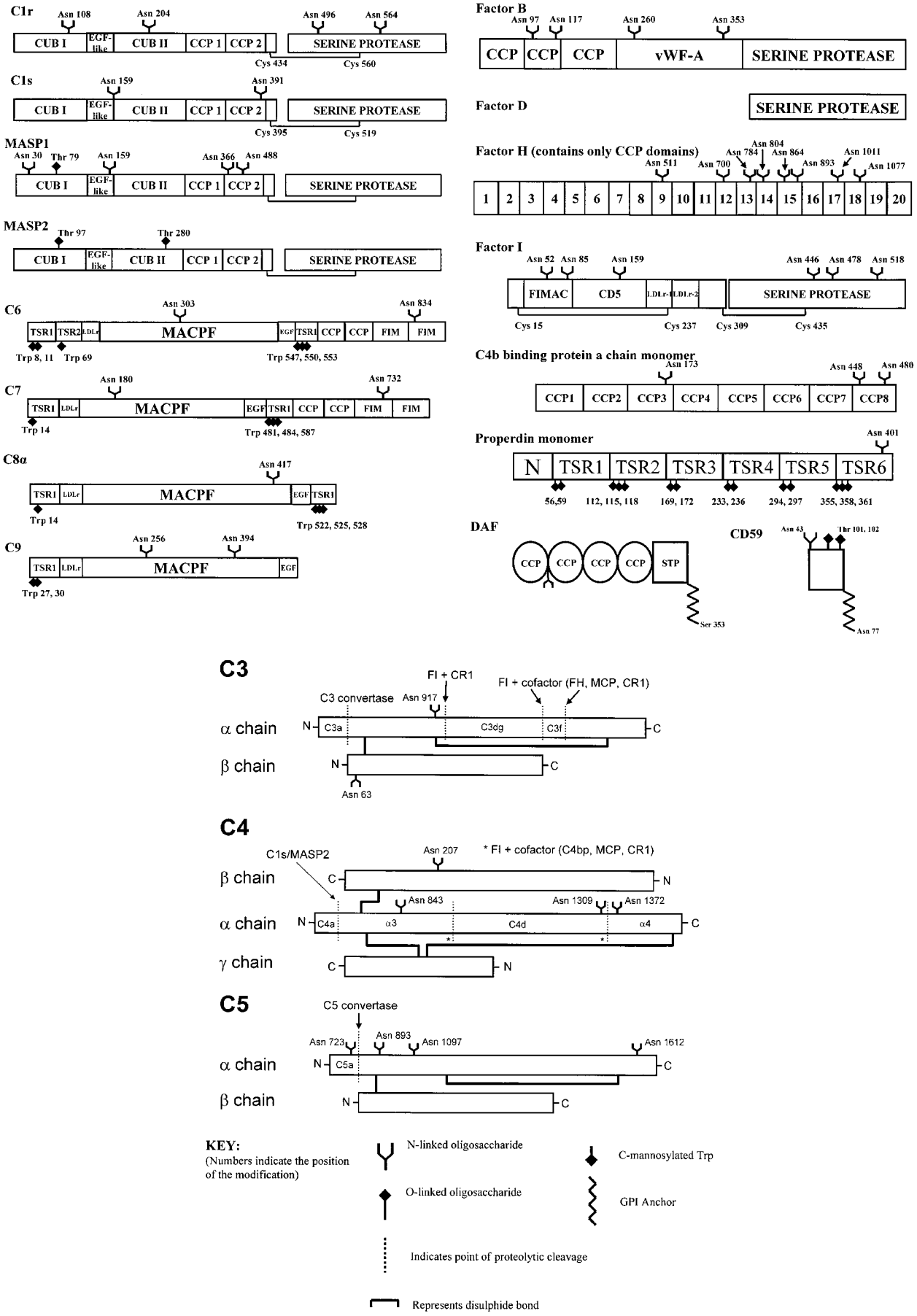


Figure 2. Domain organization of proteins in the complement cascade indicating potential glycosylation sites. C1s has two potential O-linked glycosylation sites at Ser 489 and Thr 544 in the B chain, but this chain of C1s has been shown to undergo no post-translational modification.

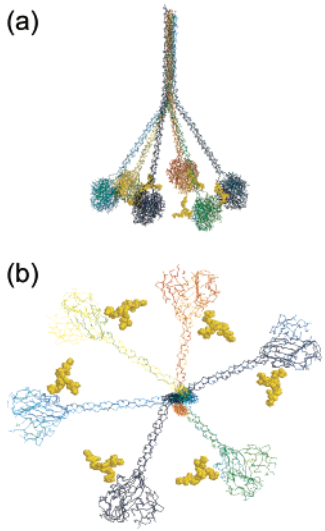


Figure 3. Molecular model of C1q. (a) Side view of hexameric C1q; (b) top view of hexameric C1q. The glycans modeled on the molecule are A2G2FS(α1,6) and are represented in yellow in the spacefill format. The chains are colored separately for clarity. The O-linked oligosaccharides of the collagen-like region are omitted for emphasis on the N-linked sugars. This represents one possible model for the position of the N-linked glycans. As panel b shows, there is potential for these oligosaccharides to be involved in regulating interactions between adjacent globular "heads". The glycans may enable these domains to maintain their ring-like orientation and so to maximize interaction with target polyanions (Fc regions of clustered IgG or IgM, or circulating DNA or lipid A). The crystal structure of the rat Acrp30 protein was used to model the globular head region of C1q; the structure of a $(Gly\text{-}Pro\text{-}Pro)_{10}$ was used to build up the collagen-like region. The N-linked glycan was attached to Gly 148 of one of the heads of the rat Acrp30 protein, the equivalent position of Asn 146 in human C1q.

withX-ray crystallography data of a fragment of C1s consisting of the second CCP module and the SP domain, which revealed electron density corresponding to the protein proximal GlcNAc and core fucose.[17] C4bp (A-chain) and the α- and β-chains of the trimeric C8 component contain fewer sialylated glycans than many of the other components made in the liver, suggesting that sialyl transferases may have limited access to the glycosylation sites; C4bp (A-chain) and C8β (Figure 7) also contain bisected structures while many other complement components do not.

2.2. C1q, Properdin, and C7 Expressed Mainly in the Lymphoid Tissue Contain Core Fucosylated Glycans

In common with components synthesized predominantly in the liver, those derived from the lymphoid organs such as C1q[18] (Figure 8), properdin, and C7 contain complex biantennary glycans. However, these glycoproteins are invariably core fucosylated, in striking contrast to those derived principally from the liver that contain only low levels of core fucosylation ranging from 0% in C3 to 17% in C1s.

2.3. C9 Expressed in *Trichoplusia ni* Cells Contains Oligomannose Glycans

Recombinant C9 has been expressed in *Trichoplusia ni* cells with the result that the complex sialylated

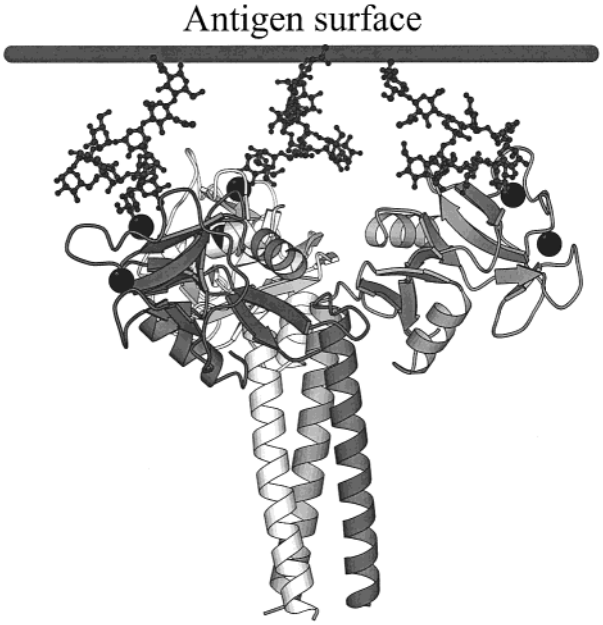


Figure 4. Model of the MBL showing the neck region and the C-type lectin domains of a MBL monomer interacting via these carbohydrate recognition domains (CRDs) with *cis*-diols located on the terminal mannose residues.[4] The dark spheres represent Ca^{2+} ions, one of which is active in coordinating to the terminal sugar residue via the equatorial 3′- and 4′-hydroxyl groups of the sugar.[49]

biantennary glycans found in C9 from human serum have been replaced by neutral high mannose glycans.[19] Such changes in the glycan composition have no direct effect on the physiological function or polymerization of C9.[19] However, this does not preclude the possibility that, as in factor H,[20] the sugars may be required in the quality control for protein folding prior to secretion.

Expression of complement proteins other than C9 in insect cell systems has shown a variety of results. Expression of FI in *T. ni* cells results in a protein with only 55% of the activity of the serum protein as measured by $C3(NH_3)$ cleavage assays;[21] this has been attributed to the changed glycosylation of the protein that resulted from expression in insect cells. Infection of *Spodoptera frugiperda* Sf-9 insect cells with recombinant baculovirus expressing C1 esterase inhibitor (C1INH) gives rise to recombinant expression of this CP complement regulatory protein.[22] The recombinant C1INH has just less than half the specific activity of the native protein purified from human serum. In contrast, expression of FH in Sf-9 insect cells[23] and expression of C8 in *T. ni* cells[24] gives rise to proteins with comparable activities to the native protein.

3. Roles for Glycans in the Function and Interactions of Complement Proteins

3.1. Active Conformation of C1q May Be Maintained by Its Oligosaccharides

The O-linked glucosyl−galactose disaccharides within the collagen-like region of C1q have been implicated in stabilizing the triple helix region of each monomer and may also stabilize the hexameric assembly.[25] A single N-linked glycosylation site exists

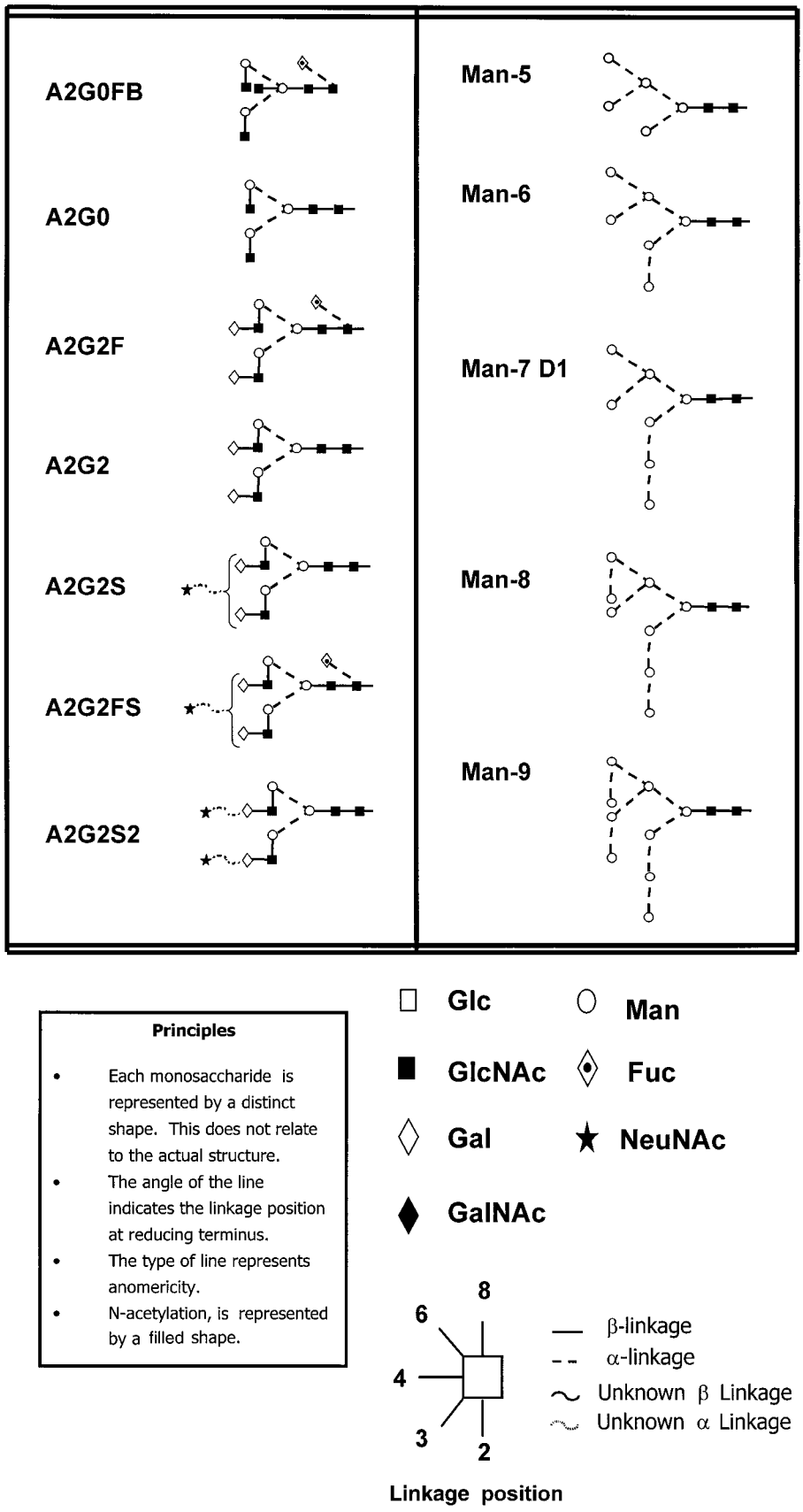


Figure 5. Commonly occurring glycans and their nomenclature A2 indicates two antennae, G0 indicates the absence of Gal, G2 indicates the presence of Gal (one on each antenna arm), F indicates the presence of core fucose, B indicates the presence of bisecting GlcNAc, and S indicates the presence of sialic acid. Most complex type glycans have the terminal sialic acid in α2,3 linkage to Gal.

on each A-chain, located in the globular head domain (Figure 3); thus, there are six N-linked glycans in fully assembled C1q. C1q is relatively flexible,[26] and for the globular heads to bind effectively to their

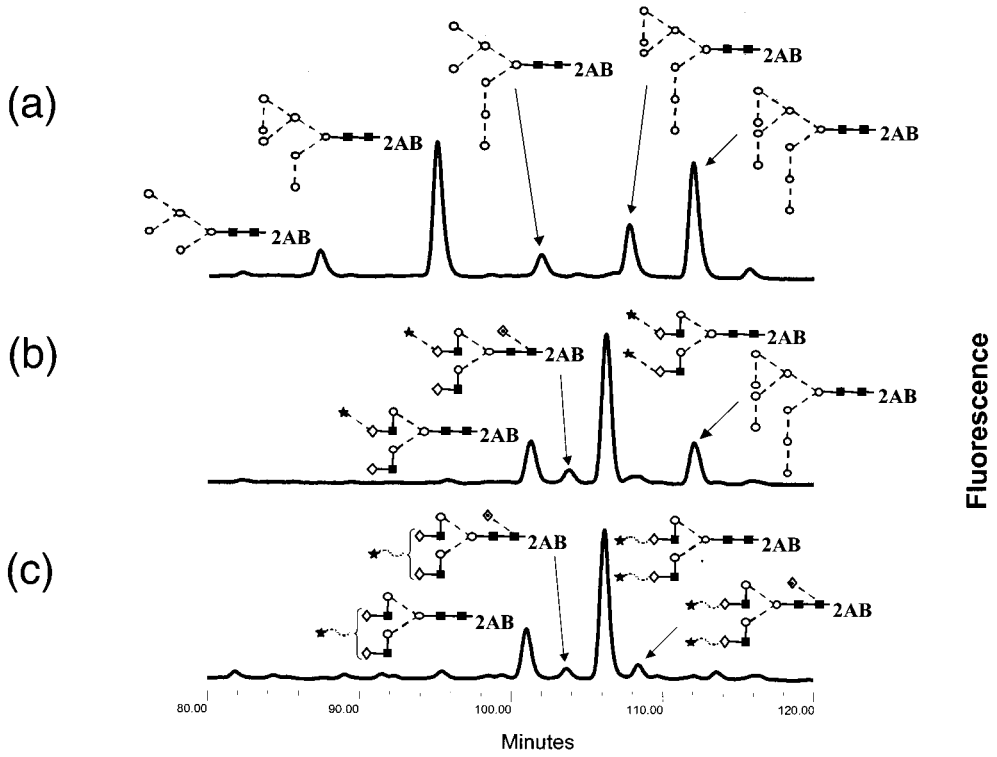


Figure 6. Comparison of glycan profiles of homologous proteins C3, C4, and C5. HPLC profiles of N-linked glycans released from (a) C3, (b) C4, and (c) C5 α-chain. Despite the homology between these proteins, their NP-HPLC glycan profiles are very different. C3 has only oligomannose structures, C4 has both oligomannose and complex glycans, while C5 has only complex glycans. Glycans were released and analyzed as described previously;[50] see legends to Figure 7 and Figure 8.

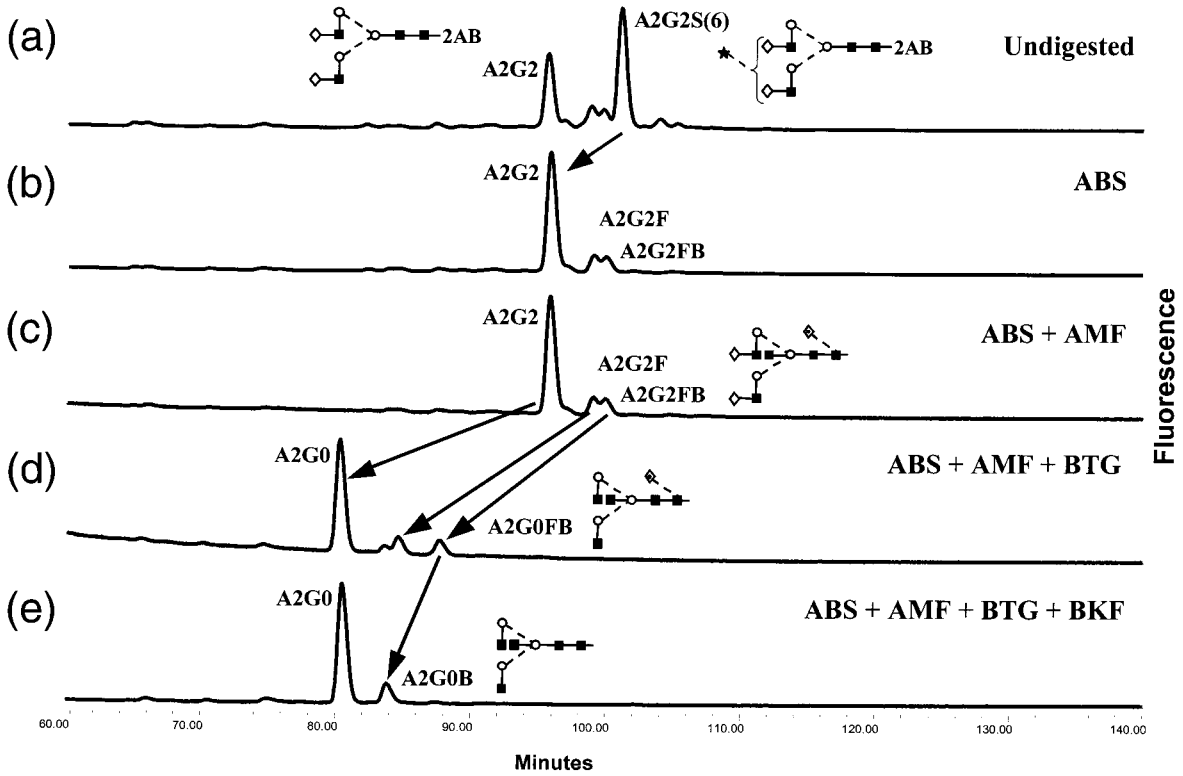


Figure 7. Bisected N-glycans released from C8β. The presence of bisected N-glycans released from the C8β chain was detected using an exoglycosidase enzyme array containing *Arthrobacter* sialidase (ABS), outer arm fucosidase (almond meal fucosidase, AMF), bovine testes β-galactosidase (BTG), *Streptococcus pneumoniae* β-*N*-acetylhexosaminidase (SPH), and core fucosidase (bovine kidney fucosidase, BKF). In panels b−e, the bisected structures are labeled. The assignments were made on the basis of the elution positions of the peaks as compared with standard glycans and also on the results of exoglycosidase digestions.

targets, there may be a requirement for keeping these head regions apart. From molecular modeling (Figure 3), it can be seen that the sugars at Asn 146 have the potential to keep these domains separate, mini-

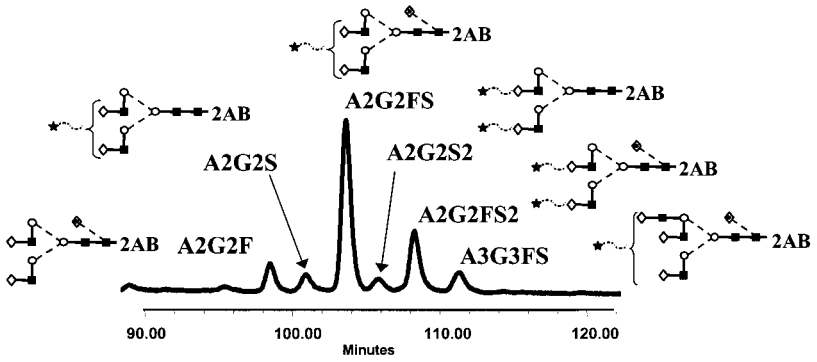


Figure 8. HPLC profile of N-glycans released from C1q showing the presence of N-glycans that predominantly contain fucose at their core. Glycans were released and analyzed according to the method previously described.[50,51] Briefly, protein bands corresponding to the C1q A chain were cut from a 10% SDS−PAGE gel stained with Coomassie blue. The sugars were cleaved from the protein by incubating the chopped up gel bands with peptide N-glycosidase F (PNGaseF). The sugars were then recovered by washing the gel pieces and were fluorescently labeled with 2-amino benzamide (2-AB) for detection on NP-HPLC. Sequential exoglycosidase digestion of the fluorescently labeled sugars using combinations of sialidase, outer arm fucosidase, β-galactosidase, β-*N*-acetylhexosaminidase, and core fucosidase and subsequent analysis on NP-HPLC enabled determination of the sequence of sugars attached to the C1q A chain. The assignments were made on the basis of the elution positions of the peaks as compared with standard glycans and also on the results of exoglycosidase digestion.

mizing nonspecific protein−protein interactions between the domains.

3.2. Structural Implications of the Differences in Glycosylation for C3 and C5

C3 and C5 have about 70% sequence homology. Despite this homology, these proteins show very different patterns of glycosylation (Figure 6). C3 contains only high mannose glycans with predominantly $Man_8GlcNAc_2 + Man_9GlcNAc_2$ on the α-chain and predominantly $Man_6GlcNAc_2$ on the β-chain.[16,27] Though C3 contains three potential N-linked sites, only two are occupied since Asn 1595 on the α-chain has been shown by conglutinin binding studies not to be glycosylated.[28] The patterns of glycosylation obtained from analysis of the α- and β-chains of C3 therefore represent the glycans present at Asn 917 and Asn 63, respectively. C5, however, contains only sialylated complex biantennary glycans. The restricted repertoire of oligomannose structures on C3 suggests that the glycosylation sites are relatively inaccessible. In contrast, the presence of complex biantennary structures on C5 indicates glycans that have been trimmed back and further processed.

3.3. Factor B Glycans Are Involved in Regulating Interactions with C3b

Factor B (FB) has four potential N-linked glycosylation sites, all containing sialylated complex biantennary sugars (Figure 9). Site directed mutagenesis studies within the vWF-A domain in FB has revealed a potential role for Asn 260. A conserved region of this domain was investigated by single and multiple mutations of various residues.[29] The single mutation of D254→A254 increased the ability of FB to initiate MAC formation (described as FB's haemolytic activity) with no change in C3b binding; mutation of N260→D260 had no effect on the haemolytic activity or the C3b binding affinity of FB. However, mutation of both D254A and N260D resulted in an increase in binding to C3b as well as in haemolytic activity. This suggests that the N-linked oligosaccharide may be involved in the regulation of FB binding to C3b. That there is increased C3b binding in the absence of the oligosaccharide suggests that the sugars sterically obstruct the C3b binding site so that C3b binding is reduced in native FB as compared to the double mutant. Strong interaction of FB with C3b followed by cleavage of FB by factor D (FD) results in formation of the C3 convertase. If this interaction was not regulated, active forms of the convertase would be formed that would inappropriately activate complement. Regulation of the FB−C3b interaction therefore limits complement activation which would lead to tissue damage and inflammation. The other N-linked site in the vWF-A domain of FB is at Asn 353. This residue is located at the opposite side of the domain to Asn 260 and close to the FD cleavage site of FB, and in common with Asn 260, the glycan at this position may sterically regulate FD binding to FB. The FD cleavage site may be concealed such that binding of FB to C3b elicits a conformational change within the vWF-A domain[30,31] that is transmitted to the other side of the domain exposing the cleavage site. The N-linked oligosaccharide at Asn 353 could act, perhaps by its dynamic flexibility, to conceal this site thus preventing inappropriate activation of the AP.

3.4. Complement Regulatory Factors

The regulatory factors (factor H, factor I, C4b binding protein, and properdin) from human serum all contain complex biantennary glycans without bisecting GlcNAc and with only low levels of core fucosylation (Figure 9).

3.4.1. Factor H and the Calnexin Pathway

In the early stages of glycan processing, glycoproteins carry the $GlcNAc_2Man_9Glc_1$ motif that enables it to engage with the lectin-like chaperone, calnexin (Clx). In its role as a quality control factor, Clx retains unfolded glycoproteins in the ER until they are correctly folded and assembled, an event that is signaled by the permanent removal of the terminal

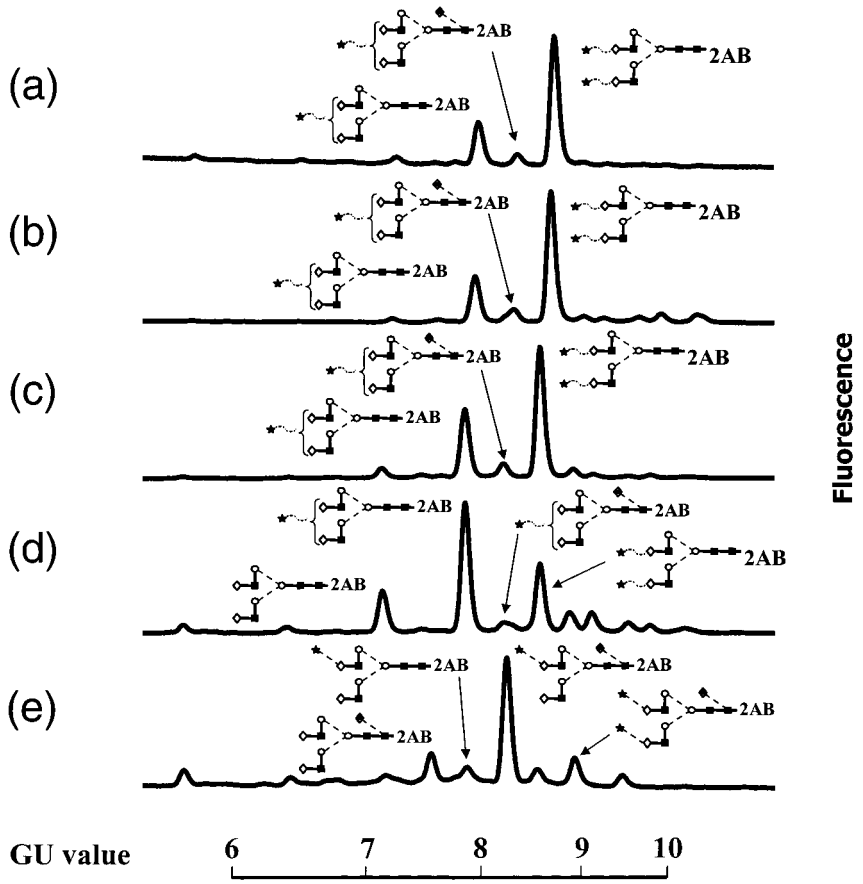


Figure 9. HPLC profile of glycans released from (a) factor H, (b) factor B, (c) factor I, (d) C4b binding protein, and (e) properdin. Glycans were released according to the method previously described;[51] see legend to Figure 8.

glucose residue by glucosidase II. The folded glycoprotein is then transferred to the Golgi apparatus where the sugars are further processed. Misfolded proteins are re-glucosylated by UDP-glucose:glycoprotein glucosyltransferase. Re-glucosylation allows unfolded proteins to rebind Clx thus entering a cyclical pathway that retains them in the ER until either they achieve their correctly folded structure and are released, or they are targeted for retrograde transport and degradation. In a child with chronic hypocomplementemic renal disease, factor H contained two mutations that each affect conserved cysteine residues and therefore protein folding. In this patient, factor H co-localises with calnexin, indicating that the sugars attached to misfolded factor H bind to the chaperone and are retained in the ER.[20]

3.4.2. Interaction of FI with C3b and FH Are Regulated in Part by FI Glycans

Factor I (FI) is involved in regulating the formation of both CP and AP C3 convertases by proteolytically cleaving C4b or C3b to an inactive form and thus inhibiting formation of an active convertase. FI circulates as an active enzyme and displays very high substrate specificity for either C4b or C3b, acting only in the presence of specific cofactors (C4bp, FH, membrane cofactor protein, complement receptor-1). FI contains predominantly complex biantennary glycans, 46% of which are disialylated and 26% of which are monosialylated glycans (Figure 9). In contrast, recombinant FI expressed in insect cells contains neutral oligomannose glycans. This reduces the haemolytic activity of FI to 55% that of the serum protein,[21] demonstrating that sugars can be directly involved in regulating protein−protein interactions. The interaction of FI with both FH and C3b is predominantly ionic in nature,[32] and so the presence of the neutral sugars may weaken the interaction between FI and its cofactors.

3.4.3. Glycans of Properdin May Impart Stability to the Protein

Properdin has a single N-linked oligosaccharide (Figure 9) located in the sixth TSR at Asn 428 (Figure 2). This site is located within a unique 25 amino acid insert in this domain. Although properdin lacking the N-linked site was secreted and functionally active, deletion of the sixth TSR in which the N-linked glycosylation site is found compromised the formation of properdin oligomers.[33]

3.5. Decay Accelerating Factor (DAF) and CD59

DAF and CD59 (Figure 10b, c) are glycosylphosphatidylinositol (GPI) anchored glycoproteins that protect cells from complement-mediated damage. The active site of CD59 is located close to the cell surface at 3.5 nm from the membrane while that of DAF is further away, at 16 nm. CD59 binds the C5b-8 component of the MAC, thus preventing the polymerization of C9 and the formation of a lytic pore. DAF dissociates both the CP and the AP C3/C5 convertases and inhibits their assembly (Figure 11).

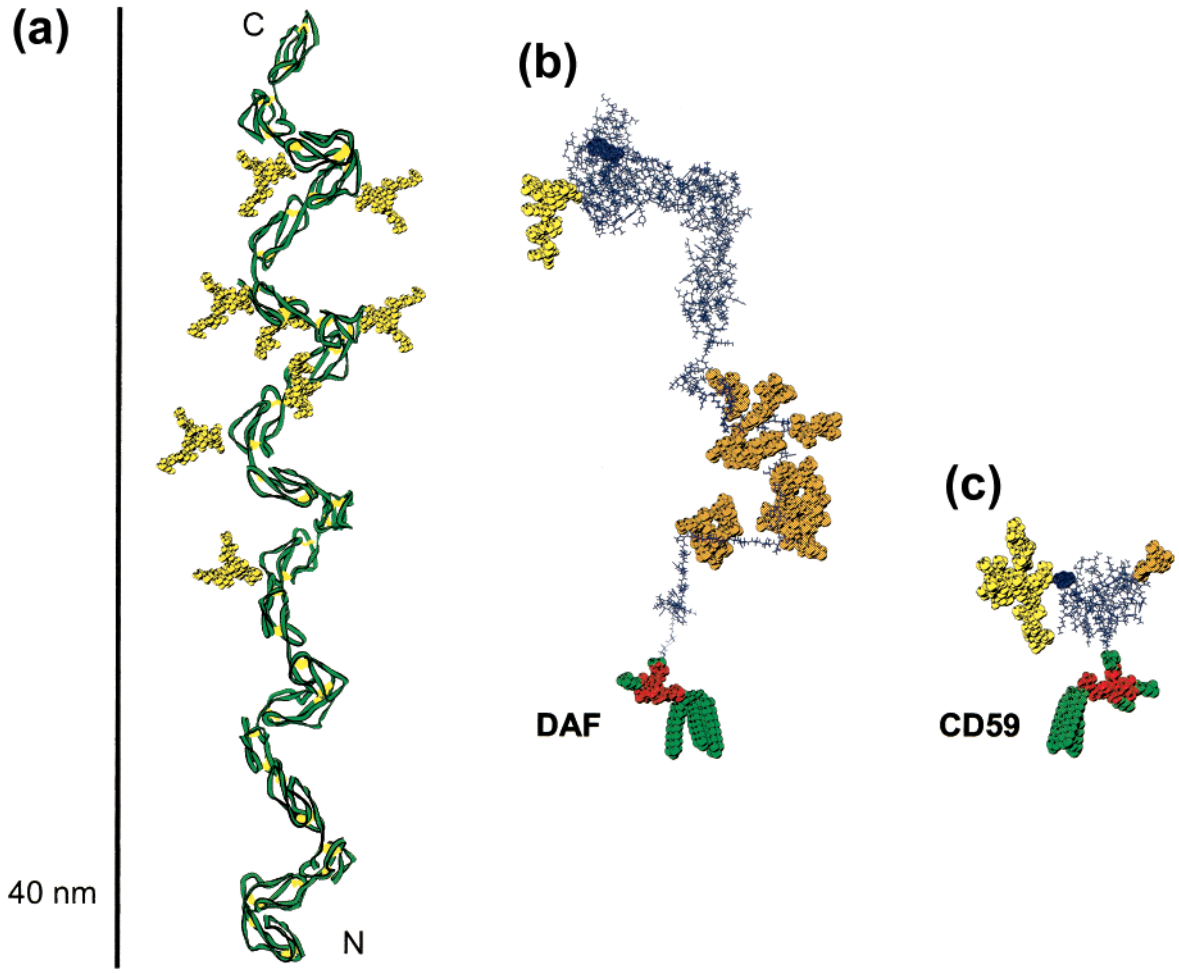


Figure 10. Molecular models of (a) factor H, (b) DAF, and (c) CD59. Panel a is a schematic representation of the FH molecule depicting all 20 CCP domains and the glycosylation sites. A recent model of FH generated from X-ray and neutron scattering data has proposed that the domains of this molecule adopt a nonlinear arrangement such that the molecule appears folded back on itself.[52] The yellow spacefill atoms represent N-linked glycans; the orange represent O-linked glycans.

DAF is composed of four CCP domains suspended above the membrane by a heavily O-glycosylated Ser/Thr-rich region of approximately 70 amino acids that is linked to the GPI anchor. The regulatory activity against C4b2a is localized in CCP domains 2 and 3 while its activity against C3bBb also involves CCP domain 4. There is one N-glycosylation site that is located between CCP1 and CCP2. The N-glycan site has been deleted without any associated loss of activity.[34] However, deletion of the Ser/Thr-rich region eliminated DAF function. This was restored in a fusion construct in which the four CCP domains were added to the HLA B44 molecule, suggesting that the O-glycosylated region serves as an important protease resistant spacer which projects the DAF functional domains above the plasma membrane.[34]

Host cells are normally protected from lysis by CD59, a cell surface glycoprotein that binds to the complement proteins C8 and/or C9 in the nascent MAC. CD59 belongs to the Ly-6 superfamily and is present on a wide variety of cell types, including leukocytes, platelets, epithelial and endothelial cells, placental cells, and erythrocytes. CD59 is attached to the surfaces of these cells by means of a GPI anchor containing three lipid chains.[35,36] Human erythrocyte CD59 consists of a heterogeneous mixture of more than 120 glycoforms of which the major single sugar is a complex glycan containing both a bisecting GlcNAc residue and a core fucose.[36] A population of sialylated O-glycans was recovered from human erythrocyte CD59; the major species that were identified were two monosialylated forms of the disaccharide Galβ1,3GalNAc. The GPI anchor glycans contain the trimannose sugar common to all mammalian anchors analyzed to date, 90% of which are not further processed.

The N-linked oligosaccharides (size range 3−6 nm in length) are attached to the disklike extracellular region of CD59 (diameter approximately 3 nm) and project away from the protein domain in the plane of the active face and adjacent to the membrane surface (Figure 10c). The glycans do not appear to restrict access to the proposed active site residues of human CD59 (Asp 24, Trp 40, Arg 53, and Glu 56) located on the membrane distal surface of the extracellular domain.[37] However, the glycans would be expected to restrict the rotational freedom of the extracellular domain around axes parallel to the membrane which may, in turn, stabilize an exposed location for the active face. Removal of the conserved N-linked glycan might therefore reduce the affinity of CD59 for the MAC without eliminating it completely. The effects of removing the N-linked glycan might therefore be expected to depend on the density

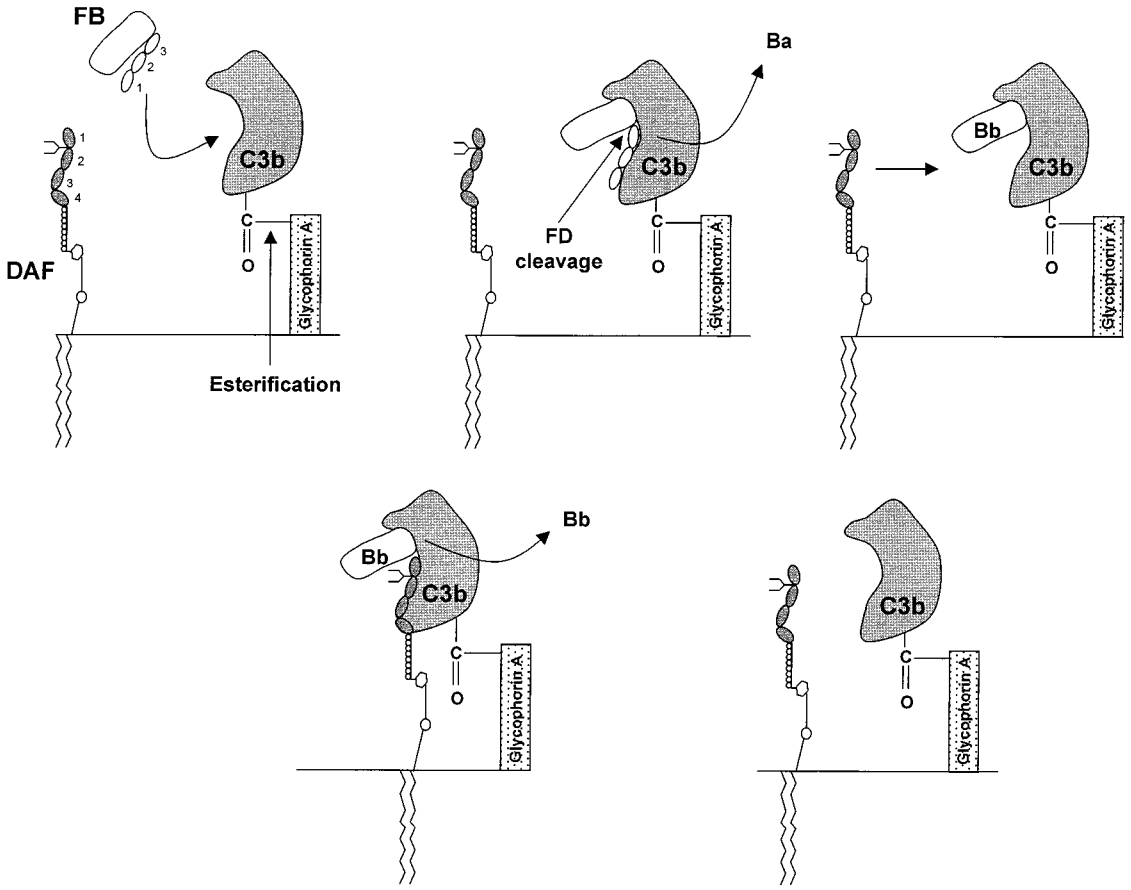


Figure 11. Schematic showing a proposed mechanism for the destabilization of C3b and Bb by DAF. In this model, DAF and the Ba fragment of factor B (FB) share common binding sites on the C3b fragment. The cleavage of FB by factor D (FD) releases Ba and reveals the underlying DAF binding site on C3b to which DAF binds and causes dissociation of Bb, resulting in protection against complement-mediated cell lysis. The regulatory activity of DAF against the CP C3 convertase (C4bC2a) has been shown to reside in CCP domains 2 and 3, while activity against the AP C3 convertase (C3bBb) is dependent additionally upon the fourth CCP domain.[53] The numbered domains are all CCP domains (Table 2), numbered from the N-terminus of DAF and FB. Adapted from ref 54.

of expression of the glycoprotein at the cell surface, and this may explain the observed variation in the activities of unglycosylated CD59.[38,39]

The heterogeneity of the sugars suggests that the glycans influence the geometry of the packing, and it is likely that they will also prevent the aggregation of CD59 molecules on the cell surface. By limiting nonspecific protein−protein interactions and controlling the spacing, the glycans may influence the distribution of CD59 molecules at the cell surface where GPI anchored proteins may associate in microdomains in dynamic equilibrium with isolated individual molecules. The large N-glycans may also be important in preventing proteolysis of the extracellular domain since N-glycosylation has been shown to increase the dynamic stability of a protein while different glycoforms variably increase its resistance to protease digestion.[40]

4. Other Types of Glycan Modifications

4.1. Nonenzymic Glycation

As well as N-linked glycosylation at Asn present in the sequon Asn-Xaa-Ser/Thr (where Xaa = any amino acid but not proline),[41] other means exist by which a protein can be modified by glycosylation. For example, oligosaccharides (often di-, tri-, or tetrasaccharides) may be attached to the hydroxyl group of surface exposed Ser or Thr. In addition, proteins can be modified nonenzymically by the addition of glucose (Glu) a process known as glycation. This often occurs when the level of Glu in the blood rises under pathological conditions, such as diabetes mellitus. Factor B and CD59 have both been shown to be glycated on specific Lys residues when there is an elevated level of blood glucose and this decreases their activity.[42,43]

4.2. Properdin and Terminal Complement Components Are C-Mannosylated

C-mannosylation[44] involves the C-glycosidic attachment of a single mannose residue to the indole moiety of Trp residues found within the sequon Trp-Xaa-Xaa-Trp[45] (where X is any amino acid; Figure 12). The enzyme involved is a microsome-associated mannosyl transferase that uses dolichol−phosphate−mannose[44] as the activated donor to form the unusual C−C bond to the protein. Properdin[46] and the terminal components of complement (C6, C7, C8, and C9) contain this modification within their TSR domains.[47] The exact role of this modification is not clear, although it is interesting that this modification is present in proteins of a system that contains the

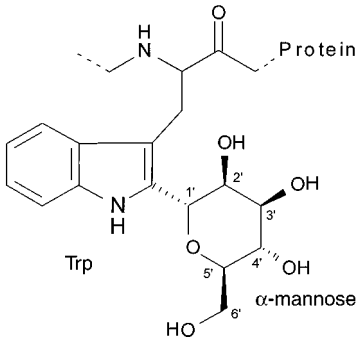


Figure 12. Structure of α-mannosyltryptophan. The α-mannose residue is attached to the indole moiety of tryptophan (Trp) via a C-glycosidic bond that involves C2 of Trp and C1 of α-mannose. C-mannosylation is an enzymatic modification carried out by a microsomal-associated mannosyl transferase. Dolichol−phosphate−mannose is used by the enzyme as the activated precursor for the addition of mannose to Trp. No enzyme has yet been identified that is able to remove α-mannose from Trp.

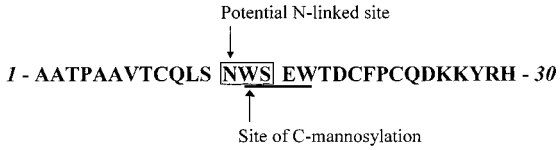


Figure 13. N-terminal sequence of C8α. Boxed sequence indicates the N-linked sequon; the underlined sequence indicates the C-mannosylation sequon.

MBL; there are also receptors that bind mannose, for example, the macrophage mannose receptor. This modification may therefore allow uptake and regulation of complement proteins. Some pathogenic microbes possess mannose-binding receptors on their surface. The presence of mannose on properdin and the terminal components of the pathway could allow localization of complement components to the surface of the microbe facilitating its destruction.

Within C8α, there is a potential N-linked glycosylation site in the first TSR that overlaps with a sequon for C-mannosylation (Figure 13) that is modified by addition of mannose.[25] The group investigating the C-mannosylation of this protein obtained a fragment corresponding to residues 10−20, which contains the two sequons, although they did not identify any attached N-linked oligosaccharides. It may be that C-mannosylation precedes the addition of N-linked sugars and thus precludes the attachment of N-glycans to these sequons.

5. Conclusions and Future Outlook

Glycosylation is one of the most common means by which proteins can be posttranslationally modified and is also the largest and most extensive type of protein modification. The key to understanding the contribution of an oligosaccharide to protein function is to view the molecule in its entirety. This requires information on the structure of the protein, the glycans, and any other post-translational modifications. Though many protein modules of which the complement components are composed have been crystallized and X-ray data are accumulating, there are still many for which there is only incomplete protein information. As these data become available, the glycans can be modeled at their appropriate locations to give insights into the ways in which the complement proteins interact with their cofactors.

One interesting possibility is that glycosylation might modify enzyme activity in the complement cascade. In the case of tPA, which initiates the protease cascade associated with remodeling of the extracellular matrix in which other glycosylated molecules such as plasminogen,[48] procollagenase, stromelysin, and gelatinases are involved, there is a 2-fold difference in fibrin-dependent generation of plasmin depending on the glycosylation status of both the tPA and the plasminogen. It is interesting to speculate that further amplification of the range of activity might result from the interactions of specific sets of glycosylated variants of all the enzymes. If it proves to be so, then differential glycosylation would allow these proteins a means to modulate the kinetics of such a cascade.

6. Abbreviations

AP	alternative pathway
Asn	asparagine
Clx	calnexin
CP	classical pathway
DAF	decay accelerating factor
Fc	constant nonantigen-binding region of antibody
FH	factor H
FI	factor I
Gal	galactose
GlcNAc	*N*-acetylglucosamine
GPI	glycosylphosphatidylinositol anchor
Glu	glucose
IC	immune complex
LP	lectin pathway
Lys	lysine
MAC	membrane attack complex
MASP	MBL-associated serine protease
MCP	membrane cofactor protein
MBL	mannose-binding lectin
NP-HPLC	normal-phase high-performance liquid chromatography
Ser	serine
Thr	threonine
Trp	tryptophan
tPA	tissue plasminogen activator
TSR	thrombospondin repeat domain

7. Acknowledgment

The authors thank Mark Wormald for his contribution to the molecular modeling in this review. P.M.R. thanks H. C. Lister for her constant inspiration.

8. References

(1) Sim, R. B.; Laich, A. *Biochem. Soc. Trans.* **2000**, *28*, 545.
(2) Law, S. K. A.; Reid, K. B. M. *Complement*, 2nd ed; IRL Press: Oxford, 1995.
(3) Malhotra, R.; Sim, R. B. *Biochem. Soc. Trans.* **1993**, *22*, 106.
(4) Weis, W. I.; Drickamer, K.; Hendrickson, W. A. *Nature* **1992**, *360*, 127.
(5) Reid, K. B. M. *Biochem. Soc. Trans.* **1983**, *11*, 1.
(6) Kishore, U.; Reid, K. B. *Immunopharmacology* **2000**, *49*, 159.
(7) Tschopp, J.; Engel, A.; Podack, E. R. *J. Biol. Chem.* **1984**, *259*, 1922.
(8) Tschopp, J. *J. Biol. Chem.* **1984**, *259*, 7857.

(9) Rudd, P. M.; Morgan, B. P.; Wormald, M. R.; Harvey, D. J.; van den Berg, C. W.; Davis, S. J.; Ferguson, M. A.; Dwek, R. A. *Adv. Exp. Med. Biol.* **1998**, *435*, 153.
(10) Thielens, N. M.; Aude, C. A.; Lacroix, M. B.; Gagnon, J.; Arlaud, G. J. *J. Biol. Chem.* **1990**, *265*, 14469.
(11) Thielens, N. M.; Enrie, K.; Lacroix, M.; Jaquinod, M.; Hernandez, J. F.; Esser, A. F.; Arlaud, G. J. *J. Biol. Chem.* **1999**, *274*, 9149.
(12) Busby, T. F.; Ingham, K. C. *Biochemistry* **1990**, *29*, 4613.
(13) Dahl, M. R.; Thiel, S.; Matsushita, M.; Fujita, T.; Willis, A. C.; Christensen, T.; Vorup-Jensen, T.; Jensenius, J. C. *Immunity* **2001**, *15*, 127.
(14) Wong, N. K.; Kojima, M.; Dobo, J.; Ambrus, G.; Sim, R. B. *Mol. Immunol.* **1999**, *36*, 853.
(15) Vivanco, F.; Munoz, E.; Vidarte, L.; Pastor, C. *Mol. Immunol.* **1999**, *36*, 843.
(16) Hirani, S.; Lambris, J. D.; Muller-Eberhard, H. J. *Biochem. J.* **1986**, *233*, 613.
(17) Gaboriaud, C.; Rossi, V.; Bally, I.; Arlaud, G. J.; Fontecilla-Camps, J. C. *EMBO J.* **2000**, *19*, 1755.
(18) Mizuochi, T.; Yonemasu, K.; Yamashita, K.; Kobata, A. *J. Biol. Chem.* **1978**, *253*, 7404.
(19) Farries, T. C.; Lachmann, P. J.; Harrison, R. A. *Biochem. J.* **1988**, *252*, 47.
(20) Ault, B. H.; Schmidt, B. Z.; Fowler, N. L.; Kashtan, C. E.; Ahmed, A. E.; Vogt, B. A.; Colten, H. R. *J. Biol. Chem.* **1997**, *272*, 25168.
(21) Ullman, C. G.; Chamberlain, D.; Ansari, A.; Emery, V. C.; Haris, P. I.; Sim, R. B.; Perkins, S. J. *Mol. Immunol.* **1998**, *35*, 503.
(22) Wolff, M. W.; Zhang, F.; Roberg, J. J.; Caldwell, E. E.; Kaul, P. R.; Serrahn, J. N.; Murhammer, D. W.; Linhardt, R. J.; Weiler, J. M. *Protein Expression Purif.* **2001**, *22*, 414.
(23) Sharma, A. K.; Pangburn, M. K. *Gene* **1994**, *143*, 301.
(24) Schreck, S. F.; Plumb, M. E.; Platteborze, P. L.; Kaufman, K. M.; Michelotti, G. A.; Letson, C. S.; Sodetz, J. M. *J. Immunol.* **1998**, *161*, 311.
(25) Bann, J. G.; Bachinger, H. P. *J. Biol. Chem.* **2000**, *275*, 24466.
(26) Perkins, S. J. *Biochem. J.* **1985**, *228*, 13.
(27) Hase, S.; Kikuchi, N.; Ikenaka, T.; Inoue, K. *J. Biochem.* (*Tokyo*) **1985**, *98*, 863.
(28) Solis, D.; Feizi, T.; Yuen, C. T.; Lawson, A. M.; Harrison, R. A.; Loveless, R. W. *J. Biol. Chem.* **1994**, *269*, 11555.
(29) Hourcade, D. E.; Mitchell, L. M.; Oglesby, T. J. *J. Immunol.* **1999**, *162*, 2906.
(30) Cooney, K. A.; Ginsberg, D. *Blood* **1996**, *87*, 2322.
(31) Hinshelwood, J.; Spencer, D. I.; Edwards, Y. J.; Perkins, S. J. *J. Mol. Biol.* **1999**, *294*, 587.
(32) Soames, C. J.; Sim, R. B. *Biochem. J.* **1997**, *326*, 553.
(33) Higgins, J. M.; Wiedemann, H.; Timpl, R.; Reid, K. B. *J. Immunol.* **1995**, *155*, 5777.
(34) Coyne, K. E.; Hall, S. E.; Thompson, S.; Arce, M. A.; Kinoshita, T.; Fujita, T.; Anstee, D. J.; Rosse, W.; Lublin, D. M. *J. Immunol.* **1992**, *149*, 2906.
(35) Davies, A.; Simmons, D. L.; Hale, G.; Harrison, R. A.; Tighe, H.; Lachmann, P. J.; Waldmann, H. *J. Exp. Med.* **1989**, *170*, 637.
(36) Rudd, P. M.; Morgan, B. P.; Wormald, M. R.; Harvey, D. J.; van-den-Berg, C. W.; Davis, S. J.; Ferguson, M. A.; Dwek, R. A. *J. Biol. Chem.* **1997**, *272*, 7229.
(37) Bodian, D. L.; Jones, E. Y.; Harlos, K.; Stuart, D. I.; Davis, S. J. *Structure* **1994**, *2*, 755.
(38) Ninomiya, H.; Stewart, B. H.; Rollins, S. A.; Zhao, J.; Bothwell, A. L.; Sims, P. J. *J. Biol. Chem.* **1992**, *267*, 8404.
(39) Davis, S. J.; van der Merwe, P. A. *Science* **1996**, *273*, 1241.
(40) Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. *Biochemistry* **1994**, *33*, 17.
(41) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683.
(42) Niemann, M. A.; Bhown, A. S.; Miller, E. J. *Biochem. J.* **1991**, *274*, 73.
(43) Acosta, J.; Hettinga, J.; Fluckiger, R.; Krumrei, N.; Goldfine, A.; Angarita, L.; Halperin, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5450.
(44) Doucey, M. A.; Hess, D.; Cacan, R.; Hofsteenge, J. *Mol. Biol. Cell* **1998**, *9*, 291.
(45) Krieg, J.; Hartmann, S.; Vicentini, A.; Glasner, W.; Hess, D.; Hofsteenge, J. *Mol. Biol. Cell* **1998**, *9*, 301.
(46) Hartmann, S.; Hofsteenge, J. *J. Biol. Chem.* **2000**, *275*, 28569.
(47) Hofsteenge, J.; Blommers, M.; Hess, D.; Furmanek, A.; Miroshnichenko, O. *J. Biol. Chem.* **1999**, *274*, 32786.
(48) Mori, K.; Dwek, R. A.; Downing, A. K.; Opdenakker, G.; Rudd, P. M. *J. Biol. Chem.* **1995**, *270*, 3261.
(49) Drickamer, K. *Curr. Opin. Struct. Biol.* **1999**, *9*, 585.
(50) Guile, G. R.; Rudd, P. M.; Wing, D. R.; Prime, S. B.; Dwek, R. A. *Anal. Biochem.* **1996**, *240*, 210.
(51) Kuster, B.; Wheeler, S. F.; Hunter, A. P.; Dwek, R. A.; Harvey, D. J. *Anal. Biochem.* **1997**, *250*, 82.
(52) Aslam, M.; Perkins, S. J. *J. Mol. Biol.* **2001**, *309*, 1117.
(53) Brodbeck, W. G.; Liu, D.; Sperry, J.; Mold, C.; Medof, M. E. *J. Immunol.* **1996**, *156*, 2528.
(54) Kuttner Kondo, L.; Medof, M. E.; Brodbeck, W.; Shoham, M. *Protein Eng.* **1996**, *9*, 1143.

CR990294A