GLYCOCONJUGATE RESEARCH / Volume I
Proceedings of the Fourth International Symposium on Glycoconjugates

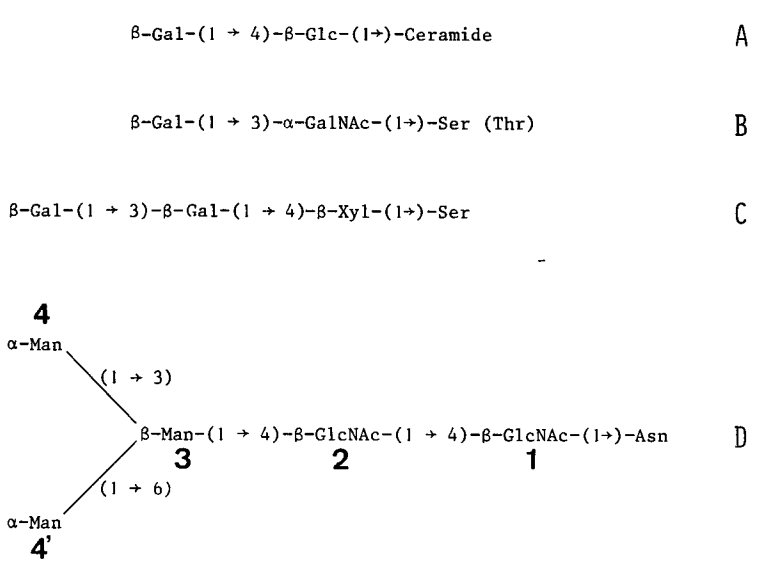
Primary Structure and Conformation of Glycans *N*-Glycosically Linked to Peptide Chains

Jean Montreuil and Johannes F. G. Vliegenthart

For a long time, our knowledge of the structure of glycoprotein glycans was limited to that of some linear glycans, such as the acidic mucopolysaccharides, and to that of some simple glycans like those of submaxillary mucins. Our ignorance was essentially due to the lack of precise and sensitive methods for structural investigation of complex glycans. But, in the past few years, important results have been accumulated because the far reaching biological importance of glycoconjugates, in general, and of glycoproteins, in particular, has become understood after the following discoveries: (a) Glycoconjugates are cell-surface antigens and their structure and function are modified in virus transformed cells and in cancerous cells. (b) They play an important role in intracellular adhesion and recognition, and in cell-contact inhibition. (c) They are receptor sites for viruses, proteins, and hormones. (d) Glycans protect the protein moiety against proteolytic attack. (e) The carbohydrate moiety may influence the conformation of the peptide chain. (f) Glycan groups permit the exit of proteins outside of the cell according to Eylar's hypothesis (1), which leads to the concept of Winterburn and Phelps (2), that glycoproteins are synthesized by cells for cells. (g) According to Ashwell (3), the sugar component regulates the catabolism of circulating proteins by different tissues, and the lifetime of proteins and cells. (h) A pathology of glycoproteins that is due to a lack of lysosomal glycosidases now exists and the term "glycoproteinosis" has been coined.

ISBN 0-12-301301-1

Thus, in the past few years, the complete primary structure of numerous glycans has been firmly established, and this rapid advance of our knowledge is due to the improvement of chemical, physical, and enzymic methods. The results obtained in the past four years entirely confirm, from a comparative biochemical point of view, the concepts that were developed in 1974 at the VIIth Symposium on Carbohydrate Chemistry held in Bratislava (4). In fact, glycan structures may be classified into families, within each of which glycan structures are very similar and present common oligosaccharide structures, even if they originate from animals, plants, microorganisms, or viruses. Consequently, a series of classes may be established: Glycans are conjugated to the peptide chains through two types of primary covalent linkages: (a) *O*-Glycosyl linkages (i) between L-serine or L-threonine and *N*-acetyl-D-galactosamine, D-mannose, or D-xylose, and (ii) between L-hydroxylysine or L-hydroxyproline and D-galactose. (b) *N*-Glycosyl linkage between L-asparagine and *N*-acetyl-D-glucosamine, which is until now the only one to have been characterized. Two types of glycans are present: (a) Linear glycans such as the acidic mucopolysaccharides; these structures are relatively simple because they result from the polymerization of a disaccharide unit; they have been known for several years. (b) Branched glycans, which present a more complex structure; some show a single branch, such as the human α_1-acid glycoprotein; others show a highly branched structure, such as the ovomucoid from hen egg white, which has 7 branching points (5).



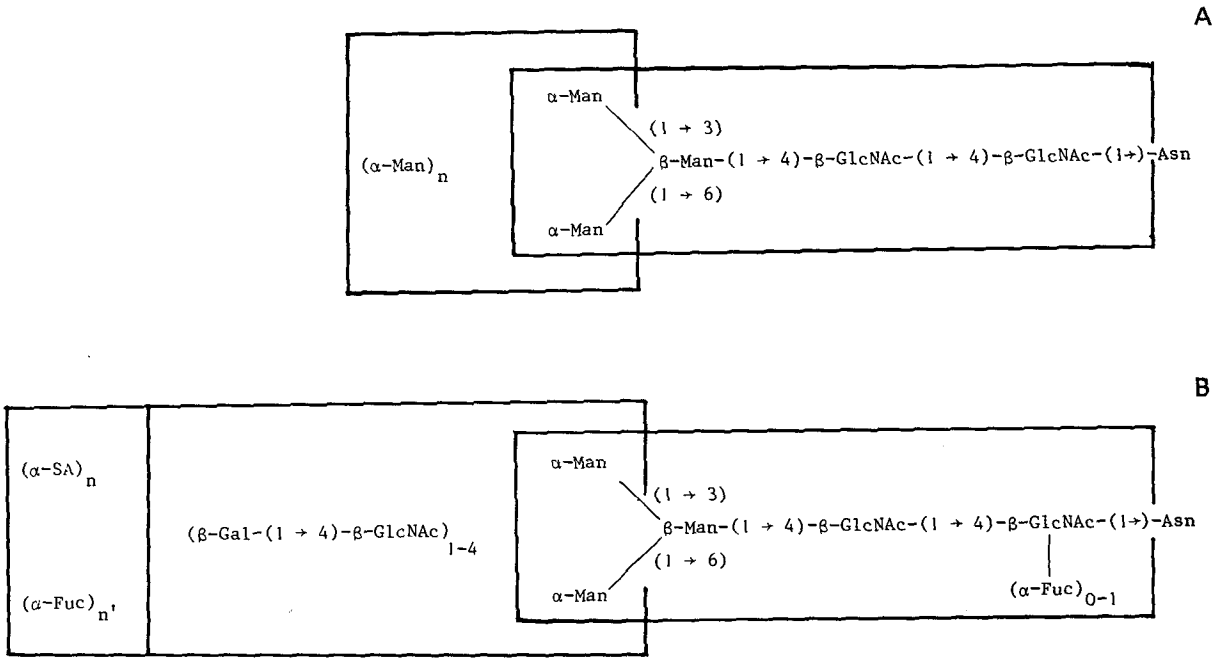
Scheme 1. Oligosaccharide cores of glycoconjugates.

The knowledge of the structure of glycoconjugates, in general, and of glycoproteins, in particular, leads to the establishment of the core concept: All glycan structures derive from substitutions at a common and nonspecific oligosaccharide core linked to the noncarbohydrate component (see Scheme 1): core A is that of glycolipids; core B is that found in numerous glycoproteins, such as MN and blood-group antigens, kappa caseins, antarctic fish antifreezing glycoprotein, IgA immunoglobulins from human serum and milk, and glycophorin; core C is that found at the reducing end of acidic mucopolysaccharides; and core D is that present in almost all *N*-glucosically conjugated glycans, which results from a mannotriose residue β-linked to a di-*N*-acetylchitobiose residue, linked itself to an asparagine residue.

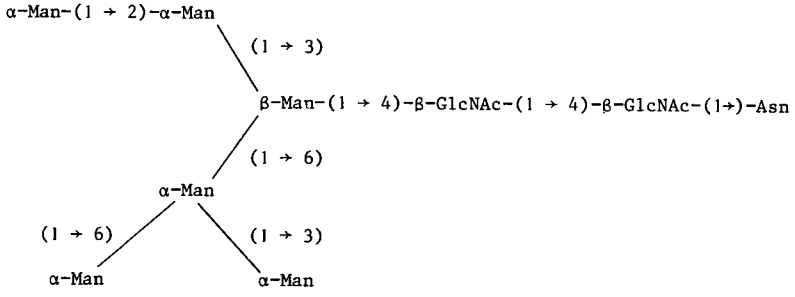
In *N*-glycosically conjugated glycans, the existence of two types of structures has been established. In the first one (Scheme 2A), the pentasaccharide core is substituted only by D-mannose residues, the term "oligomannosidic type" was proposed (4). The glycans illustrated in Schemes 3-5 belong to this type. In the second type (see Scheme 2B), the same pentasaccharide core is substituted by a variable number of *N*-acetyllactosamine, and sialic acid, or L-fucose (or both) residues. These structures belong to the "*N*-acetyllactosaminic type" the substituting and substituted *N*-acetyllactosamine residues were called "antennae" (4).

Until now, the types of substitution of the pentasaccharide core (Scheme 1) that have been demonstrated are as follows: (a) "biantennary structures" C-2 of mannose-4 and -4' residues (see Schemes 6 and 7A); (b) "triantennary structures" C-2 and C-4 of mannose-4, and C-2 of mannose-4' residue; or C-2 of mannose 4, and C-2 and C-6 of mannose-4' residue (see Schemes 7B, 8, and 9); (c) "tetraantennary structures" C-2 and C-4 of mannose-4, and C-2 and C-6 of mannose-4' residue (see Scheme 7C); (d) C-4 of mannose-3 residue substituted by an *N*-acetylglucosamine residue (Schemes 6 and 10); (e) C-6 of the first *N*-acetylglucosamine of the di-*N*-acetylchitobiose residue substituted by a fucose residue (Schemes 6, 8, 9, and 10); (f) C-3 of the *N*-acetylglucosamine residues of the *N*-acetyllactosamine residues substituted by a fucose residue (Schemes 6 and 7).

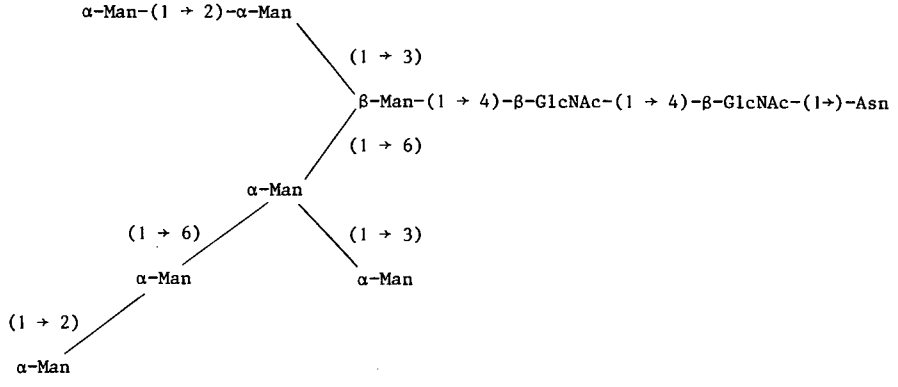
However, the concept of the existence of common cores and of similar structures in glycans must not be accepted as a dogma, because dogmas are dangerous and because some "non-orthodox" structures have been described. They generally differ from "orthodox" structures by the existence of only one *N*-acetylglucosamine residue at the "reducing terminus" and of "iso-*N*-acetylactosamine"-like structures (see Scheme 11). On the other hand, oligosialosyl sequences have been characterized, such as α-NeuAc-(2→8)-α-NeuAc-(2→6)-β-Gal in glycoproteins of various tissues (14).



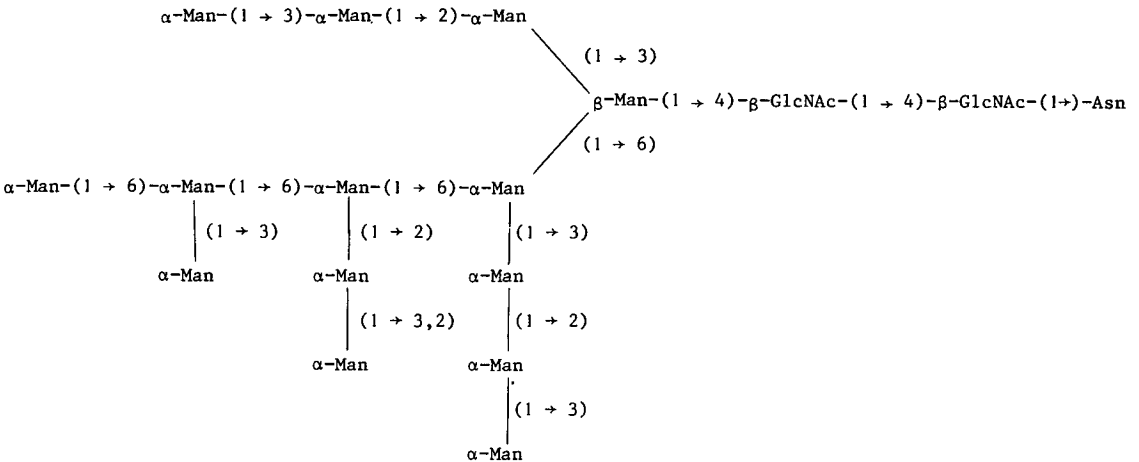
Scheme 2. General structural scheme of glycans of oligomannoside type (A) and of N-acetyllactosamine type (B).



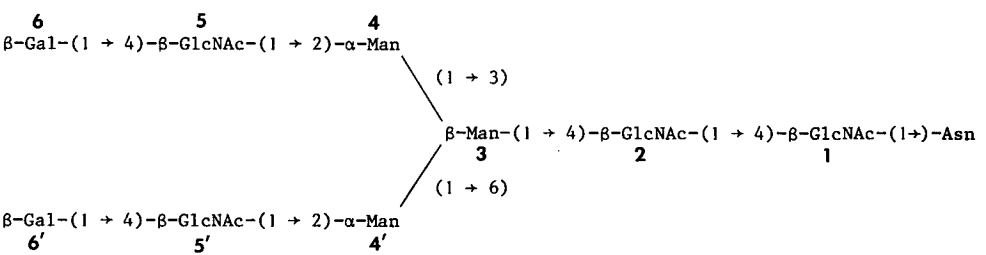
Scheme 3. Structure of ovalbumin (6) and of Taka-amylase A (7) glycans.



Scheme 4. Structure of Sindbis virus S-4 glycan.



Scheme 5. Core of S. cerevisiae *X-2180-1A wild type mannan (8).*



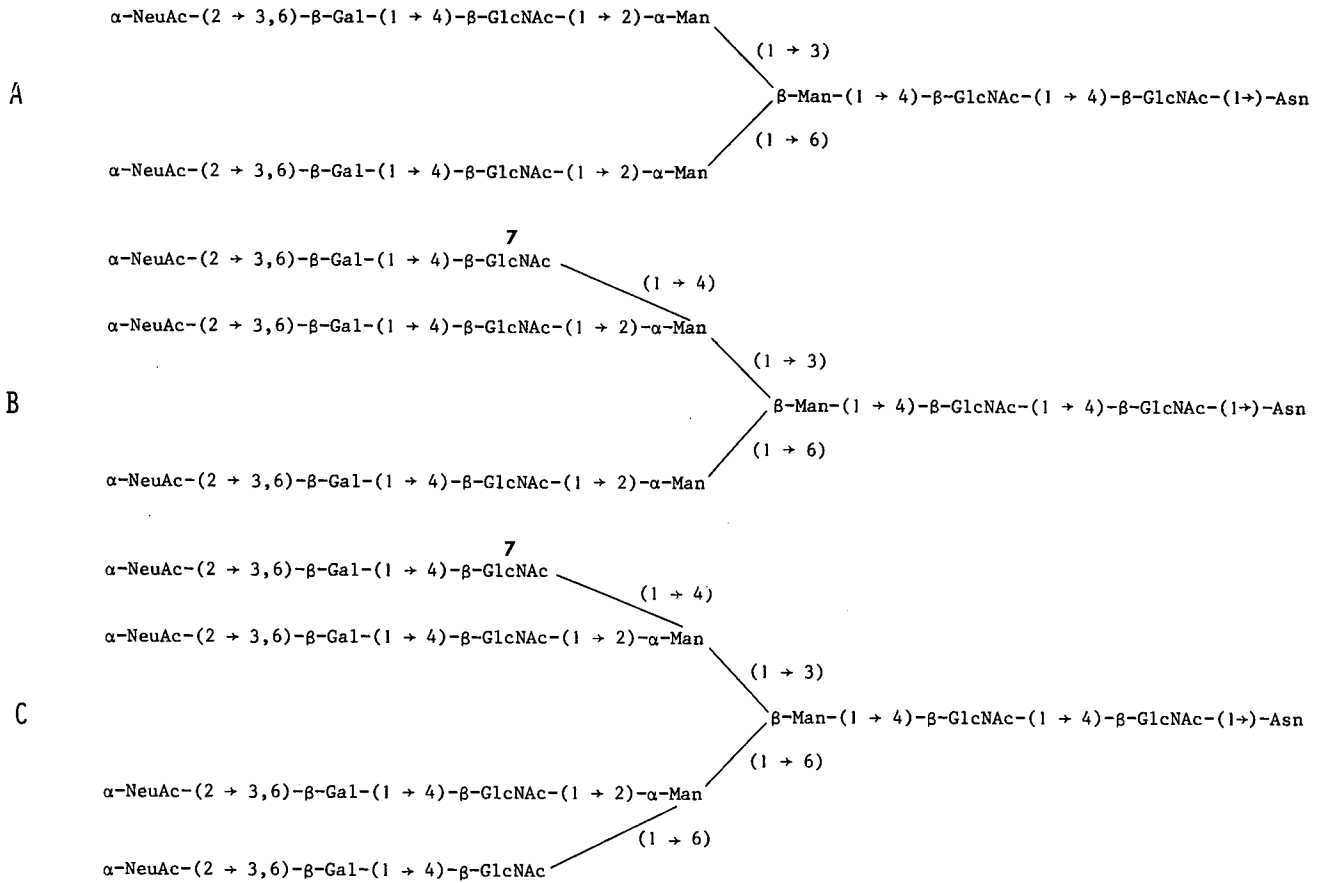
Scheme 6. Structure of biantennary glycans.

Most of the structures just described have been slowly and patiently determined by combining the use of chemical (partial acetolysis and hydrolysis, hydrazinolysis-nitrous deamination, chromic and periodate oxidation, and methylation) with enzymatic (hydrolysis by *exo*- and *endo*-glycosidases) well-defined methods. Recently, our laboratories have entered into a collaboration for exploring the conformation of human serotransferrin glycan (15) by high-resolution n.m.r. spectroscopy, by use of reference spectra obtained by enzymatic degradation of the transferrin glycan and from a large collection of oligosaccharides prepared by partial hydrolysis and acetolysis of ovomucoid (16), or isolated from urine of patients with lysosomal diseases (see later, refs. 17-26, and Strecker *et al.*, in this volume). No information concerning the conformation of the glycan molecule was obtained, as we expected, but the unexpected and exciting results obtained concerning the primary structure of glycans allow us to propose a method for determining the complete primary sequence of monosaccharides of the *N*-acetyllactosamine-type glycans on the basis of only methylation and 360-MHz ^{1}H-n.m.r. spectroscopy.

360-MHZ ^{1}H-N.M.R. SPECTROSCOPY OF CARBOHYDRATE CHAINS OF GLYCOPROTEINS

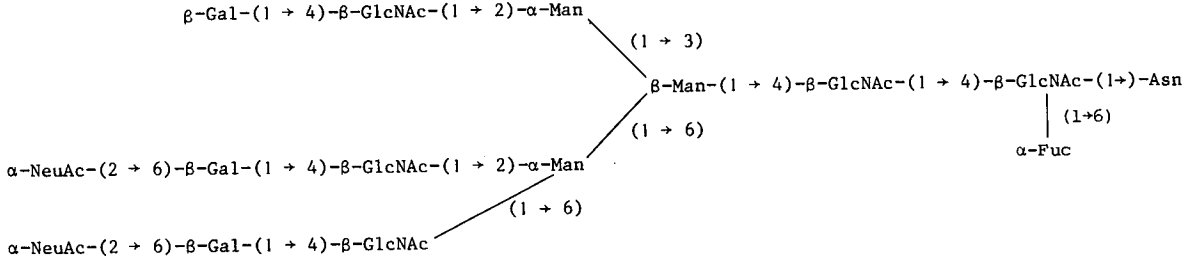
In recent years, high-resolution n.m.r. spectroscopy has become an extremely valuable technique in the study of biopolymers. In particular, for the investigation of structures, conformations, and intermolecular interactions in the fields of protein and nucleic acid chemistry, a vast amount of significant results has been described (27-31). However, the application of this technique to glycoconjugates (glycolipids and glycoproteins) is still rather limited. Several high-resolution ^{1}H-n.m.r. data are available for derivatives of mono-, oligo-, and polysaccharides; especially peracetyl (32-34), pertrimethylsilyl (35), and permethyl derivatives (36) have been studied. These spectra give valuable information on the configuration of the glycoside linkages, and on the type and conformation of the (constituting) monosaccharides.

A smaller number of papers deal with high-resolution ^{1}H-n.m.r. spectroscopy of non-derivatized carbohydrates in solution of deuterium oxide. For applications in the biochemical field, the latter system has several advantages, namely: (a) A chemical modification of the compounds can be omitted, (b) deuterium oxide is a good solvent for a wide range of carbohydrates and glycoconjugates, and (c) intramolecular hydrogen bonds are preserved, thus providing more insight into the complete structure of the compound in aqueous solution. To understand the biochemical role of the carbohydrate part of glycoconjugates, a detailed knowledge of the primary and

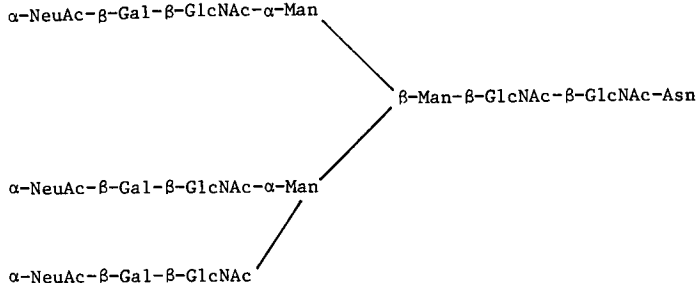


Scheme 7. *(See following page.)*

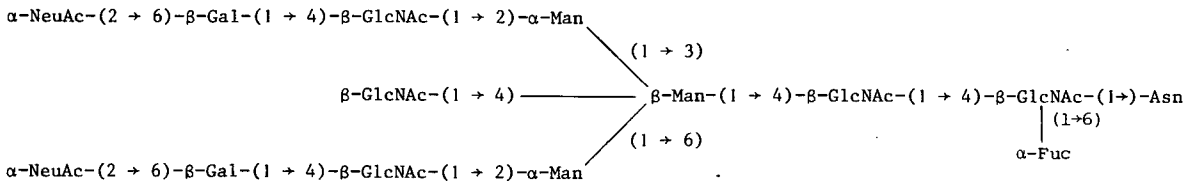
Scheme 7. Structure of bi-, tri-, and tetraantennary glycans of α_1-acid glycoprotein (B. Fournet et al., in this volume). GP-II-6: structure A; GP-II-5, GP-II-7, and GP-V-5: structure B; GP-III-6 and GP-V-4: structure C. In GP-III-5, GP-V-2, and GP-V-3 glycopeptides an additional α-Fuc-(1→3) residue is linked to the GlcNAc residue (7) of structure C.



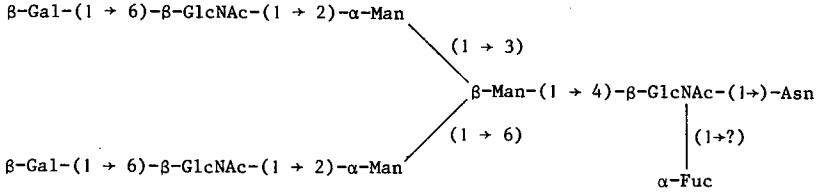
Scheme 8. Structure of unit-B type glycopeptide Gp-3 of porcine thyroglobulin (9).



Scheme 9. Structure of type A glycopeptide of Semliki Forest virus (10).



Scheme 10. Structure of human serum IgG glycopeptide (11,12).



Scheme 11. Structure of human myeloma IgM glycan (13).

preferably also of the three-dimensional structure, in aqueous solution, of these moieties is indispensable. For this purpose, n.m.r. spectroscopy is in principle one of the most powerful methods.

This report deals with the 360-MHz ^{1}H-n.m.r. spectroscopy of complex carbohydrate chains, coupled *via* an *N*-glycosyl linkage to an asparagine residue in glycoproteins. For reference purposes, the spectral data of some partial (oligosaccharide and glycopeptide) structures have been included.

The 360-MHz ^{1}H-n.m.r. spectrum of the general structural element β-GlcNAc-(1→4)-Asn is shown in Fig. 1. The resonances of the anomeric proton (H-1), the CH_2-3 residue group of the asparagine residue, and the *N*-acetyl protons of the GlcNAc residue are found clearly separated from the bulk of the non-anomeric carbohydrate protons (3.5-4.0 p.p.m.). The low-field position of the (axially oriented) anomeric proton (5.09 p.p.m.) and the rather large coupling constant $J_{1,2}$ (9.8 Hz) are due to the linkage of the amide nitrogen atom to the anomeric carbon atom (37). The two methylene-3 protons of the asparagine residue are not equivalent; their chemical shifts (2.93-2.87 p.p.m.) and the chemical shift of the methine-2 proton (3.99 p.p.m.) are in accordance with those found for the zwitterionic form of free asparagine (28,30,31). The bulk of the non-anomeric proton signals at 3.5-4.0 p.p.m. could be completely assigned, as indicated in Fig. 2, by use of specific proton-decoupling and spectrum simulation. The $^4C_1$ (D) chair conformation of the pyranosyl ring of the GlcNAc residue could be deduced from the proton coupling-constants by use of an adapted Karplus equation (38). Another structural element frequently occurring in glycoproteins is α-Fuc-(1→6)-β-GlcNAc-(1→4)-Asn, the spectrum of which is given in Fig. 3 (37). Characteristic for the fucose residue are the resonances of the anomeric proton (4.90 p.p.m., $J_{1,2}$ 3.75 Hz) and the methyl group (1.21 p.p.m.). The coupling constant of 3.75 Hz is indicative of an α-L glycosidic bond of the fucose residue. The spectrum could be completely interpreted and the ring conformation of the sugar residues determined. The attachment of fucose to C-6 of a GlcNAc residue gives rise to changes in the chemical shifts for H-4,-5, and -6 of the GlcNAc residue as compared to the shifts of β-GlcNAc-(1→4)-Asn. Also a change in the geminal coupling constant of H-6 and -6' of the GlcNAc residue (12.7 to 11.4 Hz) occurs.

The (1→6) linkage of the fucose residue was unambiguously proven by ^{13}C-n.m.r. spectroscopy (39). Only two carbon atoms of the GlcNAc residue of the glycopeptide Fuc→GlcNAc→Asn show significant shifts with respect to the resonances of the GlcNAc→Asn residue, namely a downfield shift of 6.7 p.p.m. for C-6 and an upfield shift of 0.8 p.p.m. for C-5. These shifts point directly to glycosylation of the OH-6 of the GlcNAc residue (40,41).

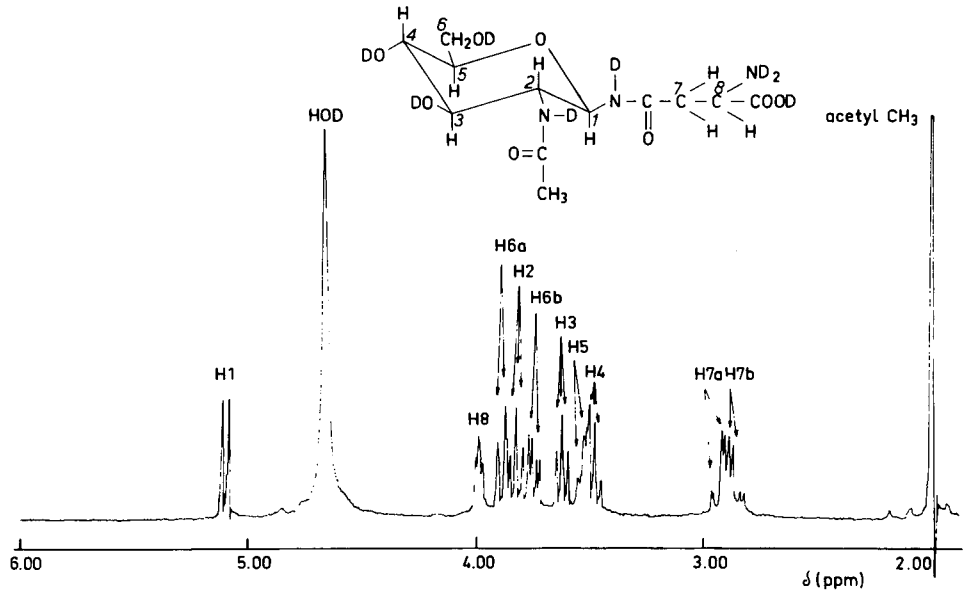


Fig 1. 360-MHz ^{1}H-n.m.r. spectrum of β-GlcNAc-(1→4)-Asn in D_2O.

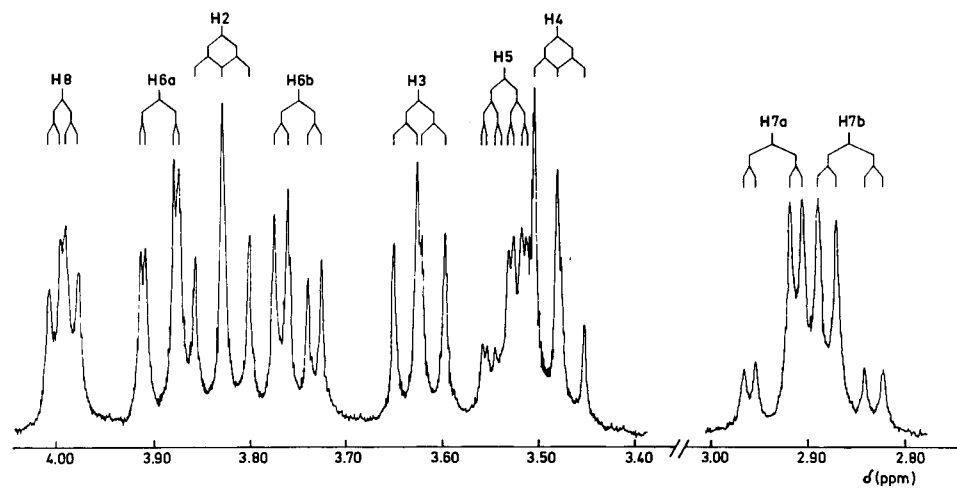


Fig. 2. 360-MHz ^{1}H-n.m.r. spectrum of the non-anomeric protons and β-methylene protons of β-GlcNAc-(1→4)-Asn.

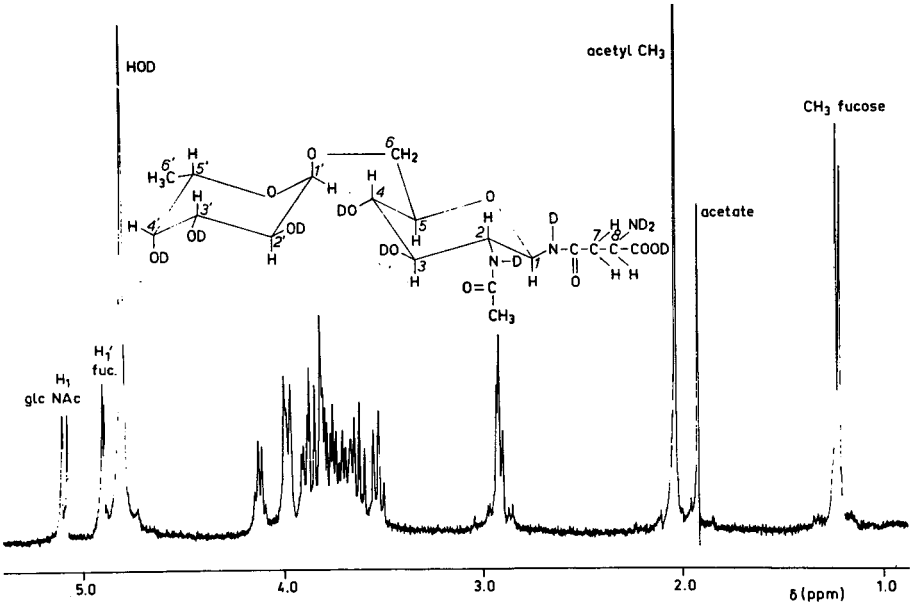


Fig. 3. 360-MHz ^{1}H-n.m.r. spectrum of α-Fuc-(1→6)-β-GlcNAc-(1→4)-Asn in D_2O.

An example of a complex Asn-glycopeptide is the asialoglycopeptide obtained from human serotransferrin. The 360-MHz ^{1}H-n.m.r. spectrum and the structure are shown in Fig. 4 (42). To illustrate the high resolution in this spectrum, the 90-MHz ^{1}H-n.m.r. spectrum is presented in Fig. 5. Comparison of both spectra makes clear that the low-resolution instrument is unsuitable for unraveling structural details of complex carbohydrate chains. The resonances shown in Fig. 4 can roughly be divided into signals from the following groups of protons: (a) Anomeric protons; at ambient temperature one anomeric-proton resonance is hidden under the HOD line. It can be visualized by recording the spectra at other temperatures. (b) H-2 protons of mannose residues. (c) Bulk of the non-anomeric protons. (d) The methylene-3 group of the asparagine residue. (e) The methyl group protons of the acetamido group. So far, it is impossible to interpret the bulk of the non-anomeric protons. Many structural details of this compound are, however, reflected in the chemical shifts of the anomeric protons and the H-2 of the mannose residues.

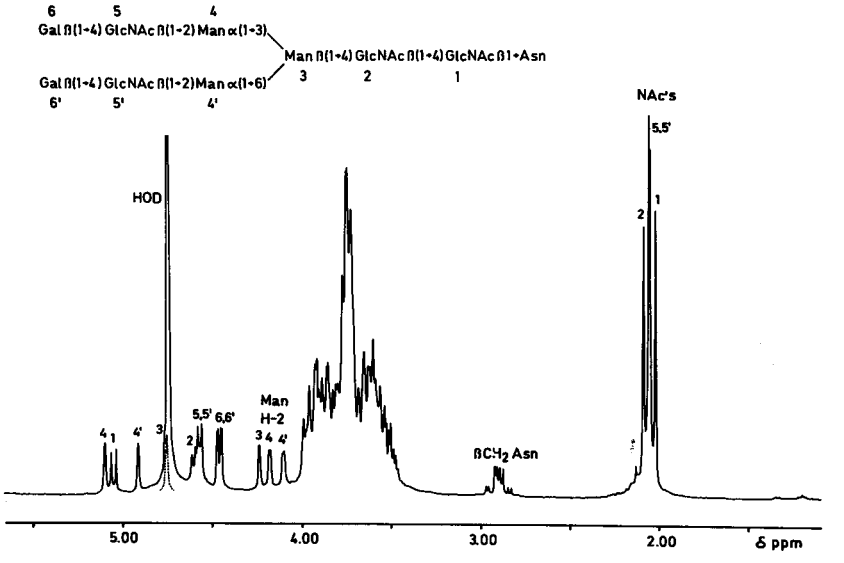


Fig. 4. 360-MHz ^{1}H-n.m.r. spectrum of the asialo-glycan-Asn isolated from human serotransferrin.

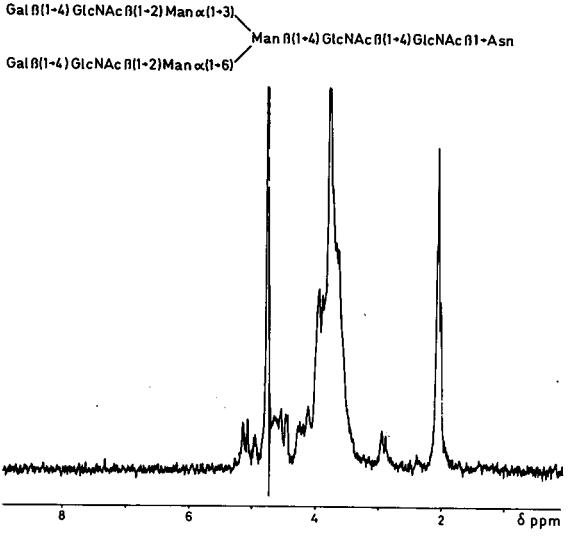


Fig. 5. 90-MHz ^{1}H-n.m.r. spectrum of the asialo-glycan-Asn isolated from human serotransferrin.

Integration showed that nine anomeric protons are present, which is in accordance with the proposed number of monosaccharide units. In fact, integration of peak areas is an accurate method for the determination of the number of constituting monosaccharides. An additional control for the number of amino sugar residues may be obtained from the integration of the NAc signals. For assignment of the anomeric protons, the spectra of a large series of partial structures has been recorded. A few relevant representatives of this group of compounds are given in Fig. 6 together with a schematic presentation of some spectral data.

The anomeric signal of the GlcNAc residue (*1*) can easily be recognized; among others, the large coupling constant is typical. Comparison of the structures *B*, *E*, and *F* yields the anomeric signal of the GlcNAc residue (*2*) ($J_{1,2}$ *ca.* 7 Hz, H-1β). The anomeric signals of the mannose residues follow from correlation of the spectra of *B* and *C*, taking into account that an H-1 bound to a C in α-D-glycosidic linkage resonates at a field lower than that of an H-1 bound to a β-D-linked C. Comparison of the spectra of *D*, *E*, and *F* makes clear that the anomeric protons of the GlcNAc residues (*5*) and (*5'*) resonate both at *ca.* 4.57 p.p.m. The anomeric protons of the residue (*6*) and (*6'*) are also indistinguishable and resonate both at *ca.* 4.47 p.p.m., as is evident from comparison of the spectra of *E* and *F*. By selective irradiation, the signals of the H-2 of the mannose residues could be assigned as indicated. The *N*-acetyl resonances could be interpreted on the basis of a set of reference compounds. The expanded region of the interpretable parts of the spectrum of the asialoglycan→Asn glycopeptide of human serotransferrin is depicted in Fig. 7.

A few conclusions can be drawn: (a) The chemical shifts of the various anomeric protons in the intact glycopeptide and the partial structures thereof occur at characteristic positions. (b) The primary structure and the type of glycosidic linkages are reflected by the chemical shifts and the coupling constants of the anomeric protons of the various monomers. (c) The total n.m.r. spectrum can be used as a fingerprint, *e.g.*, on this basis the occurrence of the "asialoglycan→Asn" part as a structural element in rabbit serotransferrin, human lactotransferrin, orosomucoid, IgA, and IgG could be demonstrated. (d) The mannotriose branching-core, surrounded by GlcNAc residues, can be recognized on the basis of the pattern of the mannose H-2 resonances.

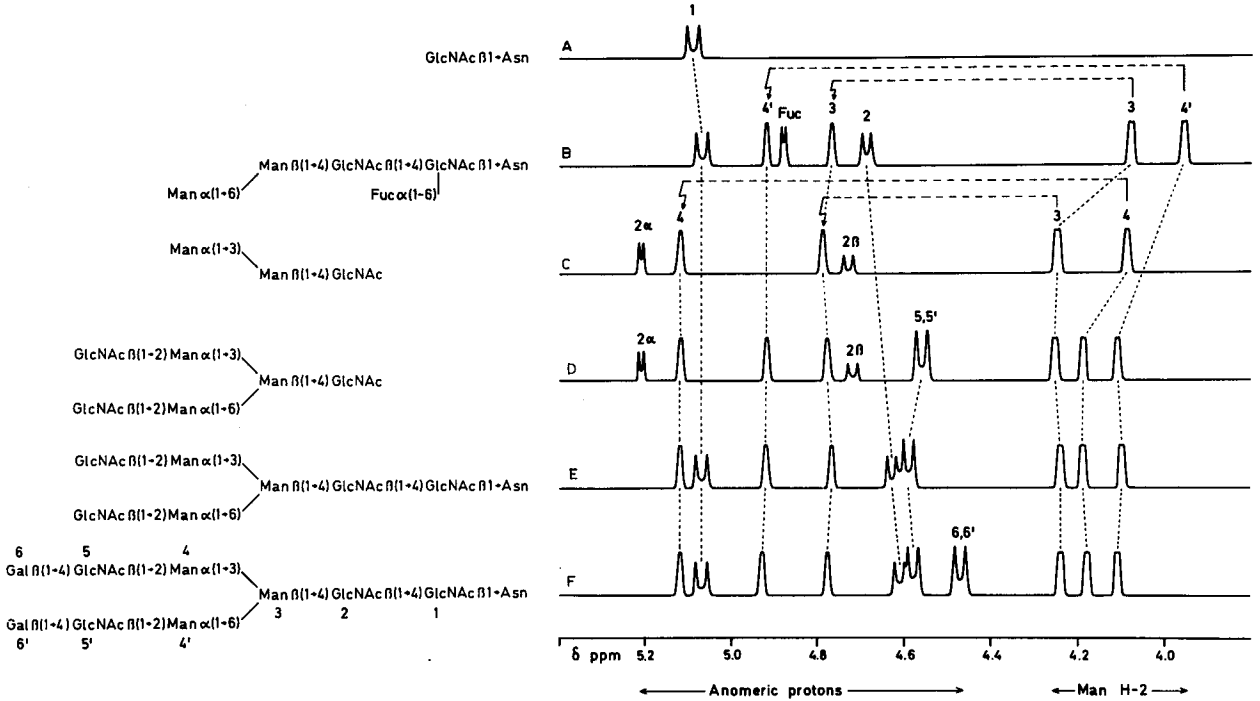


Fig. 6. Schematic representation of 360-MHz ^{1}H-n.m.r. data of the asialo-glycan-Asn and some reference compounds.

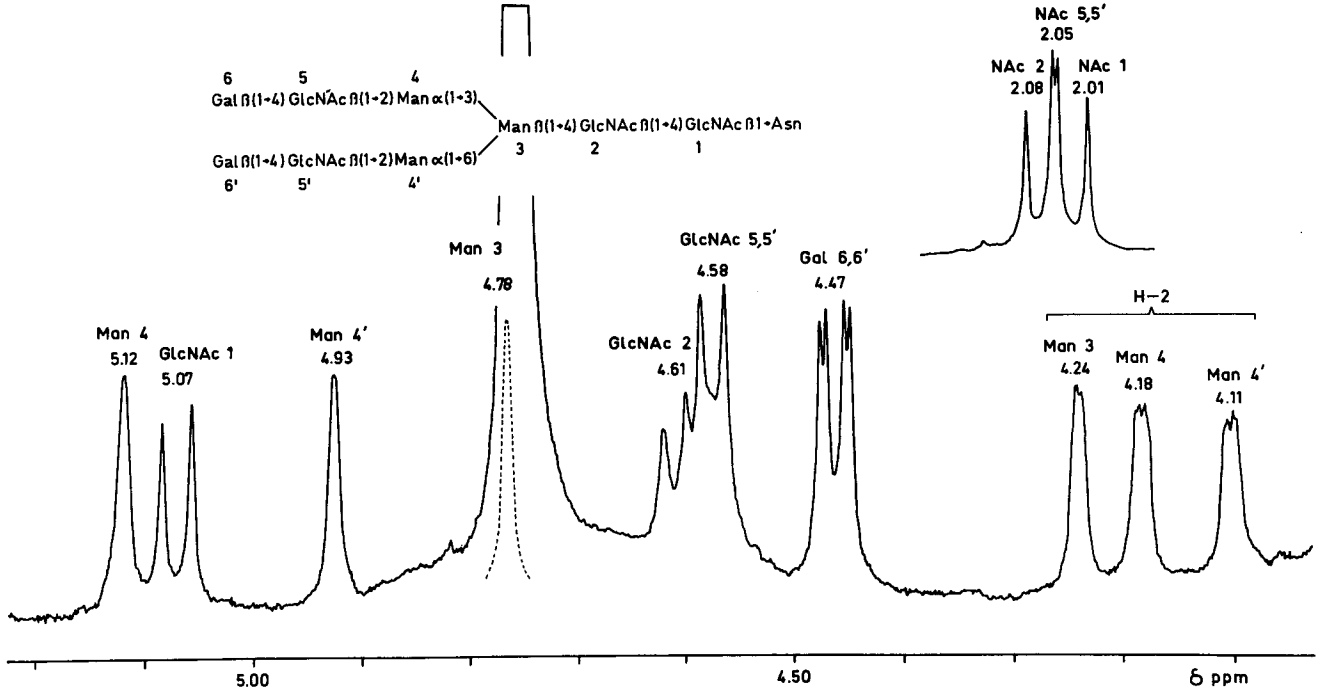


Fig. 7. Expanded regions of the anomeric protons, the mannose H-2 protons, and the N-acetyl protons of the 360-MHz ^{1}H-n.m.r. spectrum of the asialo-glycan-Asn.

The occurrence of sialic acid residues in terminal position of the biantennary structure has a remarkable effect on the spectrum, depending on the type of glycosidic linkage. First, the (2→6)-linked sialic acid residue occurring in the glycan of human serotransferrin will be considered. In Fig. 8, the 360-MHz ^{1}H-n.m.r. spectrum of this glycan Asn-Lys chain is presented. The significant additional signals in this spectrum are those of the H-3*eq* and H-3*ax* protons of sialic acid. These are typical resonances for sialic acid, which do not coincide with the bulk of the non-anomeric protons. Their chemical shifts may be used to assign the position of the glycosidic linkage as will be illustrated later. The introduction of a sialic acid residue at C-6 of the galactose residues gives rise to a few small, but characteristic shift increments for some anomeric protons, as indicated in Fig. 9. Downfield shifts are observed for the anomeric protons of the Man (*4*) and (*4'*) residues as well as of the GlcNAc (*5*) and (*5'*) residues, whereas the anomeric protons of the Gal (*6*) and (*6'*) residues show upfield shifts. These shift increments may be used to determine the position of sialic acid in a biantennary structure having only one sialic acid residue.

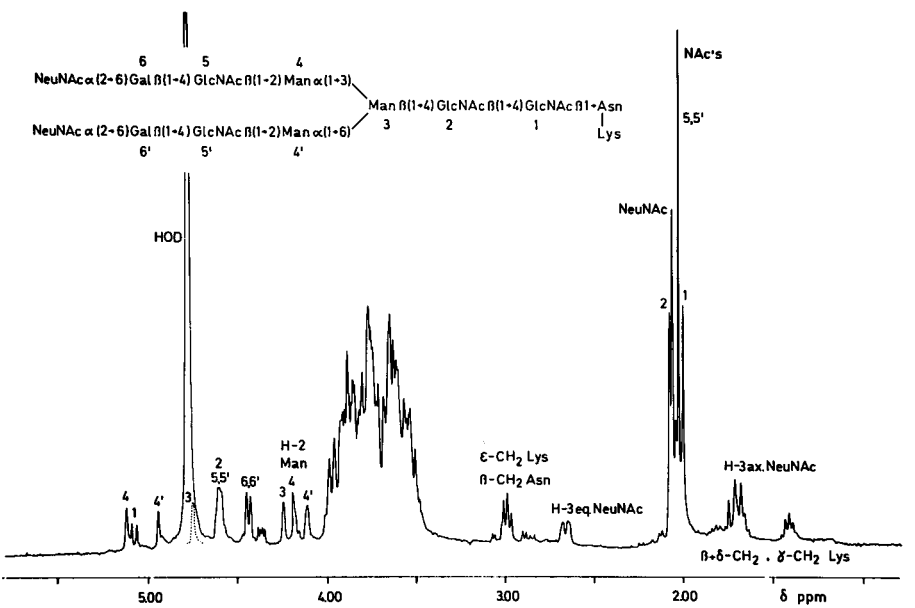


Fig. 8. 360-MHz ^{1}H-n.m.r. spectrum of the glycan-Asn-Lys isolated from human serotransferrin.

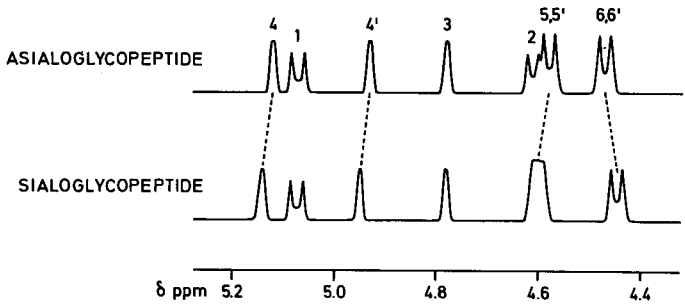
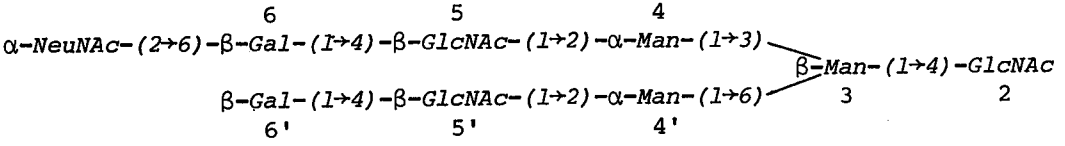


Fig. 9. Comparison of the anomeric regions of the asialo glycopeptide and the sialo glycopeptide.

In Table I, the 360-MHz ^{1}H-n.m.r. spectral data of asialoglycan→Asn and the oligosaccharide V, which has the sialic acid residue attached to the upper branch, are compared. The shift increments of the anomeric protons of neither the Gal (*6*) and (*6'*) residue, nor of the GlcNAc (*5*) and (*5'*) residues provide structural information, since it is not known which signals stem from the upper or the lower branch. However, it is easy to distinguish the resonances of the anomeric protons of the Man (*4*) and (*4'*) residues as only the anomeric proton of the Man (*4*) residue undergoes a shift when a sialic acid residue is linked to the upper branch (see Table I). This resonance is therefore indicative of the position of the sialic acid residue. On the basis of these data and those of a set of reference compounds, the following conclusions may be drawn: (a) The attachment of a sialic acid residue to one branch in a biantennary structure has no effect on the chemical shifts of the anomeric protons in the other branch. (b) The (2→6)-linked sialic acid residue has a long distance effect on the mannose residue that occurs in the same branch. (c) The chemical shifts of the mannose (*4*) and (*4'*) anomeric protons may be used to establish the position of a (2→6)-linked sialic acid residue. Sialic acid residues may also occur in a (2→3)-glycosidic linkage to D-galactose residues. Such a sialic acid residue may easily be recognized from the chemical shifts of the H-3*eq* and -*ax* protons, as indicated in Table II. The only anomeric proton that undergoes a significant shift upon attachment of a sialic acid residue to C-3 of the galactose residue is that of the galactose residue itself. In contrast to the (2→6)-linked sialic acid residue, it is not possible to distinguish whether the (2→3)-linked sialic acid residue is present in the upper or the lower branch because the anomeric protons of Gal (*6*) and (*6'*) residues have the same chemical shift, and the chemical shifts of the mannose (*4*) and (*4'*) anomeric protons are unaffected.

Table I. Changes in the Chemical Shift of Anomeric Protons Due to the Presence of Sialic Acid in the Upper Branch of the Biantennary Structure

Oligosaccharide V



Structures	*Chemical shifts (δ) of H-1 of residues*					
	4	5	6	4'	5'	6'
Asialoglycan→Asn	*5.119*	*4.581*	*4.470*	*4.926*	*4.581*	*4.470*
Oligosaccharide V	*5.131*	*4.603*	*4.447*	*4.929*	*4.583*	*4.468*

Table II. Comparison of the Chemical Shifts of H-3eq. and H-3ax. of Sialic Acid Residue (2→3)- or (2→6)-linked to the Galactose Residue of N-Acetyl-lactosamine

Structures	*Chemical shifts (δ)*	
	H-3eq	*H-3ax*
α-NeuNAc-(2→6)-Gal	*2.67*	*1.72*
α-NeuNAc-(2→3)-Gal	*2.76*	*1.80*

In a biantennary structure having one (2→3)-linked and one (2→6)-linked sialic acid residue, the occurrence and position of both types may easily be deduced from the 360-MHz ^{1}H-n.m.r. spectrum. As shown in Fig. 10, two sets of H-3 protons characteristic for the two types of sialic acid residues are present. The signal of the anomeric proton of the mannose (*4*) residue is shifted from 5.12 to 5.14 p.p.m., thus indicating the presence of a (2→6)-linked sialic acid residue in the upper branch. Consequently, the (2→3)-linked sialic acid residue is attached to the lower branch. In more complex structures, for example in a triantennary structure, it is often possible to draw useful conclusions on the position of sialic acid residues. As shown in Fig. 11, the 360-MHz ^{1}H-n.m.r. spectrum of a triantennary structure having 3 sialic acid residues, the integration of the H-3-*eq* and -*ax* signals of the sialic acid residue indicates that the ratio of (2→3)- to (2→6)-linked sialic acid residues is 1:2. The resonance position of the anomeric protons of the mannose (*4*) and (*4'*) residues indicate the occurrence of a (2→6)-linked sialic acid residue in the corresponding branches. Therefore, it must be concluded that the (2→3)-linked sialic acid residue is located in the additional branch. Characteristic for the triantennary structure are the chemical shifts of the H-2 protons of the mannose residues, as will be discussed later.

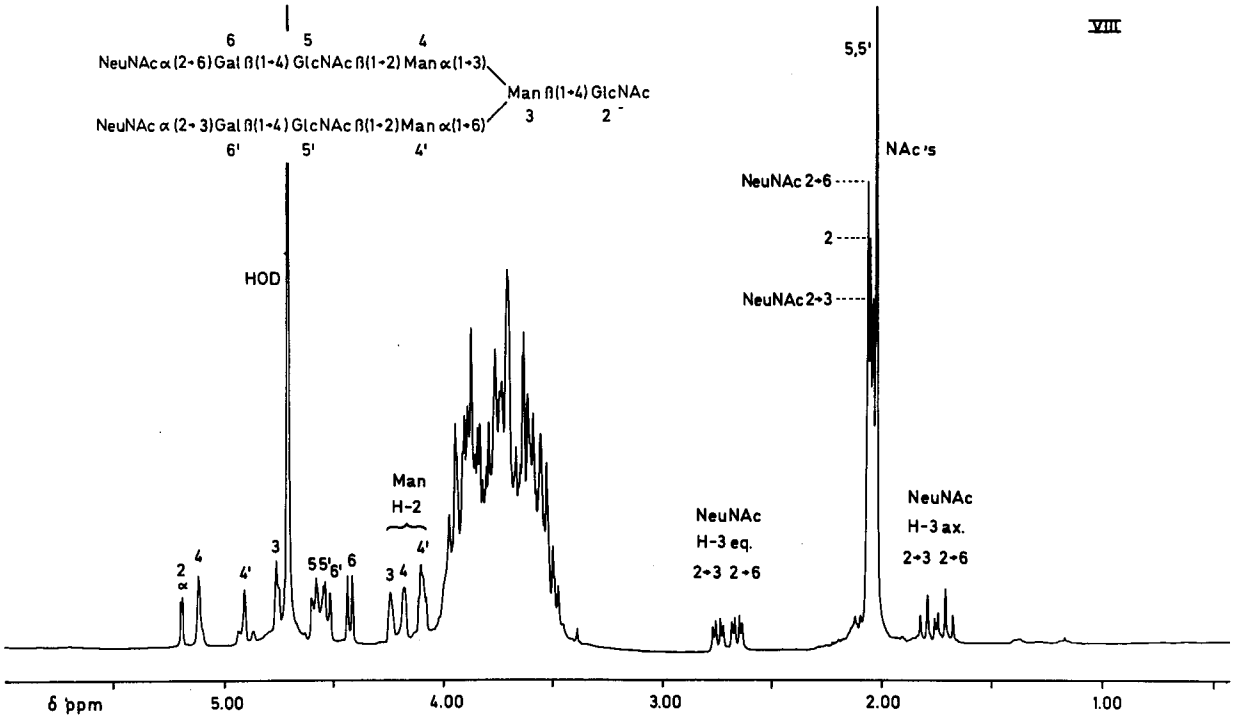


Fig. 10. 360-MHz ^{1}H-n.m.r. spectrum of oligosaccharide VIII, isolated from urine of a sialidosis patient, having two different linkages of the sialic acid residues.

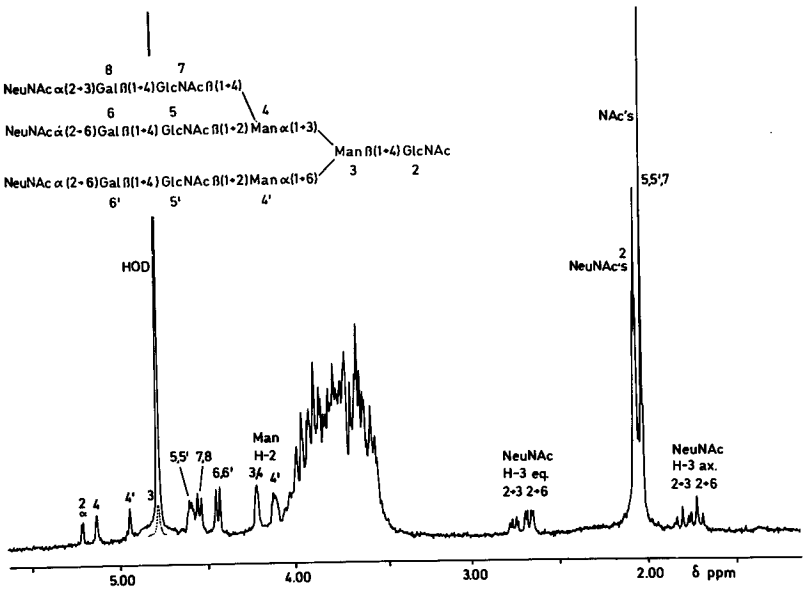


Fig. 11. 360-MHz ^{1}H-n.m.r. spectrum of an urinary oligosaccharide having a triaantennary structure.

Tetraantennary structures occur in α_1-acid glycoprotein. The 360-MHz ^{1}H-n.m.r. spectrum of such a structure in the asialo form is shown in Fig. 12. The quantitative composition of this glycopeptide can easily be derived from integration of the peak areas: The 5 *N*-acetyl signals account for 6 *N*-acetyl groups. The methyl signals (2 doublets) indicate the presence of 1 fucose residue, besides 1 threonine residue. Furthermore, 3 mannose and 4 galactose residues are present. Characteristic for this structure is the extension of the mannotrioside core with two additional branches. This is reflected, in the ^{1}H-n.m.r. spectrum, in the chemical shifts of the H-1 and H-2 protons of the mannose residues. These chemical shifts are summarized in Table III for the bi-, tri-, and tetra-antennary structures. On the basis of these data, it is possible to deduce the substitution pattern of the mannotrioside core, provided that only the structures reported here are taken into consideration. It is to be expected that other structures give rise to other spectral data.

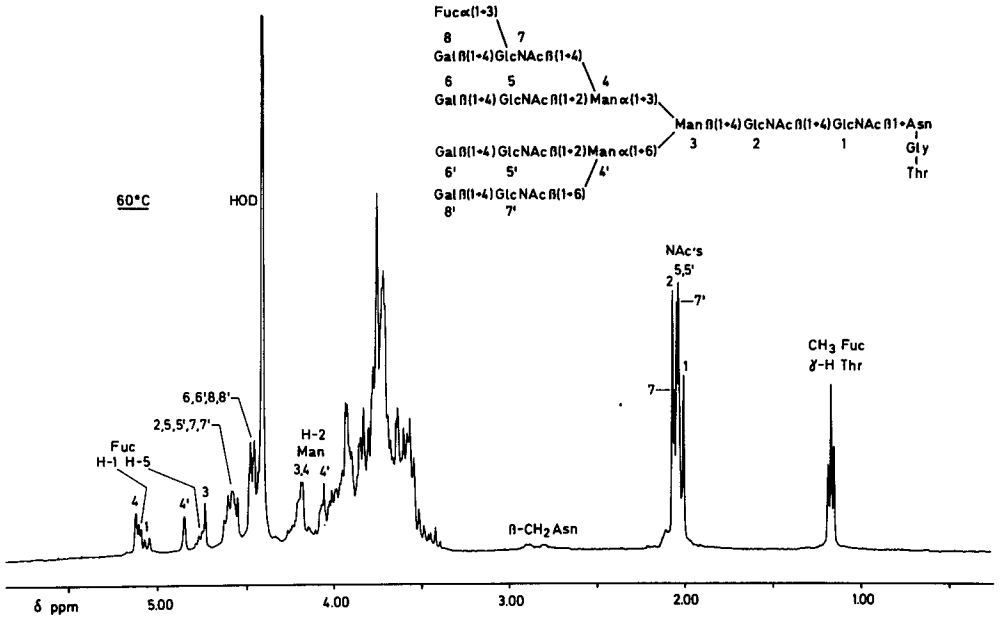
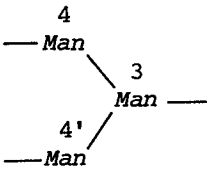
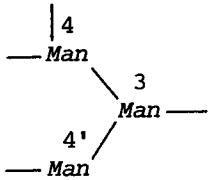
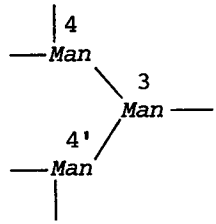


Fig. 12. 360-MHz ^{1}H-n.m.r. spectrum of a glycopeptide, isolated from α_1-acid glycoprotein, having a tetraantennary structure.

Table III. Chemical Shift Data of the Mannose H-1 and H-2 Protons in Bi-, Tri-, and Tetra-antennary Structures

Structures	Chemical shift (δ) of H-1 Mannose residue			H-2 Mannose residue		
	3	4	4'	3	4	4'
	4.77	5.12	4.93	4.24	4.18	4.11
	4.76	5.12	4.93	4.21	4.21	4.11
	4.77	5.12	4.86	4.22	4.22	4.09

The position of the L-fucose residue is an interesting feature of the tetraantennary structure illustrated in Fig. 12. The unusual chemical shift of the H-5 proton of the fucose residue suggests a location of this residue in vicinal position to the galactose residue, both being substituents of an *N*-acetylglucosamine residue. A similar abnormal behavior of the signal of the H-5 proton of a fucose residue has been reported for the trisaccharide α-Fuc-(1→4)-[β-Gal-(1→3)]-GlcNAc having Lewis^a blood-group activity (43,44). A small but significant difference exists, however, between the chemical shifts of the H-5 in the tetraantennary structure and that of the Lewis^a structure. Furthermore, the tetraanntennary structure is completely devoid of Lewis activity, which excludes the occurrence of such a structural element in α_1-acid glycoprotein. Interestingly, the attachment of a fucose residue to the tetraantennary structure has a typical effect on the chemical shift of one of the *N*-acetyl groups. By comparison of the spectra of reference compounds, as illustrated in Fig. 13, this *N*-acetyl group was identified as being part of the GlcNAc (7) residue. Therefore, it was concluded that the fucose residue is located in the most upper branch, as indicated in Fig. 12.

In summary, high-resolution ^{1}H-n.m.r. spectroscopy is an extremely powerful tool for the elucidation of carbohydrate structures of glycopeptides. As it is a nondestructive technique, it can easily be incorporated in the usual analysis routes, provided that sufficient material of high purity is available.

PREDICTION OF GLYCAN STRUCTURES

All the described structures are in good agreement with the present concepts on the biosynthesis of glycoproteins. It is known that glycoprotein biosynthesis begins in the rough endoplasmic reticulum, where a part of the *N*-acetylglucosamine residues and the totality of the mannose residues are conjugated, and terminates in the smooth reticulum where the remaining *N*-acetylglucosamine residues and the totality of galactose, sialic acid, and fucose residues are conjugated.

On the other hand, the study of oligosaccharides, isolated from the urine of patients with lysozomal diseases, has furnished some interesting information on the catabolism of glycoproteins. These diseases result from a deficit of *exo*-glycosidases, which leads to an accumulation reaching often several grams per liter of oligosaccharides in tissues and urines. Schemes 12-16 illustrate the structures of oligosaccharides extracted from urines of patients with deficits in galactosidase (GM_1 gangliosidosis type I), mannosidase (mannosidosis), *N*-acetylglucosaminidase (Sandhoff's disease),

fucosidase (fucosidosis), and neuraminidase (mucolipidosis I, mucolipidosis II or I-Cell disease, mucolipidosis III, and mucolipidosis De P. and N). All these structures possess an *N*-acetylglucosamine residue at the reducing terminal position, to which a β-D-mannose residue is linked at C-4.

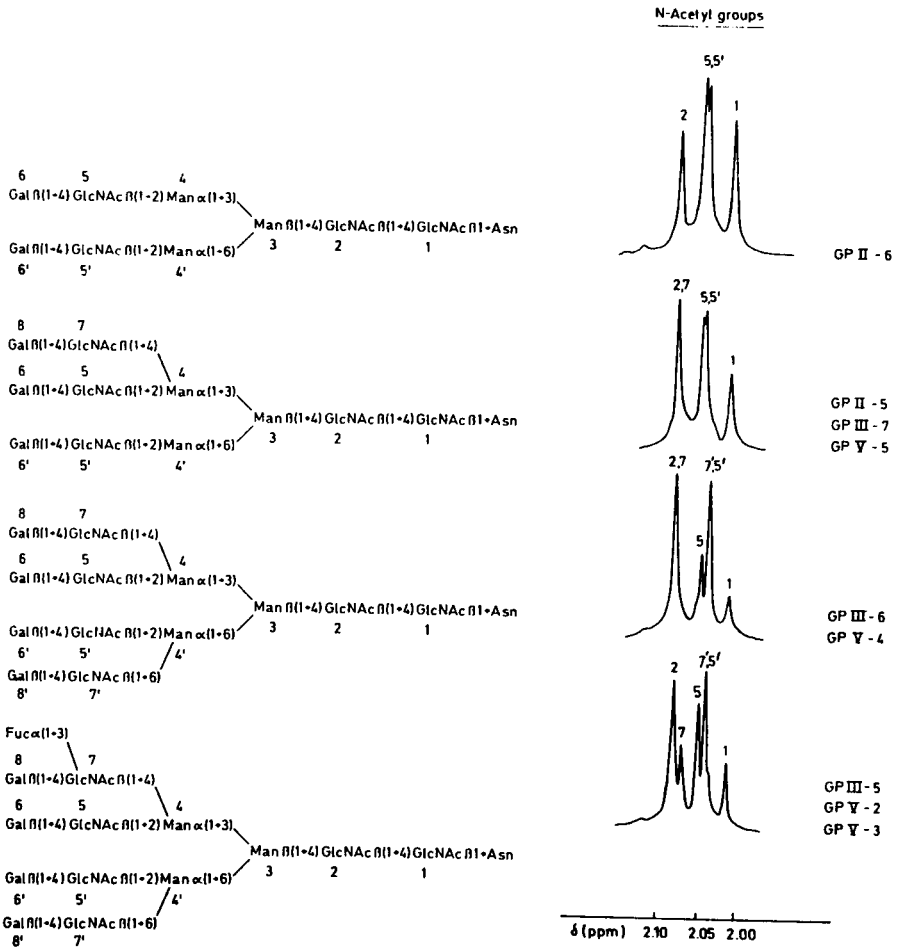
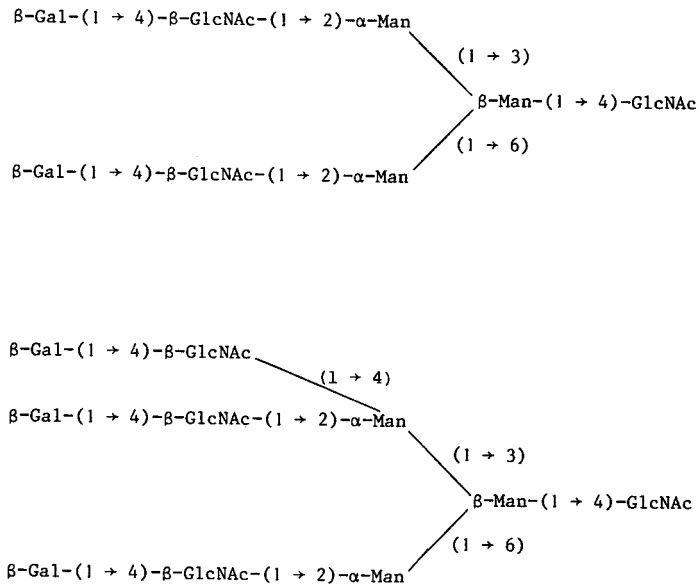
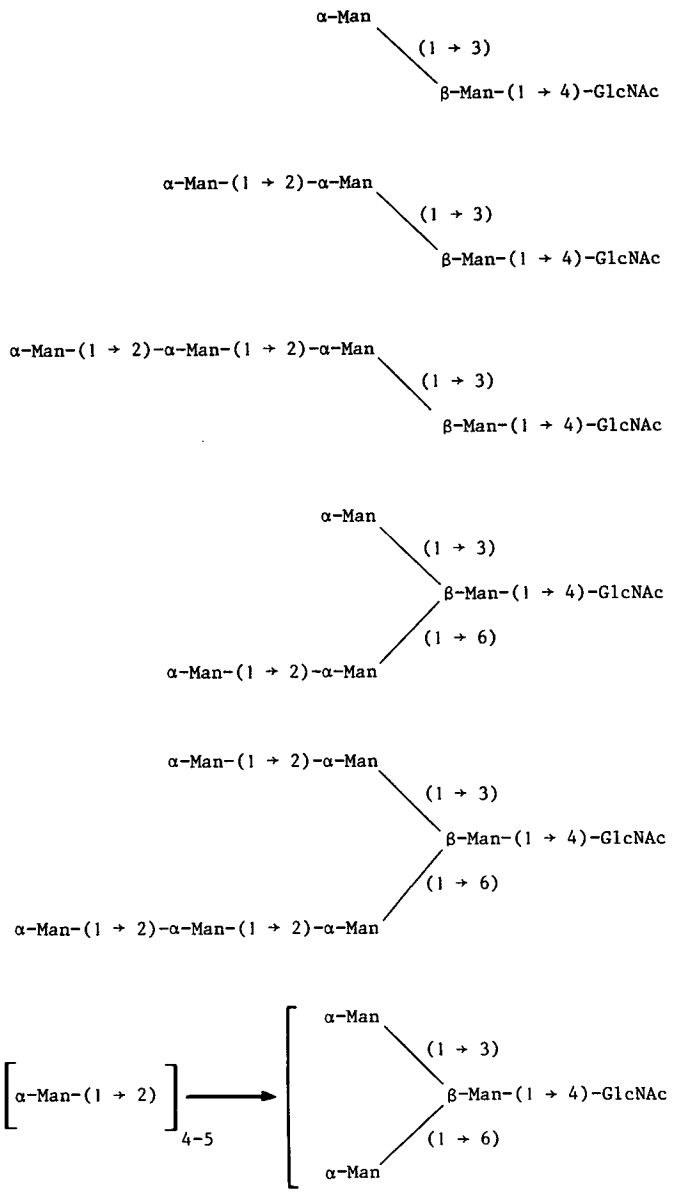


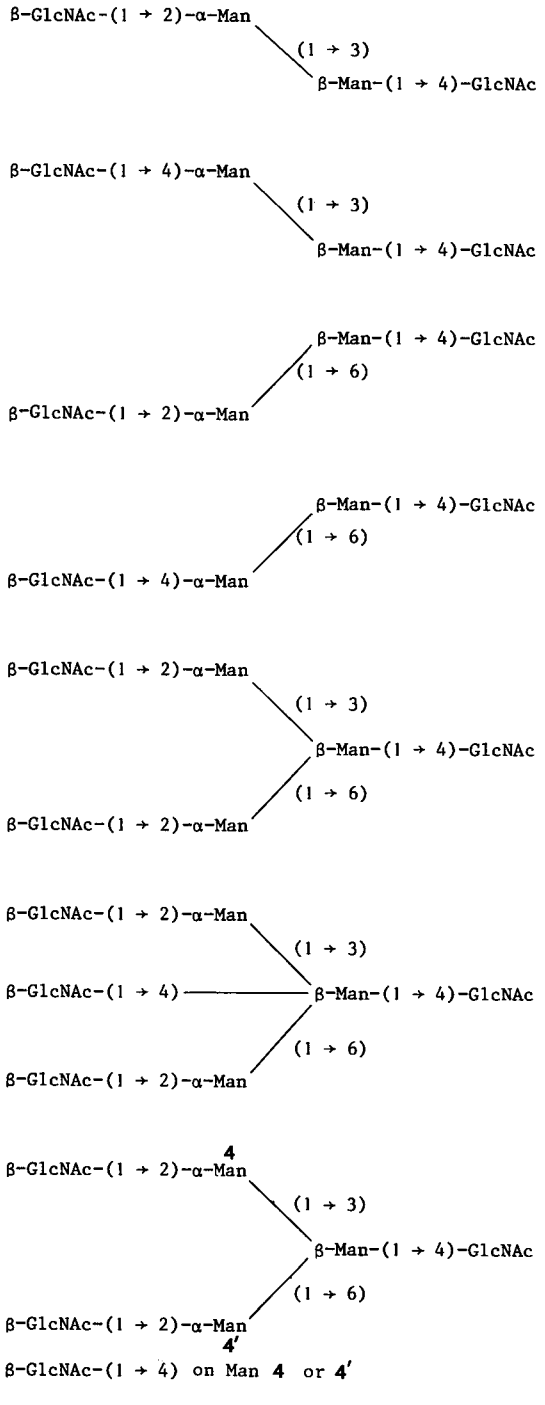
Fig. 13. Influence of the fucose residue on the N-acetyl protons in the tetraantennary structure.



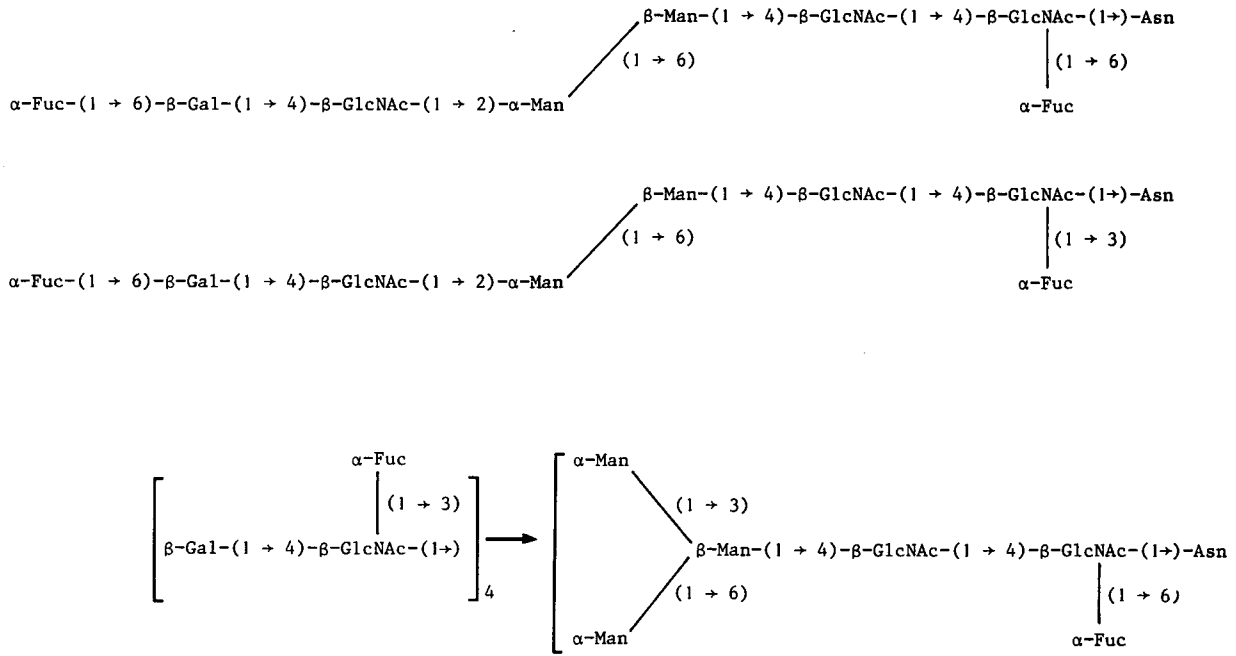
Scheme 12. Structure of oligosaccharides isolated from the liver of GM_1-gangliosidosis type I patients (45).



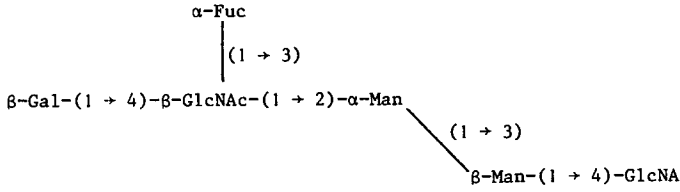
Scheme 13. Structure of oligosaccharides isolated from urine of mannosidosis patients (oligosaccharides 1 to 3, ref. 18; oligosaccharides 5 to 8, ref. 20).

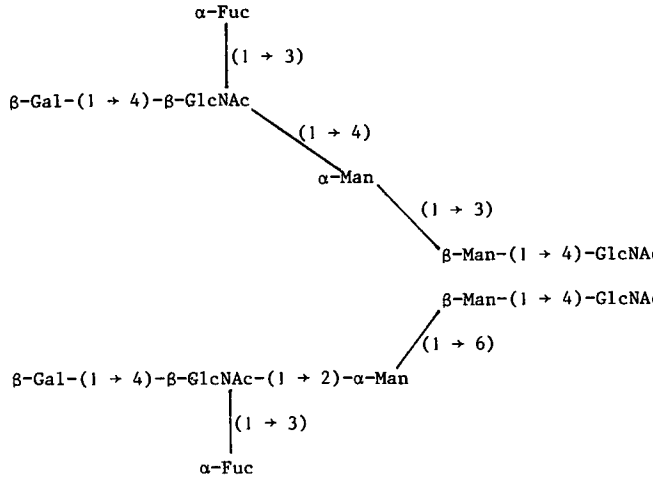


Scheme 14. Structure of oligosaccharides isolated from urine of Sandhoff's disease patient (23).

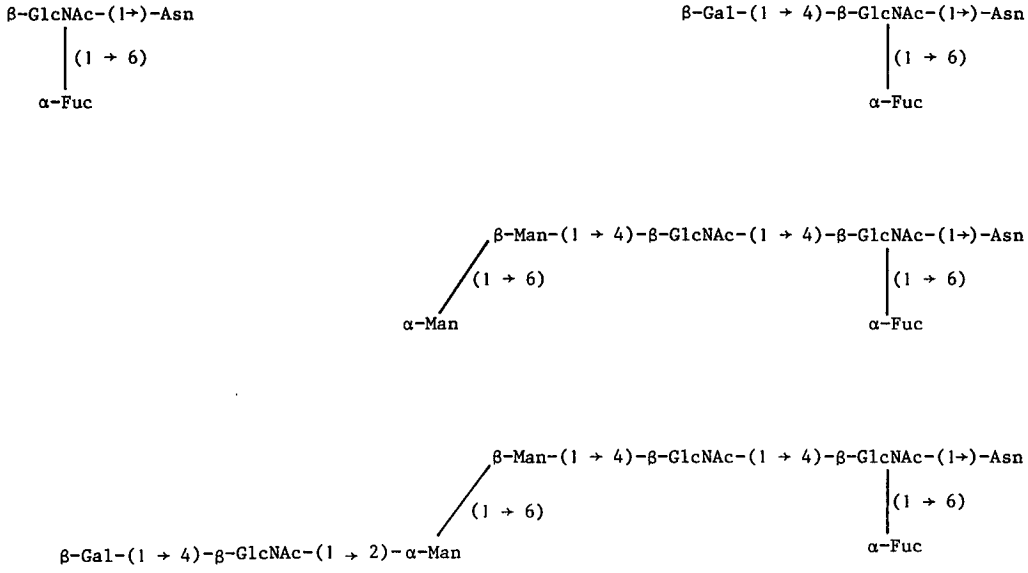


Scheme 15a. Structure of oligosaccharides isolated from the urine of a fucosidosis patient (21).

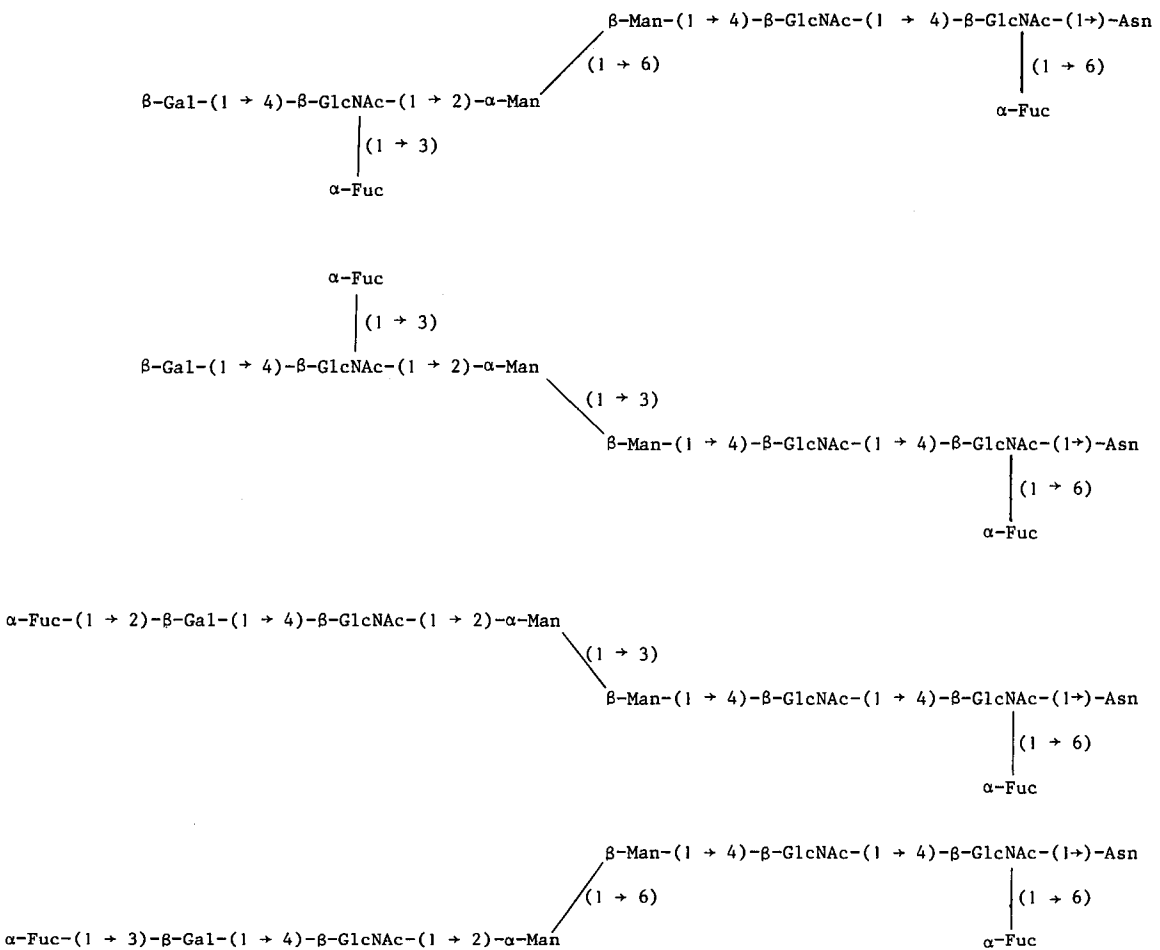




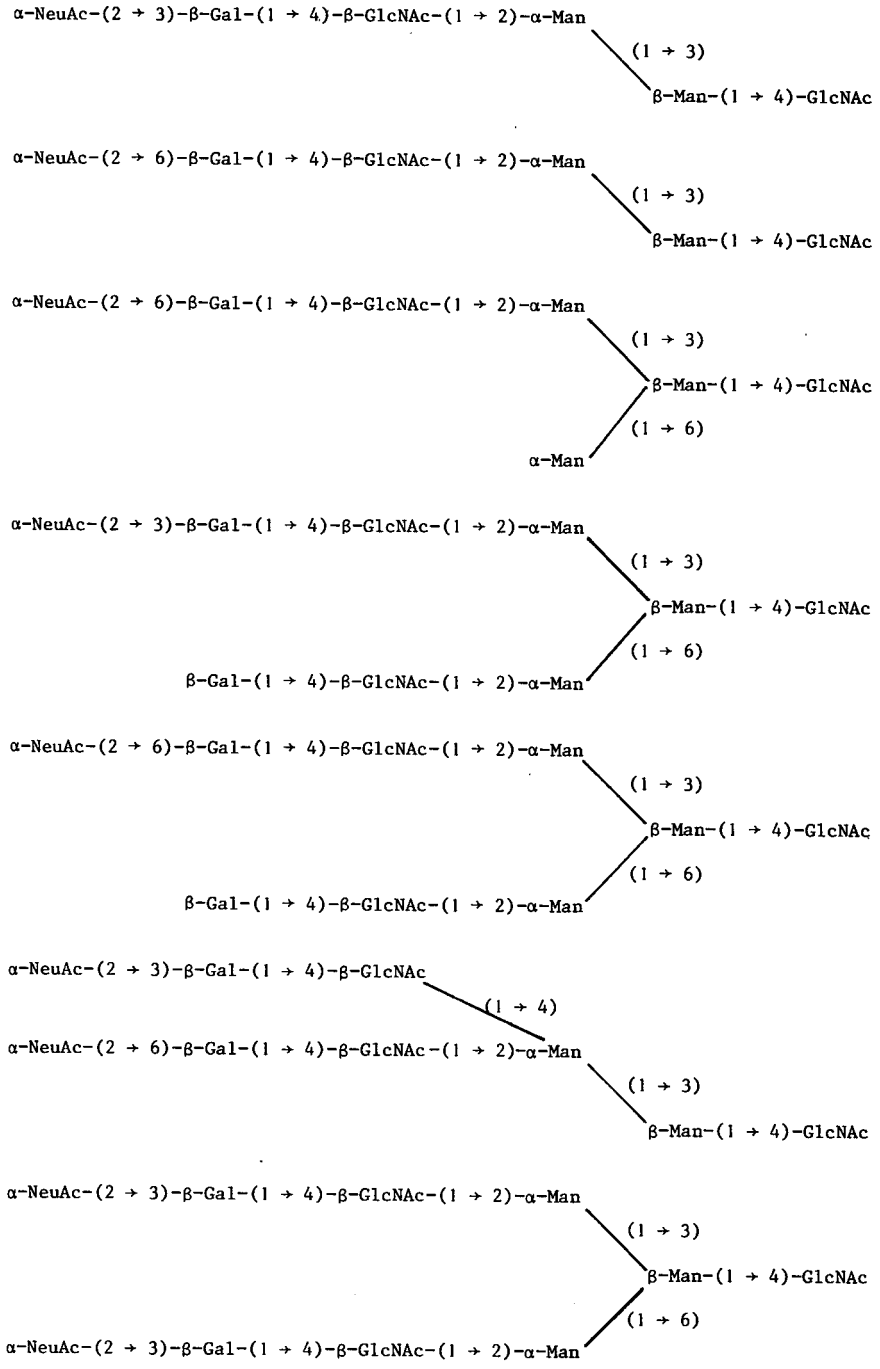
Scheme 15b. Structure of oligosaccharides isolated from the urine of a fucosidosis patient (21).



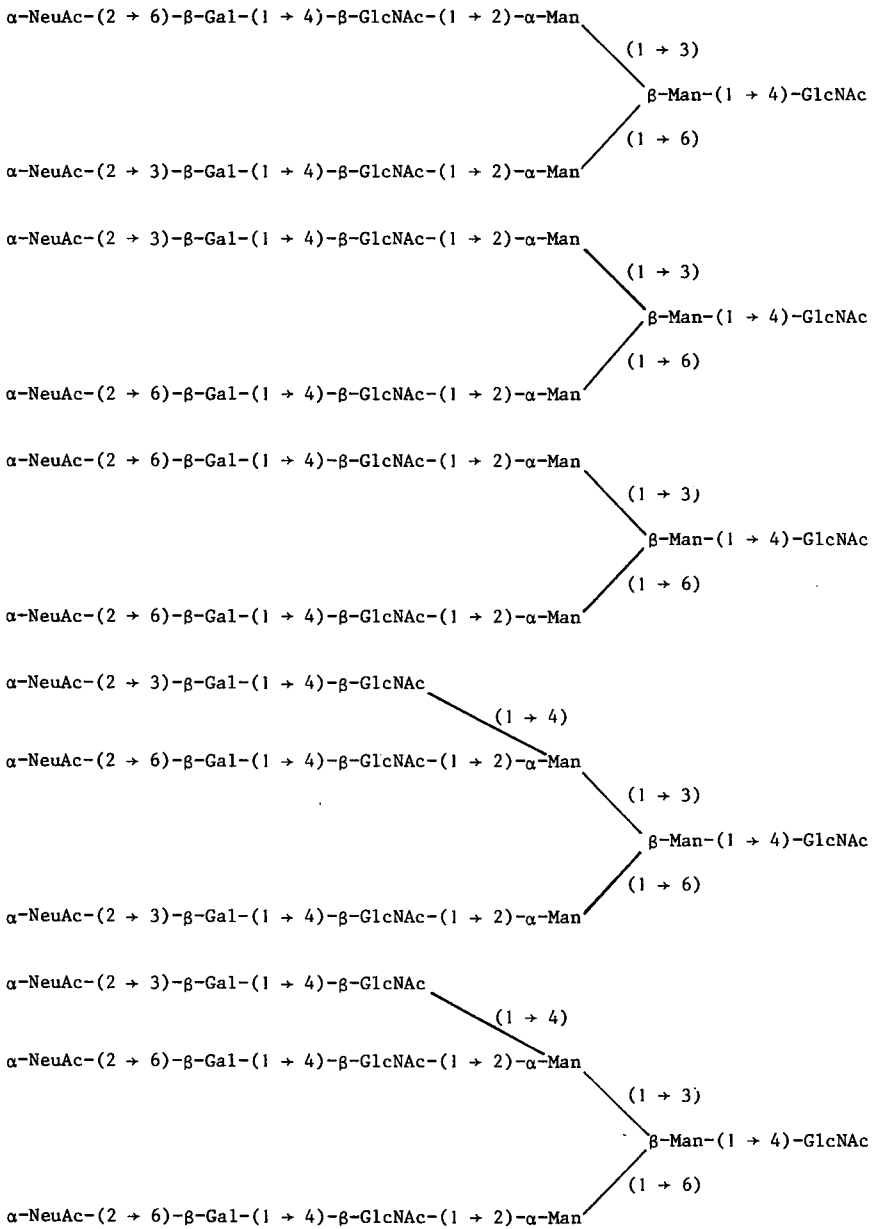
Scheme 15c. Structure of oligosaccharides isolated from the urine of a fucosidosis patient (21).



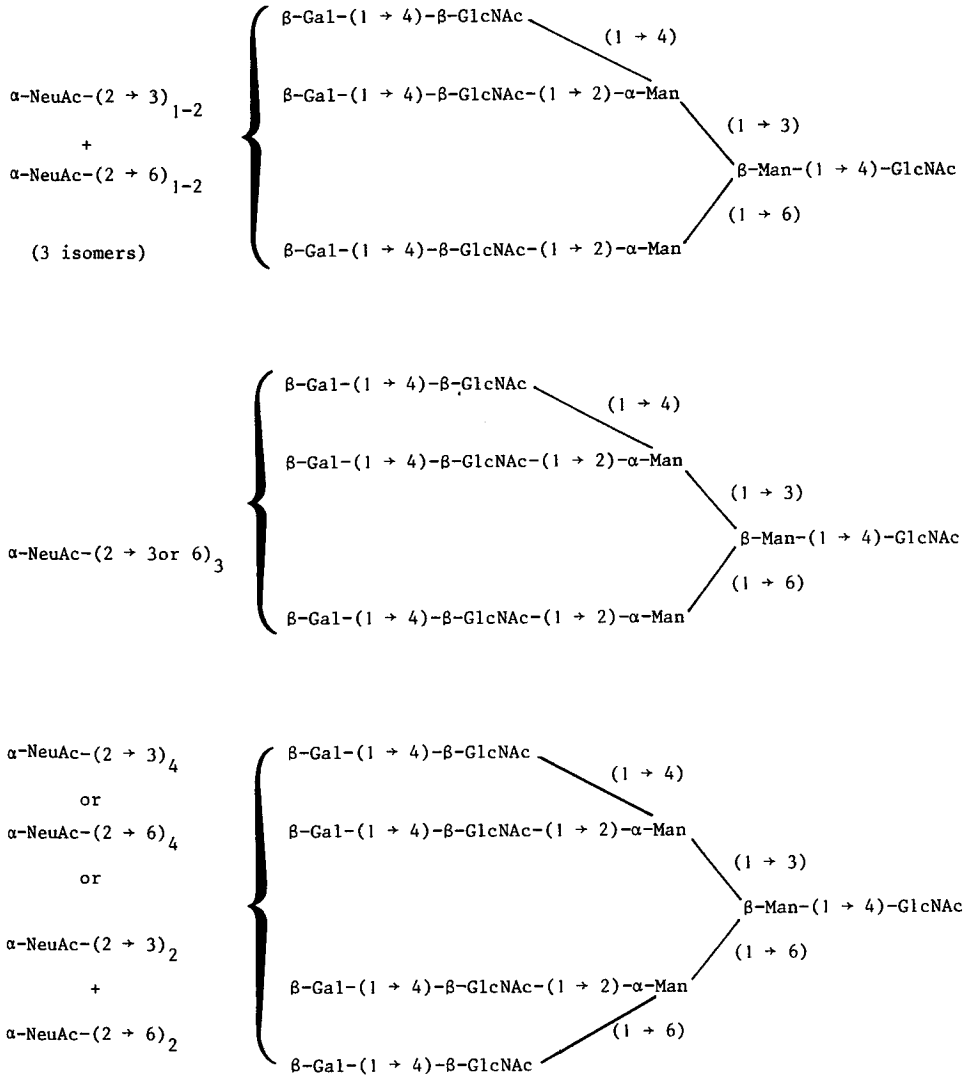
Scheme 15d. Structure of oligosaccharides isolated from the urine of a fucosidosis patient (21).



Scheme 16a. Structure of oligosaccharides isolated from the urine of a sialidosis patient (17,22,25,26).



Scheme 16b. Structure of oligosaccharides isolated from the urine of a sialidosis patient (7,22,25,26).



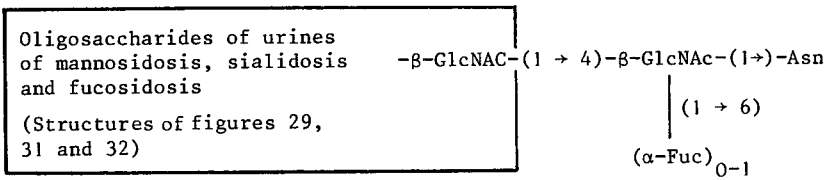
Scheme 16c. Structure of oligosaccharides isolated from the urine of a sialidosis patient (7,22,25,26).

Fragments of glycan structures, as they exist in numerous glycoproteins, are present. For example, the structures of the oligosaccharides of Sandhoff's disease may be considered as parts of the structures of serotransferrin (Fig. 8), of α_1-acid glycoprotein (Scheme 7) and of IgG (Scheme 10) glycans, respectively, and the oligosaccharides shown in Scheme 16 as parts of the structures of serotransferrin glycan or of α_1-acid glycoprotein *A*, *B*, or *C* glycans (Scheme 7).

On the basis of these observations, we have proposed the 3 following hypotheses: (a) The oligosaccharides from urines or tissues of lysozomal diseases originate from glycoprotein glycans, and we proposed to call this type of disease "glycoproteinoses" (4,19,25). (b) Consequently, we have reconstituted the first step of glycan catabolism, postulating that it begins with the action of *endo*-β-*N*-acetylglucosaminidases that cleave the di-*N*-acetylchitobiose residue (4,19,25). Thus, oligosaccharides having an *N*-acetylglucosamine residue in terminal reducing position would be liberated. *Endo*-β-*N*-acetylglucosaminidases acting on glycans of the oligomannoside type have been already characterized (46,47), but enzyme splitting sialo- and fuco-glycans of the *N*-acetyllactosamine type remain to be discovered. (c) Correlatively, we must postulate that the structures corresponding to the oligosaccharides of mannosidosis, sialidoses, and fucosidosis pre-exist in glycoprotein structures, even if these structures have not yet been characterized. For example, by adding the sequence β-GlcNAc-(1→4)-Asn to the tenth oligosaccharide of Scheme 16, we are able to reconstitute the complete glycan of serotransferrin. By adding the same sequence (Scheme 17) to all the oligosaccharides, accumulation of which is due to a lack in "terminal" exoglucosidases (fucosidases, neuraminidases, and mannosidases), we are able to reconstitute unknown but foreseen glycan structures. That such glycans have not yet been characterized might be due to the fact that they exist in too low quantities, probably in the cytoplasm or the cell membrane, or both. As the products of the action of *endo*-β-*N*-acetylglucosaminidases are being protected because of the lack of *exo*-glycosidases, they accumulate in the cells, and then in the urine.

PREDICTION OF GLYCAN CONFORMATION

Is it possible, on the basis of our knowledge of glycan primary structure, to imagine the spatial conformation of glycans, and is this conformation compatible with the biological role of glycoproteins, in particular with their role of recognition signals?



Scheme 17. General scheme of foreseeable glycan structures.

Fig. 14A illustrates the disposition in space of the biantennary glycan of human serotransferrin and shows that one can distinguish two parts in this structure. The first one is a compact zone, linked to the protein, and consisting of the pentasaccharide core itself. This part, common to numerous glycoprotein molecules, is non-specific and invariable. This was the reason to call it the invariant (inv) part (*4*), as in the case of the invariable peptide component of immunoglobulins. On the other hand, the trisaccharide fragment Man→GlcNAc→GlcNAc is practically flat (see Fig. 15). Moreover, this core is rigid because hydrogen bonds stabilize the structure.

To the compact and rigid core are attached the antennae. This part of the molecule is variable (var) and directs the specificity of the glycans. These antennae could occupy two positions in the space. The first one leads to the "Y conformation" (Fig. 14A), and the second one to the "T conformation" (Fig. 14B). The results from X-ray diffractometry obtained in collaboration with R. Fouret's group in our University are in favor of the second conformation. In fact, the analysis of the crystalline trisaccharide α-Man-(1→3)-β-Man-(1→4)-GlcNAc shows that the α-mannose residue is in a position perpendicular to that of the β-Man-GlcNAc residue.

The last information given by the construction of molecular models concerns the conformation of antennae which, after the formation of hydrogen bonds, appears helical (Fig. 15).

In conclusion, I would like to go back 52 years to present from Levene's book "Hexosamine and Mucoproteins" (48) the first structures of chondroitin sulfate and of fibrin "hyaloidin" (Schemes 18 and 19) so that we may contemplate the route that has been travelled. These structures are of course erroneous, however we should not smile but remember the pionners' works that during one century have lead us to the present knowledge of glycoprotein structure.

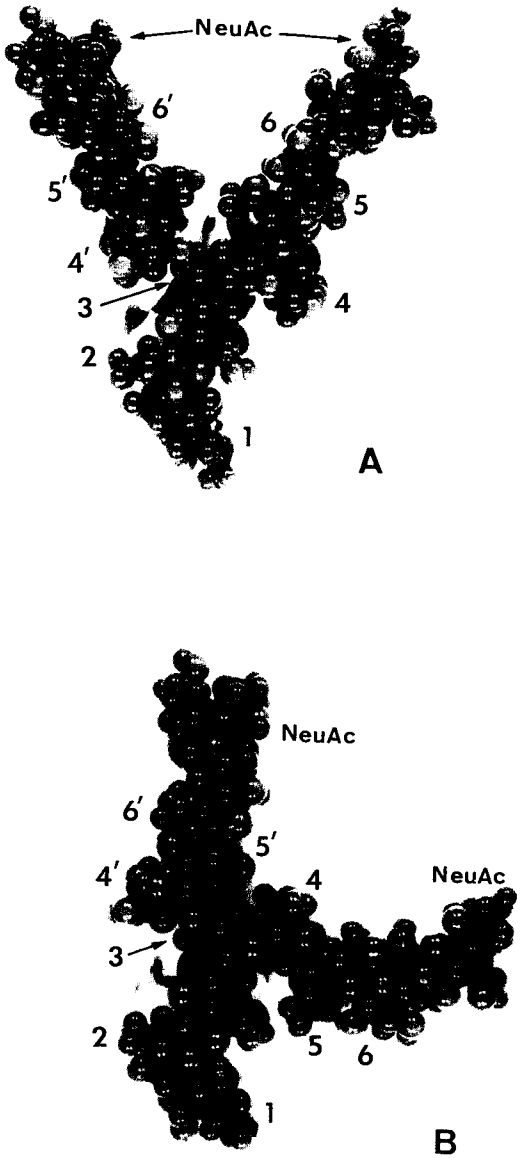


Fig. 14. Molecular model of human serotransferrin biantennary glycan: A,Y conformation; B,T conformation. The numbers correspond to those of Scheme 6.

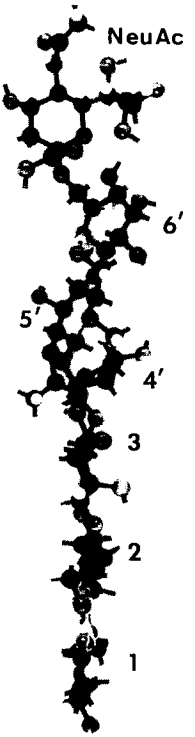
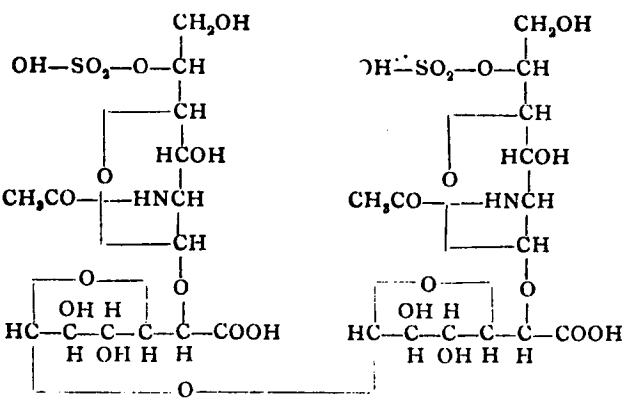
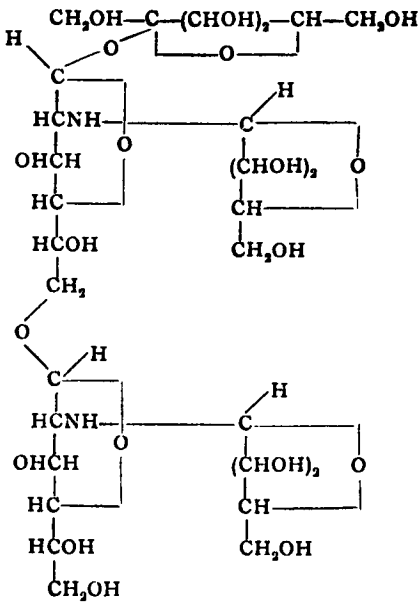


Fig. 15. Lateral view of the molecular model of N-*acetyl-lactosamine type. The numbers correspond to those of Scheme 6.*



Scheme 18. Schmiedeberg's structure of chondroitin sulfuric acid (in Levene, ref. 48).



Fibrin hyaloidin.

Scheme 19. Schmiedeberg's structure of fibrin hyalosidin (in Levene, ref. 48).

ACKNOWLEDGMENTS

The authors are much indebted to Drs. L. Dorland, B. Fournet, J. Haverkamp, B. L. Schut, G. Spik and G. Strecker for their enthusiastic contributions. These investigations were supported by The Netherlands Foundation for Chemical Research (SON) with financial aid from The Netherlands Organization for the Advancement of Pure Research (ZWO), the Centre National de la Recherche Scientifique (Laboratoire Associé n° 217), and the Université des Sciences et Techniques de Lille.

REFERENCES

1. Eylar, H., *J. Theor. Biol. 10*, 89 (1965).
2. Winterburn, P.-J., and Phelps, C.F., *Nature (London) 236*, 147 (1972).

3. Ashwell, G., and Morell, A. G., *in* "Glycoproteins of Blood Cells and Plasma" (G. A. Jamieson and T. J. Greenwalt, eds.), p. 173, J. B. Lippincott Co., Philadelphia (1971).
4. Montreuil, J., *Pure Appl. Chem. 42,* 431 (1975).
5. Bayard, B., and Montreuil, J., *C.N.R.S. Int. Symp. 221,* 209 (1974).
6. Tai, T., Yamashita, K., Ogata-Arakawa, M., Koide, N., Muramatsu, T., Shintaro, I., Inoue, Y., and Kobata, A., *J. Biol. Chem. 250,* 8569 (1975).
7. Yamaguchi, H., Ikenaka, T., and Matsushima, Y., *J. Biochem. (Tokyo) 70,* 587 (1971).
8. Nakajima, T., and Ballou, C. E., *J. Biol. Chem. 249,* 7685 (1974).
9. Kondo, T., Fukuda, M., and Osawa, T., *Carbohydr. Res. 58,* 405 (1977).
10. Pesonen, M., and Renkonen, O., *Biochim. Biophys. Acta 455,* 510 (1976).
11. Baenziger, J., and Kornfeld, S., *J. Biol. Chem. 249,* 7260; 7270 (1974).
12. Kearns, D. R., and Shulman, R. G., *Accounts Chem. Res. 7,* 33 (1974).
13. Miller, F., *Immunochemistry 9,* 217 (1972).
14. Finne, J., Krusius, T., and Rauvala, H., *Biochem. Biophys. Res. Commun. 74,* 405 (1977).
15. Spik, G., Bayard, B., Fournet, B., Strecker, G., Bouquelet, S., and Montreuil, J., *FEBS Lett. 50,* 296 (1975).
16. Bayard, B., Fournet, B., Bouquelet, S., Strecker, G., Spik, G., and Montreuil, J., *Carbohydr. Res. 24,* 445 (1972).
17. Michalski, J.-C., Strecker, G., and Fournet, B., *FEBS Lett. 79,* 101 (1977).
18. Norden, N.E., Lundblad, A., Svensson, S., Ockerman, P. A., and Autio, S., *J. Biol. Chem. 248,* 6210 (1973); *Biochemistry 13,* 871 (1974).
19. Strecker, G., *in* "J.-P. Farriaux, Les oligosaccharidoses" (Crouan et Roques, eds.), p. 13, Lille (1977).
20. Strecker, G., Fournet, B., Bouquelet, S., Montreuil, J., Dhondt, J. L., and Farriaux, J.-P., *Biochimie 58,* 579 (1976).
21. Strecker, G., Fournet, X., Spik, G., Montreuil, J., Durand, P., and Tondeur, M., *C.R. Acad. Sci. Ser. D. 284,* 84 (1977); and unplublished results.
22. Strecker, G., Hondi-Assah, T., Fournet, B., Spik, G., Montreuil, J., Maroteaux, P., Durand, P., and Farriaux, J.-P., *Biochim. Biophys. Acta 444,* 349 (1976).
23. Strecker, G., Herlant-Peers, M-C., Fournet, B., and Montreuil, J.; Dorland, L., Haverkamp, J., and Vliegenthart, J.F.G.; Farriaux J.-P., *Eur. J. Biochem. 81,* 165 (1977).
24. Strecker, G., and Lemaire-Poitau, A., *J. Chromatog. (Biochem. Applic.) 1,* 553 (1977).

25. Strecker, G., and Montreuil, J., *Biochimie,* in press.
26. Strecker, G., Peers, M.-C., Michalski, J.-C., Hondi-Assah, T., Fournet, B., Spik, G., Montreuil, J., Farriaux, J.-P., Maroteaux, P., and Durand, P., *Eur. J. Biochem. 75,* 391 (1977) and unpublished results.
27. Bovey, F. A., "High Resolution NMR of Mâcromolecules", Academic Press, New York, 1972.
28. Dwek, R. A., "Nuclear Magnetic Resonance (NMR) in Biochemistry, Applications to Enzyme Systems" Clarendon Press, Oxford (1973).
29. Kearns, D. R., and Shulman, R. G., *Acc. Chem. Res. 7,* 33 (1974).
30. Roberts, G. C. K., and Jardetzky, O., *Adv. Protein Chem. 24,* 447 (1970).
31. Wüthrich, K., "NMR in Biological Research: Peptides and Proteins", North-Holland Pub. Co., Amsterdam (1976).
32. Durette, P. L., and Horton, D., *Adv. Carbohydr. Chem. Biochem. 26,* 49 (1971).
33. Hall, L. D., *Adv. Carbohydr. Chem. 19,* 51 (1964).
34. Inch, T. D., *Ann. Rev. NMR Spectroscop. 2,* 35 (1969).
35. Streefkert, D. G., Thesis, University of Utrecht, 1973.
36. Haverkamp, J., Thesis, University of Utrecht, 1974.
37. Dorland, L., Schut, B. L., Vliegenthart, J. F. C., Strecker, G., Fournet, B., Spik, G., and Montreuil, J., *Eur. J. Biochem. 73,* 93 (1977).
38. Streefkerk, D. G., De Bie, M. J.A., and Vliegenthart, J. F. G., *Tetrahedron 29,* 833 (1973).
39. Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., Strecker, G., Fournet, B., Spik, G., and Montreuil, J., to be published.
40. Usui, T., Yamaoka, N., Matjuda, K., Tuzimura, K., Sugiyama, H., and Seto, S., *J. Chem. Soc. Perkin I,* 2425 (1973).
41. Colson, P., and King R. R., *Carbohydr. Res. 47,* 1 (1976).
42. Dorland, L., Haverkamp, J., Schut, B. L., Vliegenthart, J. F. C., Spik, G., Strecker, G., Fournet, B., and Montreuil, J., *FEBS Lett. 77,* 15 (1977).
43. Lemieux, R. U., Bundle, D. R., and Baker, D. A., *J. Am. Chem. Soc. 97,* 4076 (1975).
44. Lemieux, R. U., and Driguez, H., *J. Am. Chem. Soc. 97,* 4063 (1975).
45. Wolfe, L. S., Senior, R. G., and Ng Ying Kin, N. M. K., *Fed. Proc. 32,* 484 (1973); *J. Biol. Chem. 249,* 1828 (1974).
46. Muramatsu, T., *J. Biol. Chem. 246,* 5535 (1971).
47. Tarentino, A. L., Plummer, T. H., Jr., and Maley, F. J., *J. Biol. Chem. 248,* 5547;(1973); Tarentino, A. L., and Maley, F., *J. Biol. Chem. 249,* 811 (1974).
48. Levene, P. A., "Hexosamine and Mucoproteins", Longmans Green, and Co., London (1925).