

Review

The complexity of glycoprotein-derived glycans

By Johannes F. G. Vliegenthart^{*1,†}

(Communicated by Kunihiko SUZUKI, M.J.A.)

Abstract: A brief review is presented of our studies on the structure of glycoprotein-derived glycans. The emphasis is on the introduction of high-resolution ¹H-NMR spectroscopy for the unambiguous determination of primary structures. For this purpose, we developed the structural reporter group concept. Structural reporters are defined as unique markers of structural elements in the NMR spectra. Application of this concept led to the discovery of numerous new structures. Furthermore, a number of structures presented in the literature could be corrected. The results are relevant for insight in the various steps in glycan metabolism in health and disease, for the function and mode of action of glycans *in vivo* and for the interpretation of structural information obtained through other techniques. The strength of the approach is further shown for several highly complex glycoproteins, carrying very heterogeneous and complicated glycans.

Keywords: N-glycans, O-glycans, proteo-glycans, immuno-determinants, blood groups, NMR spectroscopy, structural reporter group concept

Introduction

Glycosylation is one of the most prominent post-/co-translational modifications of proteins. The attachment of covalently linked glycans to a protein is a non-template-driven process, carried out by interplay of sequentially acting specific glycosyltransferases and glycosidases. The latter enzymes play a main role in trimming of intermediary structures in the biosynthesis. The glycans can be attached to the protein at one or more glycosylation sites via N, O, C or S-atoms in the amino acid side chains. In the main N-glycoproteins the linkage occurs via the amide side chain of asparagine that forms part of the consensus sequence AsnXxxSer/Thr, wherein Xxx can be any amino acid except Pro. O-glycans are linked via the side chain of hydroxy amino acids.¹⁾⁻³⁾ Up to now, C-linked mannose is the only observed example of a C-linked glycan. Mannose is exclusively attached to the indole nucleus of tryptophan.^{4),5)} S-glycosylation is rare, and only observed in glycopeptides.⁶⁾ Due to

its nature, the biosynthesis of glycoproteins can give rise to the formation of a plethora of compounds that have in common identical protein chains, but differ in structure of the glycan chains and/or in occupancy of the glycosylation sites. This feature constitutes the so-called (micro)-heterogeneity of glycoproteins. By consequence a glycoprotein comprises usually more than one molecular structure. The heterogeneity complicates single-molecule studies and the assignment of a biological or physical function to a specific glycan structure. A further complication in assigning a specific function to a single glycan structure is due to the occurrence in various types of glycoproteins. These biomacromolecules may be excreted, function as membrane constituents or may be part of the extracellular matrix. Therefore, glycans having the same primary structure might differ in function due to their different presentation to the environment. Defining an *in vivo* function at the molecular level may require knowledge of the molecular context/architecture, wherein the glycan exerts its function.

For gaining fundamental insight in the structure-function relation of glycans of a glycoprotein and of the glycoprotein as a molecular entity, structural information is needed. A first step is the determination of the primary structure. This comprises essential structural elements like the description of a number of parameters including *i*) the

^{*1} Bijvoet Center for Biomolecular Research, Utrecht University, The Netherlands.

[†] Correspondence should be addressed: J. F. G. Vliegenthart, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands (e-mail: j.f.g.vliegenthart@uu.nl).

Dedicated to the memory of Dr. Yoshitaka Nagai.

structure of the protein chain, *ii*) the structure of glycan chains, *iii*) the location of the glycosylation site(s) in the protein and *iv*) the degree of occupation of the glycosylation sites. Finally, since the molecular environment at the place of action has influence on the function, definition of the molecular embedding is a relevant aspect; this holds in particular for membrane-bound glycoproteins.

This brief review presents a selection of the contributions of our research group to the study of the structure of glycans of glycoproteins to illustrate the large variety in glycan-structure and in possible function. The attention is focused mainly on results we obtained by our introduction of high-resolution $^1\text{H-NMR}$ spectroscopy for structural analysis.⁷⁾ This method opened a new approach in structural studies.

Primary structure of glycans. An important issue concerns the isolation of a 'pure' glycoprotein as starting material. This includes also the proper selection of well-defined cells and/or tissues. In recent years, the development of advanced chromatographic methods gave great improvements, although the aforementioned (micro)heterogeneity leads almost by definition to a family of compounds. With the increased sensitivity and resolution power of up-to-date analytical techniques, it is still of crucial importance to distinguish properly, natural heterogeneity from heterogeneity arising from contaminating glycoproteins. Glycan analyses can rarely be performed on the intact glycoprotein. Therefore, the glycoprotein is degraded to partial structures by means of enzymatic or chemical reactions. Then the fragments, which may be oligosaccharides, oligosaccharide-alditols or glycopeptides, are isolated, purified and the structure determined. Subsequently, the glycoprotein structure can be deduced if sufficient fragments have been analysed.

For a long time structure analysis of the glycan moieties made only a rather slow progress, due to limitations of the used analytical methods. The breakthrough in this area resulted from the development of high-resolution instrumentation and analytical techniques like nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS). Sophistication of the methodology was a prerequisite to enable the structure determination of complex glycans, since a large number of parameters have to be established to identify the structure:

- * Nature and number of the constituting monosaccharides.
- * Ring size and absolute configuration of the monosaccharides.

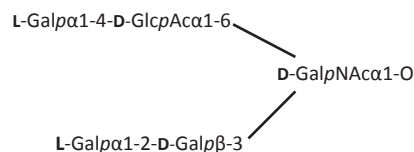


Fig. 1. Structure of glycopeptide CeTx, obtained from *Conus consors*. Note that the glycan contains D- as well as L-Gal residues.

- * Sequence of the monosaccharides.
- * Type and anomeric configuration of the glycosidic linkages.
- * Type of the carbohydrate-amino acid linkage and position of the corresponding amino acid in the protein.
- * Type and position of non-carbohydrate substituents.
- * Occupation of the glycosylation sites.

We focused on the development of an integral methodology leading to unambiguous answers for the structure of glycans. To achieve this goal, in general information from more than one independent method is needed. We have chosen for NMR and if feasible MS, in combination with chemical techniques. We established that methanolysis of glycoproteins, followed by monosaccharide analysis *e.g.*, by GLC of suitable derivatives is adequate for establishing the identity of the constituting monosaccharides.⁸⁾ A new method was designed for determining the absolute configuration of monosaccharides (D or L) based on GLC of diastereomers. These are obtained through derivatization with (–)-2-butanol.^{9),10)} This step is particularly relevant for non-mammalian systems, since even both enantiomers may occur in one glycan. This is illustrated in Fig. 1 for a *O*-glycopeptide isolated from the venom of the piscivorous cone snail *Conus consors*, wherein indeed both D and L galactose residues occur.¹¹⁾

For determination of the position of glycosidic linkages, we applied methylation analysis.¹²⁾ Analysis of the occupation of glycosylation sites is performed by analysis of (glyco)peptides.

Mass spectrometry. The advances of mass spectrometry are impressive. In particular, the progress made in the mass range that can be analysed, the developments in ionization techniques and importantly, the advancement in glyco-bio-informatics rendered analysis of carbohydrates by mass spectrometry feasible. This could be shown by us for the analysis of mono- and oligosaccharides as pertrimethylsilyl, or permethyl derivatives with electron impact ionisation. A significant number of

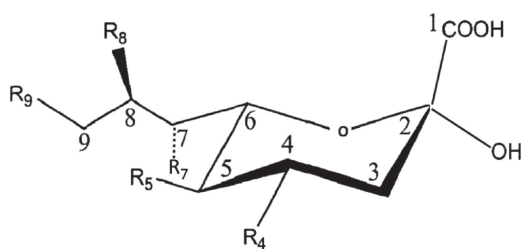


Fig. 2. Structure of neuraminic acid. Possible substituents are: R4: OH or OAc. R5: NH₂, NAc, NGI or OH. R7: OH or OAc. R8: OH, OAc, OCH₃, SO₄ or Sia. R9: OH, OAc, O-lactyl, PO₄, SO₄ or Sia.

structural details can be elucidated.¹³⁾ The introduction of Fast Atom Bombardment (FAB) ionisation enlarged further the mass range of compounds that can be analysed. We could show this for FAB mass spectrometry in the positive as well as negative mode for underivatized oligosaccharides and glycopeptides obtained from glycoproteins.¹⁴⁾ In the past years, main steps forward as to instrumentation and ionization methods are realized, enabling the study of intact glycoproteins. For example, the molecular multiplicity of glycoproteins has recently been illustrated for intact chicken ovalbumin. High-resolution native electrospray ionization mass spectrometry on a modified Exactive Orbitrap mass analyser demonstrated qualitatively and semi-quantitatively 59 proteoforms in the natural protein, induced by 45 different glycan structures and a number of phosphorylation sites.¹⁵⁾

In a joint program with R. Schauer *et al.*, we applied MS in conjunction with ¹H-NMR successfully for identification of the structure of sialic acid residues.^{16),17)} Sialic acids proved to form a large family of compounds that arises from variation in substituents at different positions in the basic structure.¹⁸⁾ This is shown in Fig. 2.

One dimensional (1D) ¹H-NMR spectra. For elucidation of glycan structures, we introduced high-resolution (360–900 MHz) ¹H-NMR spectroscopy. NMR spectroscopy is non-destructive, thereby leaving all possibilities open for additional analyses. The NMR spectra disclose the greater part of the many structural details and allow mostly the deduction of unambiguous structures. The starting set of ¹H-NMR data, in terms of chemical shifts and coupling constants could be collected and verified thanks to the availability of so many partial structures of glycoprotein-glycans isolated from the urine of patients with inborn errors of glycan-metabolism like oligomannosidosis, sialidosis, fucosidosis,

Gaucher's and Sandhoff's diseases. To record ¹H-NMR spectra we dissolve glycan samples in ²H₂O to exchange labile ¹H for ²H. Subsequently, spectra of glycans are recorded in ²H₂O solution. For the interpretation of these spectra we created the 'structural reporter group concept'.^{19)–24)} The 1D ¹H-NMR spectra contain a composite, so-called bulk signal, from about δ 3.2–3.9 ppm, mainly representing ring protons with similar chemical shifts that cannot be assigned easily to individual atoms. For unravelling the bulk signal and other complex signals, techniques like homo- and heteronuclear 2D NMR spectroscopy are applied. Outside of the bulk region, well-defined resonances can be distinguished, which are characteristic of structural elements. These signals represent the structural reporter group signals that can be used as structural identifiers. The main structural reporter groups comprise:

- * Anomeric protons.
- * Protons shifted out of the bulk-region due to glycosylation shifts or due to the influence of substituents, such as sulfate, phosphate, alkyl- and acyl-groups.
- * Deoxy sugar protons.
- * Protons of alkyl and acyl substituents like methyl, acetyl, glycolyl *etc.*

Examples of 1D ¹H NMR spectra are shown in Fig. 3a,b.

General features of N-glycoproteins and the corresponding ¹H-NMR data. In the biosynthesis of eukaryotic N-glycoproteins the Glc₃Man₉GlcNAc₂ oligosaccharide is exclusively transferred *en bloc* to Asn in a consensus sequence of the protein chain. The N-linked glycans are subsequently stepwise modified by the action of glycosidases and glycosyl-transferases. Chemically the N-glycans are generally characterised by a common pentasaccharide core, consisting of an N,N-diacetylchitobiose unit extended with a branched tri-mannosyl entity as shown in Fig. 4a. Characteristic for this structural element are the reporter group chemical shifts of Man H1, H2 and the GlcNAc NAc groups as summarized in refs. 19–24. The terminal mannoses can be elongated with exclusively mannose residues giving rise to oligomannose type as shown in Fig. 4b. The biosynthesis can further give rise to structures that can be conceived as extensions of the terminal Man residues of the tri-mannosyl core with N-acetylglucosamine moieties, arising from the sequential interplay of glycosidases and glycosyl transferases. This process may afford di-antennary compounds, further branching can give rise to tri-, tri'- and tetra-antennary structures, as

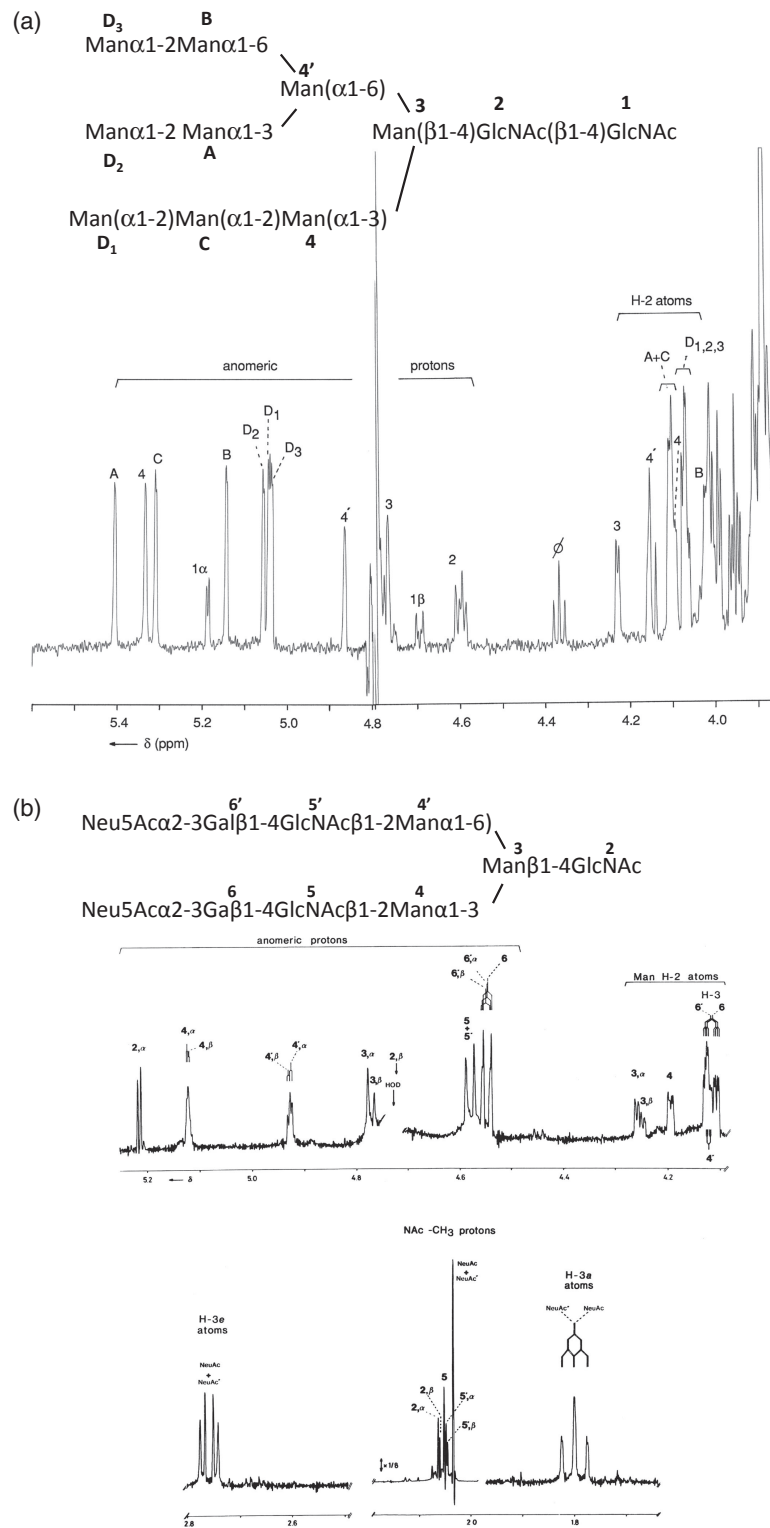


Fig. 3. (a) 1-D ^1H -NMR spectrum of the $\text{Man}_9\text{GlcNAc}_2$ recorded in $^2\text{H}_2\text{O}$. The structural reporter groups of the individual Man residues are assigned. (b) 1-D ^1H -NMR spectrum of the disialylated diantennary *N*-glycan, recorded in $^2\text{H}_2\text{O}$. The relevant reporter group signals like NH-, anomeric-, Man-H2, N-acetyl, and Neu5Ac H3 signals are indicated. Note the long distance effect of the reducing GlcNAc-2.

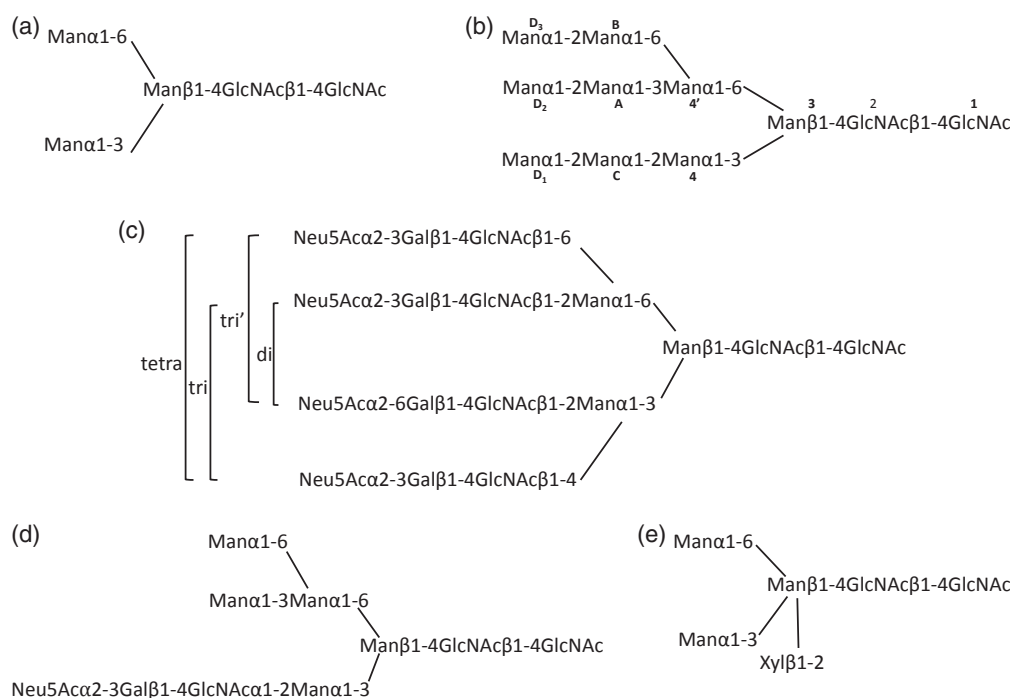


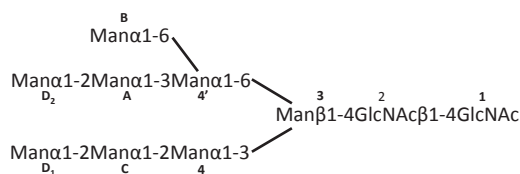
Fig. 4. (a) Common pentasaccharide core of *N*-glycans, containing a (β 1-4) linked Man residue, which acts as branching point. (b) Oligomannose type of *N*-glycan including numbering of Man residues in the Man₉ structure. (c) *N*-acetyllactosamine type of *N*-glycan, with sialylated di-, tri-, tri' and tetra-antennae as extensions of the core. (d) Hybrid type of *N*-glycans. The Man(α 1-6) branch is extended with the oligomannose type and the Man(α 1-3) arm with the *N*-acetyllactosamine type of structure. (e) A branched trimannosyl entity extended with a Xyl residue (β 1-2) linked to β Man.

illustrated in Fig. 4c. The branching pattern and the antennae can be identified on the basis of their reporters namely the Man H-1 and H-2 signals as summarized in Table 5 of ref. 23. For the lactosamine units, the anomeric signals of Gal and GlcNAc and the NAc resonances of GlcNAc, together with glycosylation shifts are characteristic as discussed in ref. 23. These glycans are designated as complex or *N*-acetyllactosamine type. Extension of the trimannosyl core with *N*-acetyllactosamine moieties at the Man-3 arm and with Man residues at the Man-6 arm yields the so-called hybrid type of chain, as shown in Fig. 4d. The spectra comprise a combination of the NMR signals of the corresponding antennae as presented in Table 6 in ref. 23.

Extensions of Asn-linked GlcNAc may be L-Fuc in (α 1-6) and/or (α 1-3) linkage. Fuc may also occur at peripheral parts of the glycan in (α 1-2), (α 1-3), and/or (α 1-4) linkage. As summarised in Table 8 of ref. 23 the data of the structural reporter group signals for Fuc H1 and CH₃ are characteristic for the various linkages.

Bisecting GlcNAc concerns a GlcNAc residue attached in (β 1-4) linkage to the branching Man **3** of the core. Introduction of this residue shows the own signals and the effects on the structural reporter groups Man H-1, H-2 and NAc as presented in Table 6 of ref. 23. Another example is the occurrence of Xyl residues as shown in Fig. 4e. Sialic acids can occur as terminal residues at antennae. The type of glycosidic linkage (α 2-3), (α 2-6) or (α 2-8) can be deduced from the chemical shifts of H3_{eq} and H3_{ax}, as shown in Table 8 of ref. 23. Substituents of the glycan like alkyl and acyl groups are identified by the corresponding ¹H NMR chemical shifts. Inorganic substituents like sulfate and phosphate induce typical downfield shifts of neighbouring protons.

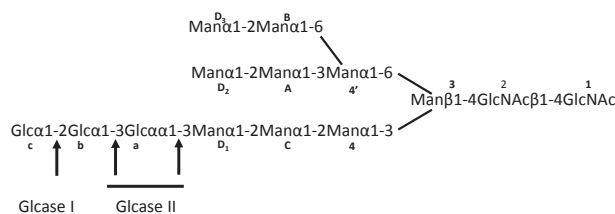
Oligomannose glycans. As result of the biosynthesis the Man₉GlcNAc₂ can be conceived as a mature glycan for the eukaryotic system. Interestingly, for soybean agglutinin it was described that two types of *N*-glycans would occur each consisting of Man and GlcNAc in a molar ratio of 9 : 2, but differing in branching pattern.²⁵⁾ However,

Fig. 5. Man_8 structure missing Man-D_3 .

^1H NMR studies showed that only the well-known $\text{Man}_9\text{GlcNAc}_2$ sequence as presented in Fig. 4b is occurring.²⁶⁾ In many glycoproteins from plant and animal origin, also partial structures of the ‘mature’ Man_9 chain occur. In Fig. 4b numbering of the Man residues has been given. In legume storage proteins *e.g.*, in kidney bean glycoprotein II the structures ranged from $\text{Man}_9\text{GlcNAc}_2$ to $\text{Man}_6\text{GlcNAc}_2$ and in 75-soybean glycoprotein from $\text{Man}_8\text{GlcNAc}_2$, to $\text{Man}_6\text{GlcNAc}_2$.²⁷⁾ In $\text{Man}_8\text{GlcNAc}_2$, D_3 is absent as shown in Fig. 5, in the other compounds partially also D_2 and D_3 are missing. In bovine lactotransferrin is besides the $\text{Man}_9\text{GlcNAc}_2$ structure also $\text{Man}_8\text{GlcNAc}_2$ found, missing D_1 .²⁸⁾ In lysosomal α -mannosidase from porcine kidney Man_9 is present and Man_6 , lacking D_1 , D_2 and D_3 , and also two isomers of Man_5 lacking D_1 , D_2 , D_3 and A as well as a compound missing D_1 , D_2 , D_3 and C .²⁹⁾

Also in urine of some patients with inborn errors of metabolism such partial structures occur. For example, in urine of Gaucher’s disease the presence has been demonstrated of Man_5 , lacking the Man residues D_1 , D_2 , D_3 and C ; Man_4 lacking the Man residues D_1 , D_2 , D_3 , C and $\mathbf{4}$, and Man_2 lacking the Man residues D_1 , D_2 , D_3 , C , $\mathbf{4}$, A and B .³⁰⁾ In all cases the reporter group signals allow the complete assignment of structures. The partial structures arise from the catabolic route and are helpful in reconstructing the pathway of the stepwise degradation. It should be noted that also partial structures are present with $\text{Fuc}(\alpha\text{-}1\text{-}6)$ linked to $\text{GlcNAc-}1$, probably derived from lactosamine-type glycans.

The precursor oligosaccharide in *N*-glycan biosynthesis. From the precursor oligosaccharide $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ (see Fig. 6), which is transferred in the biosynthesis to the growing protein chain, also partial structures can be found. Analysis of the glycans of glycoproteins isolated from the ovary of the starfish *Asteria rubens* (L.) shows the presence of the $\text{Man}_9\text{GlcNAc}_2$, and $\text{Man}_8\text{GlcNAc}_2$ structure. The latter missing specifically Man-D_3 as is evidenced by the absence of the corresponding structural reporter group signals. Furthermore, gluco-oligosaccharide compounds could be demonstrated,

Fig. 6. $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ structure, the consecutive points of degradation by glucosidase I and II, respectively, are indicated by arrows.

consisting of 9 Man residues and 1-3 Glc residues.³¹⁾ The monogluco-oligosaccharide only carrying Glc_a , is characterized by the structural reporter groups of the Man_9 moiety in conjunction with the anomeric proton of Glc. Comparison of the NMR data with those of the Man_9 structure shows that only the H-1 and H-2 signals of Man D_1 are affected by the presence of Glc indicating that Glc is attached to Man D_1 . Furthermore, di- and tri-gluco-oligosaccharide structures occur, characterized by the anomeric signals of the Glc residues. The assignment of these signals was confirmed by comparison of the synthetic oligosaccharide $\text{Glc}(\alpha\text{-}1\text{-}2)\text{Glc}(\alpha\text{-}1\text{-}3)\text{-Glc}(\alpha\text{-}1\text{-}3)\text{Man}(\alpha\text{-}1\text{-}2)\text{Man}$.³²⁾ This identification of a unique collection of glycans provided insight in the metabolic route starting with $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$. A further example of such a structure was obtained from the lipid-linked precursor oligosaccharide isolated from porcine thyroid tissue. The oligosaccharide was released from dolichol pyrophosphate by mild acid hydrolysis, studied by ^1H -NMR and the structure compared to that deduced for other organisms.³¹⁾ The reducing *N,N*-diacetylchitobiose unit and nine mannose residues are present, the latter in equimolar amounts.³³⁾ The chemical shifts are very similar to those we derived for $\text{Man}_9\text{GlcNAc}_2\text{Asn}$. The three additional anomeric signals are assigned to Glc residues. The values correspond very well with those of the synthetic oligosaccharide $\text{Glc}(\alpha\text{-}1\text{-}2)\text{Glc}(\alpha\text{-}1\text{-}3)\text{-Glc}(\alpha\text{-}1\text{-}3)\text{Man}(\alpha\text{-}1\text{-}2)\text{Man}$.³²⁾ In addition to the NMR evidence, the branch location of the trigluco-syl unit could be elegantly demonstrated after treatment with jack bean α -mannosidase. From the removal of the Man A , D_2 , B , and D_3 residues and the identification of the resulting $\text{Glc}_3\text{Man}_5\text{GlcNAc}(\beta\text{-}1\text{-}4)\text{-GlcNAc}\alpha/\beta$ it is evident that the trigluco-syl unit is attached to the Man-($\text{D}_1\text{-C-}4$) branch. This precursor oligosaccharide is a universal compound in the eukaryotic kingdom. In the regular biosynthesis, the precursor trigluco-oligosaccharide is stepwise degraded after transfer to the protein. This process starts with

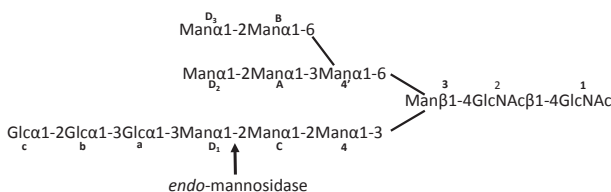


Fig. 7. Formation of the tetrasaccharide Glc_3Man as a result of cleavage by *endomannosidase* of the $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ glycan. Point of cleavage is indicated.

the removal of the Glc residues by glucosidase-I of $\text{Glc}(\alpha 1-2)$ and then of the two $\text{Glc}(\alpha 1-3)$ residues by glucosidase-II. Interestingly, in case of a defect in N-glycosylation, due to glucosidase I deficiency, we found in the urine the oligosaccharide $\text{Glc}(\alpha 1-2)\text{Glc}(\alpha 1-3)\text{Glc}(\alpha 1-3)\text{Man}(\alpha/\beta)$. This glycan arises from the alternative pathway by the action of *endo- α -1,2-mannosidase* as shown in Fig. 7.³⁴⁾ The NMR data of fragments of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ are compiled in Table 2 of ref. 23.

Di- to tetra-antennary complex type N-glycans in $\alpha 1$ -acid glycoprotein. The structure elucidation of the asialo-glycans of $\alpha 1$ -acid glycoprotein was among the first glycoprotein projects we started. In this study, NMR spectroscopy resulted in a breakthrough by affording unambiguous structures. In combination with methylation analysis, the glycans were identified as di-, tri-, and tetra-antennary lactosamine type compounds.^{35)–39)} Novel was also the observation that tetra-antennary glycans occur with Fuc($\alpha 1-3$) linked to an external GlcNAc. The major constituent contains Fuc($\alpha 1-3$) attached to GlcNAc 7 of the tetra-antenna, yielding a Le^x epitope. In addition, two minor compounds are present in the tetra-antennary fraction having Fuc($\alpha 1-3$) linked to one of the external GlcNAc residues in the ($\alpha 1-6$) antenna;⁴⁰⁾ see Fig. 8.

The di- to tetra-antennary chains are unevenly distributed over the five potential glycosylation sites. Terminal sialic acid can be linked in ($\alpha 2-3$) or ($\alpha 2-6$) linkage and can be assigned on guidance of the structural reporter groups H-3_{ax} and H-3_{eq} . Attachment of $\text{Sia}(\alpha 2-3)$ to the Le^x determinant affords SiaLe^x .

Penta-antennary N-glycan in hen ovomucoid. The hen ovomucoid exhibits a large heterogeneity in its N-glycans. In addition to a series of more frequently occurring glycans, a novel bisected penta-antennary structure, which could be identified by $^1\text{H-NMR}$, in combination with methylation analysis and partial acid hydrolysis.⁴¹⁾ The Man-H-1 and

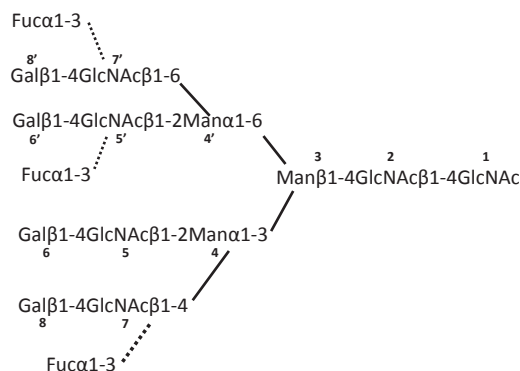


Fig. 8. Tetra-antennary compounds, extended with a Fuc residue at alternative positions as indicated by dotted lines. In the major compound Fuc is linked to GlcNAc 7.

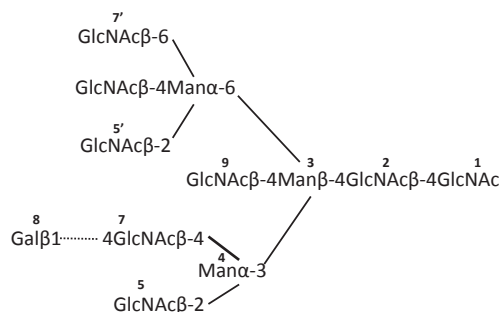


Fig. 9. A penta-antennary structure, which can be extended by a Gal residue at GlcNAc 7 as indicated by a dotted linkage.

H-2 and NAc chemical shifts are typical for the bisected structure. The tri-substitution of Man 4' comes further to expression in the chemical shift of H-4 of Man 4'. Subsequently, we observed that GlcNAc 7 could be extended with Gal($\beta 1-4$) as followed from the chemical shift of the Gal-anomeric proton and from the two significant changes in the GlcNAc reporter group signals in comparison to the agalacto compound.⁴²⁾ The structures are shown in Fig. 9.

Trisialyl di-antennary N-glycan in rat plasma hemopexin. Glycopeptides obtained by pronase digestion of rat plasma hemopexin contain in addition to mono- and di-sialyl diantennary constituents, a trisialyl diantennary compound, bearing three Neu5Ac residues. The localization of the sialic acid residues is as follows: To Gal 6 Neu5Ac($\alpha 2-3$) is linked and to both GlcNAc 5 and Gal 6' Neu5Ac($\alpha 2-6$) is attached. Gal 6 is attached in a ($\beta 1-3$) linkage to GlcNAc 5 as shown in Fig. 10. In the $^1\text{H-NMR}$ spectra, the reporter group signals of the three Neu5Ac residues are indicative.⁴³⁾ Interestingly, by methylation analysis it could be

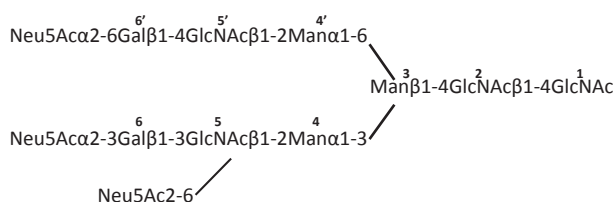


Fig. 10. Structure of a trisialyl diantennary compound as found in the *N*-glycan of rat hemopexin.

shown that also tiny amounts of Neu5Ac(α 2-8) are present. The structures are not unique, but also present in other glycoproteins. However, in this study for the first time the NMR parameters were documented.⁴³⁾

***O*-Acetylated sialic acids in equine fibrinogen.** The glycans of equine fibrinogen consist of mono- and disialo diantennary lactosamine type of structures in a molar ratio of approximately 2:3. The (α 2-6) linked *N*-acetylneuraminic acids are partially *O*-acetylated at C-4. The ratio of Neu5Ac to Neu4,5Ac₂ is about 3:2. The mono-sialo compounds may in the (1-6) arm lack Neu5Ac or Neu4,5Ac₂ and even Gal **6'**. 4-*O*-acetylation of sialic acid introduces a large additional heterogeneity. In the disialo compounds the heterogeneity is even larger, since in both antennae the partially *O*-acetylated sialic acids occur, as presented in Fig. 11. This sialic acid residue is identified by its structural reporter groups namely, H-3_{ax}, H-3_{eq}, H-4, H-5, NAc, and OAc.⁴⁴⁾ This 4-

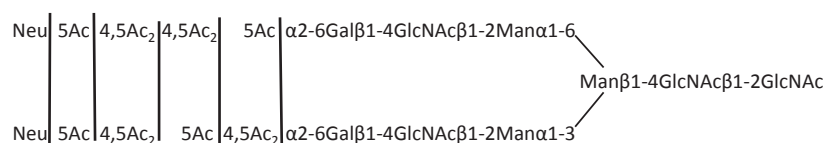


Fig. 11. 4-*O*-acetylated sialic acid in glycans of equine fibrinogen. Note that the composite structure indicates an impressive heterogeneity resulting from 4-*O*-acetylation of sialic acid and from the presence/absence of sialic acid as well as of Gal **6'**. By lines the various combinations of terminal residues are depicted.

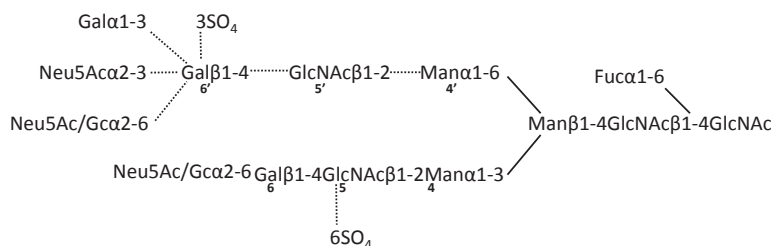


Fig. 12. A composite structure of the glycans occurring in porcine thyroglobulin showing fucosylated diantennary compounds. In the Man α 1-6 branch the heterogeneity at Gal **6'** is indicated by dotted lines. The antenna may also end on GlcNAc **5'** or Man **4'**. In the Man α 1-3 branch the presence or absence of a SO₄ group and the occurrence of *N*-glycolyl or *N*-acetyl sialic acid are other sources of heterogeneity.

O-acetylation is a characteristic feature of equine glycoproteins, but later studies showed that this substitution is not unique.

Gal(α 1-3) and NGc epitopes as well as 3-*O*-SO₄Gal-6' and 6-*O*-SO₄GlcNAc-5 in porcine thyroglobulin. The major acidic compounds of porcine thyroglobulin are mono- or disialylated, fucosylated diantennary compounds. The Man(α 1-6) antenna shows an impressive heterogeneity since it can have as terminus Man-4', GlcNAc-5' or Gal-6'. Furthermore, Gal-6' can be extended with Neu5Ac(α 2-3), with sulfate at C-3 or with the Gal(α 1-3) epitope. At Gal-6' Neu5Ac or Neu5Gc can be attached in (α 2-6) linkage, as shown in Fig. 12. The Gal(α 1-3) as well as the *N*-glycolylsialic acid entities have characteristic structural reporter groups.²³⁾ Importantly, both epitopes are not compatible with the human system. This is relevant for biologics, biosimilars and eventual xenotransplantation. A further heterogeneity stems from the presence or absence of sulfate attached to C6 of GlcNAc **5** in the Man(α 1-3) antenna. As mentioned before, the extension with sulfate gives in general rise to a downfield shift of the geminal and vicinal proton signals. In a minor fraction compounds containing tri-sialylated, fucosylated tri-antennary structures are present.⁴⁵⁾

4-*O*-SO₄GalNAc in human urokinase. Urinary-type plasminogen activator (u-PA) is a serine protease, playing important biological roles. The

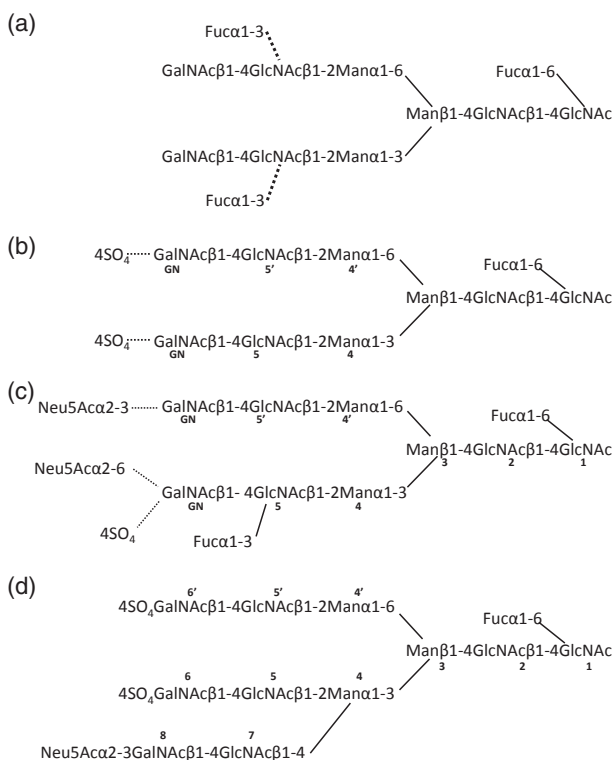


Fig. 13. (a) Fucosylated diantennary glycan carrying GalNAc β 1-4[Fuc α 1-3]GlcNAc as novel epitope in urokinase. By the dotted lines the partial presence is shown. (b) Sulfated glycan found in urokinase. 4-SO₄GalNAc has besides in urokinase, also been identified in a number of other glycoproteins. By the dotted lines the partial presence is shown. (c) The acidic sialylated, sulfated fraction in urokinase exhibits a considerable heterogeneity as summarised in this composite structure. By dotted lines possible linkages are indicated. (d) Triantennary compound in the acidic glycan fraction from urokinase, carrying sialic acid at GalNAc-8 and 4-SO₄GalNAc at the other termini.

protein has a single *N*-glycosylation site at Asn-302, which exhibits a large heterogeneity and an *O*-linked Fuc at Thr-18 in the epidermal growth factor domain.⁴⁶⁾ The *N*-glycans are (α 1-6)-fucosylated diantennary complex structures, wherein GalNAc has replaced Gal in both antennae. We observed that either GlcNAc 5 or GlcNAc 5' can be substituted with Fuc(α 1-3), yielding GalNAc(β 1-4)[Fuc(α 1-3)]GlcNAc(β 1-2) as novel structural element as depicted in Fig. 13a.⁴⁷⁾ In the acidic *N*-glycans, GalNAc in one branch or in both branches may be 4-*O*-sulfated, as shown in Fig. 13b. To a minor extent, the (α 1-3) antenna can be capped with Neu5Ac(α 2-6). Sialylation in (α 2-3) linkage can occur in the (α 1-6) branch, in conjunction with 4-*O*-sulfation of GalNAc in the (α 1-3) antenna. The major glycan is a (α 1-6)fucosylated, diantennary

structure with 4-SO₄GalNAc(β 1-4)GlcNAc(β 1-2) termini. A composite structure gives a summary in Fig. 13c.

Also fully sulfated tri'-antennary compounds are present as well as a tri-antennary compound carrying 4-*O*-sulfate at GalNAc 6 and 6' in combination with Neu5Ac(α 2-3) at GalNAc 8, as shown in Fig. 13d.⁴⁸⁾

HNK1 epitope in P0. P0 glycoprotein is the most abundant protein constituent of peripheral myelin. It consists of a single extracellular immunoglobulin-like domain, a transmembrane part and a cytoplasmic tail. P0 appears at the initial stage of myelination and functions in the formation and maintenance of myelin compaction as an adhesion molecule. Severe neurological disorders are associated with mutations in the P0 gene. P0 isolated from bovine sciatic nerves has a single *N*-glycosylation site. P0 is the first glycoprotein wherein the HNK1 epitope consisting of 3-*O*-SO₄GlcA(β 1-3)Gal(β 1-4)GlcNAc(β 1-) could be established by ¹H-NMR.⁴⁹⁾ We could further show that this glycosylation site exhibits a formidable heterogeneity. The major part of the glycan structures could be identified. Basically, the glycans can be conceived as hybrid type or as fucosylated, diantennary complex compounds, with or without a bisecting GlcNAc. The amazing diversity stems from the non-reducing termini attached to the Man(α 1-3) and Man(α 1-6) residues. At the Man(α 1-3) branch GlcNAc 5 can be partially sulfated at C-6. Further extension may give rise to 6-*O*-sulfo HNK1 and 6-*O*-sulfo sialyl Le^x as remarkable elements. Alternatively, Gal 6 may be extended with Neu5Ac(α 2-3), Neu5Gc(α 2-3), Neu5Ac(α 2-8)Neu5Ac(α 2-3) or Gal(α 1-3)Gal(β 1-4)6-*O*-SO₄-GlcNAc. The epitope Gal(α 1-3)Gal(β 1-4)6-*O*-SO₄-GlcNAc was here observed for the first time. The same holds for Neu5Ac(α 2-8)Neu5Ac(α 2-3)Gal(β 1-4)6-*O*-SO₄-GlcNAc.

The Man(α 1-6) antenna can be extended with Man(α 1-3), [Man(α 1-6)]Man(α 1-3), GlcNAc(β 1-2) or HNK-1. In a composite formula shown in Fig. 14, the various termini are summarized in rectangles. The characterization of these structures required the combination of advanced NMR and MS techniques due to the small amounts of material and the complexity of the structures.^{49),50)} The glycan moiety of P0 plays a role in cell-cell adhesion, possibly via homophilic binding. At least one of the glycoforms is important for the function of the molecule, it may be a HNK-1 carrying epitope. P0 exposes appropriate glycans for recognition, including its own glycans. The precise molecular mechanism of this auto-recognition remains to be established. The biological

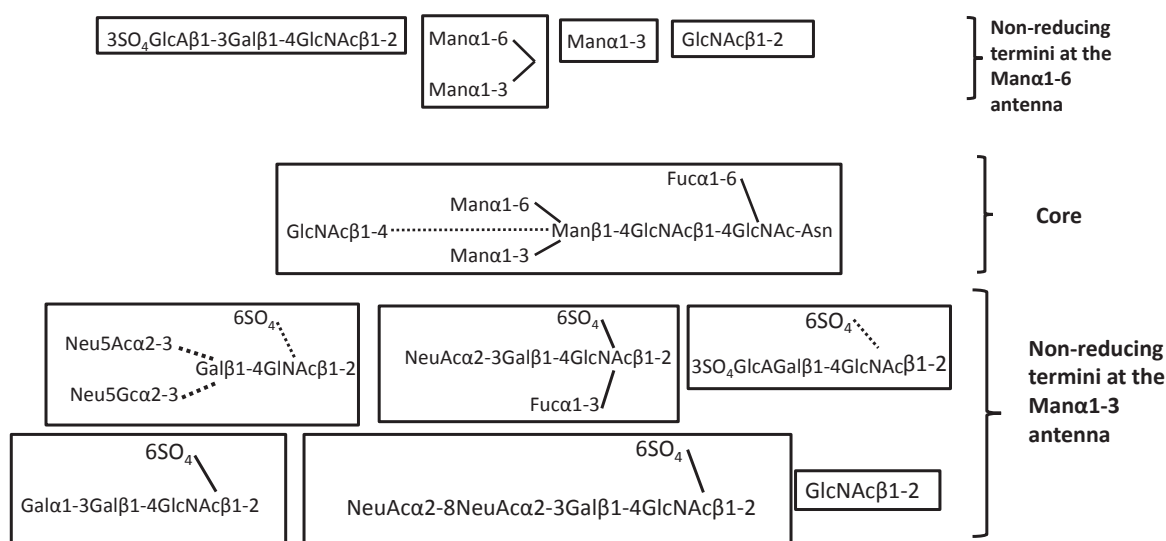


Fig. 14. In this composite structure the enormous heterogeneity is shown for the glycans of P0, attached to the single *N*-glycosylation site. In the various rectangles the structural elements are indicated that are identified at the antennae as possible extensions of the core structure. The dotted lines indicate that these groups can be present or absent. The different possible combinations of elements constitute the actual basis for the observed formidable heterogeneity.

significance of the diversity and of the individual epitopes remain intriguing questions.

Sda determinant in Tamm Horsfall Glycoprotein. The Tamm Horsfall Glycoprotein was first recognized as a urinary mucoprotein.⁵¹⁾ It is now characterized as the most abundant glycoprotein in normal urine. The excretion level may be as high as 100 mg/day. When isolated from the urine of pregnant women the glycoprotein is called uromodulin, but it has still the same amino acid sequence.⁵²⁾ The glycoprotein is expressed in the thick ascending limbs of the loop of Henle and in the early distal convoluted tube of the nephron in the kidney.⁵³⁾ It is a membrane bound, GPI anchored glycoprotein from which the signal peptide of 24 amino acids and the propeptide of 26 amino acids have been split off.⁵⁴⁾ By proteolysis it is cleaved from the membrane. The protein consists of three EGF domains, one DC8 and one ZP domain as shown in Fig. 15. It is heavily *N*-glycosylated up to 25–30% of the molecular mass. The protein is difficult to purify due to its gel-like behavior and the easy polymerization to filaments or matrices. These two features are ascribed to the ZP domain.⁵⁵⁾ The glycosylation pattern of this glycoprotein is highly complex as we showed in a few studies.^{56),57)} The protein obtained from the urine of a single male donor, contains 8 consensus sequences for *N*-glycosylation. Counting from amino acid residue 1 in the secreted form, Asn 14 is not glycosylated in contrast to the sites at Asn 52, 56, 208, 251, 298, 372

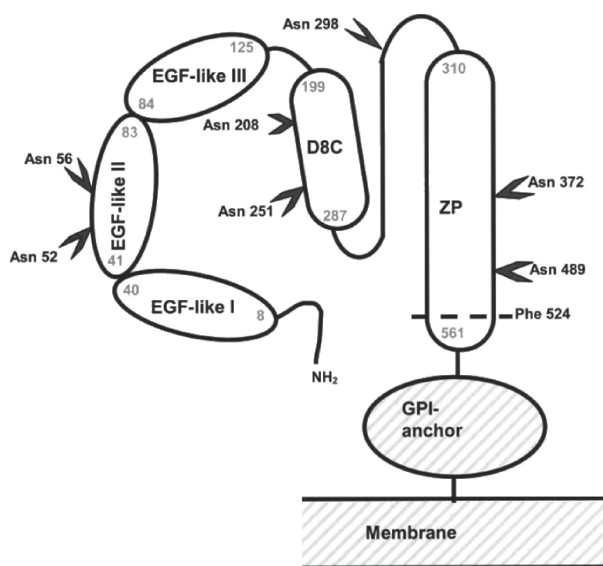


Fig. 15. The backbone structure is shown of the protein chain of Tamm-Horsfall glycoprotein. The different protein domains and the *N*-glycosylation sites are indicated. The protein is connected to the membrane via a GPI-anchor.

and 489. Oligomannose type chains, ranging from $Man_5GlcNAc_2$ to $Man_8GlcNAc_2$ and representing about 5% of the total carbohydrate content are only occurring at glycosylation site 251.⁵⁸⁾ We obtained more than 150 glycan-containing fractions, representing the combined (micro)heterogeneity of the complex glycosylation sites. By ¹H NMR analysis

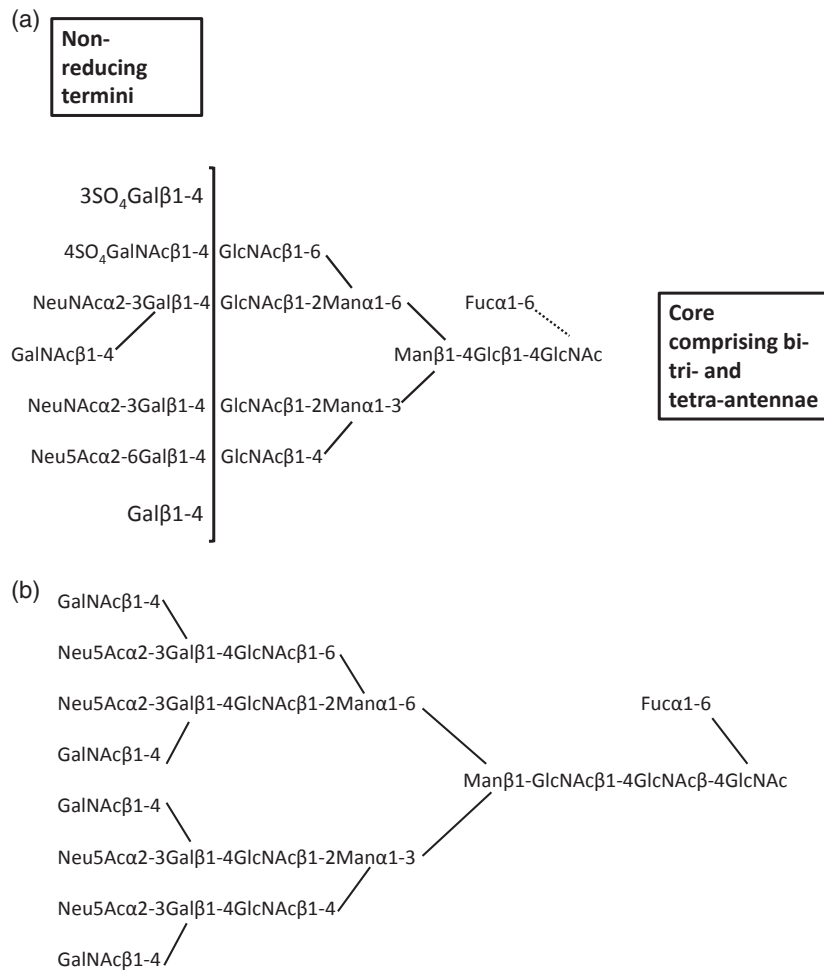


Fig. 16. (a) The identified non-reducing termini found in Glycans from Tamm-Horsfall glycoprotein are listed. The core starts with the common pentasaccharide unit that can be extended to di-, tri- and tetraantennae. To GlcNAc-1 can be attached Fuca1-6. (b) One of the most complicated tetraantennary structures in the glycan of Tamm Horsfall glycoprotein is presented. At the termini Sd^a determinants occur.

of the isolated glycans we established the structure of compounds ranging from a nonfucosylated, mono-sialylated diantennary structure to sialylated, di-, tri- and tetra-antennary compounds containing fucose and sulfate groups as well as motifs like additional *N*-acetyllactosamine units and terminal Sd^a epitopes. The sulfates occur as 3-*O*-sulfated Gal and as 4-*O*-sulfated GalNAc in mono- to trisulfated *N*-glycans. The results can hardly be summarized in a single composite structure. In Fig. 16a a presentation is given for part of the structures in terms of non-reducing termini and core structures. In Fig. 16b one of the most complicated structures is shown.

The occurrence of donor-specificity in glycan structures of the Tamm Horsfall glycoprotein, could be demonstrated for *N*-acetyllactosamine units,

oligomannosyl components, and Sd^a epitopes, respectively. We studied this in detail for the Sd^a determinant: NeuAc(α 2-3)[GalNAc(β 1-4)] Gal(β 1-4)GlcNAc(β 1-3)Gal as follows. This determinant can be released by *endo*- β -galactosidase digestion of the glycoprotein, in conjunction with the formation of the tetrasaccharide Neu5NAc(α 2-3)Gal(β 1-4)GlcNAc(β 1-3)Gal and trisaccharide Gal(β 1-4)GlcNAc(β 1-3)Gal. By analysis of THp from 4 unrelated healthy male donors we have shown that the molar ratio of released Sd^a pentasaccharide to tetrasaccharide and thereby the Sd^a-related glycosylation is donor-specific. The donor-specificity holds also for the total content of the Sd^a pentasaccharide plus tetrasaccharide.⁵⁹⁾ The question arose whether in genetically identical individuals the donor-specificity

would be the same. In fact, a study of THp from two monozygotic pairs of twins showed the qualitative and quantitative identity of the Sd^a content, thereby providing additional evidence for the donor-specificity.⁶⁰⁾

The THp isolated from pregnant women is designated uromodulin in view of the presumed immunomodulatory properties.⁵²⁾ Since the amino acid sequence has not been altered, it is obvious to ascribe eventual changes in properties to differences in glycan structures. We investigated the glycan structures for uromodulin from 3 pregnant women at different stages of pregnancy. For the *N*-glycans the negative charge distribution of sialic acid and sulfate remained virtually constant. By consequence it was presumed that the branching pattern remained unaltered. Pregnancy leads to slight changes in type of oligo-mannosyl structures.⁶¹⁾ However, it is not certain that these differences arise from pregnancies, because the donor specificity can already account for significant variation of the content of oligo-mannose structures. In one publication, the occurrence of *O*-linked glycans was reported.⁶²⁾ These *O*-linked compounds should undergo changes in structure during pregnancy. We could not confirm the occurrence of *O*-linked glycans.^{55)–57)} (Rohfritsch, Ph.H. *et al.* unpublished). Furthermore, we showed that for our preparations the immunosuppressive activity of the glycan pools was invariant during pregnancy.

Glycans occurring on human epidermal growth factor receptor. Many details are known of the epidermal growth factor receptor (EGFR). EGFR is a transmembrane glycoprotein with 11 potential *N*-glycosylation sites. The *N*-glycosylation is essential for its biological function. Here we could focus for the first time on the structure of the glycans of human non-recombinant EGFR. The human epidermoid carcinoma cell line A431 secretes the extracellular domain of EGFR as a soluble 105-kDa glycoprotein. We investigated the structure of the *N*-glycans of the secreted glycoprotein by NMR and MS. The glycans, released by PNGase F show a large heterogeneity. The oligomannose constituents vary from Man₈GlcNAc₂ to Man₅GlcNAc₂ and account for about 17% of the glycans, but Man₇GlcNAc₂ and Man₆GlcNAc₂ are the main compounds.

The complex carbohydrates comprise di-, tri-, and tetra-antennary glycans, both neutral and (α 2-3)sialylated compounds, representing 24% and 59%, respectively, of the total glycan pool. The antennae can carry A, H, Le^x, Le^y, ALe^y or sialyl-Le^x blood-group antigens. From the enormous heterogeneity 55

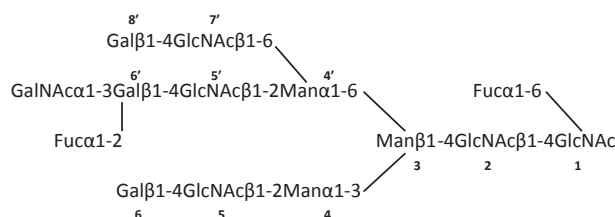


Fig. 17. Out of the large number of *N*-glycan-structures attached to sEGFR, an example is given of a tri'-antennary compound containing a bloodgroup A epitope.

glycan structures were assigned and 32 as novel.^{63),64)} In Fig. 17 a representative tri'-antennary structure is presented, carrying the bloodgroup A epitope.

Two types of Fuc attached to GlcNAc-1 in honeybee venom phospholipase A₂. The core structure of the *N*-glycans of honeybee venom phospholipase A₂ can show the occurrence of both Fuc(α 1-3) and Fuc(α 1-6) at the Asn-bound GlcNAc. This was deduced from the structural reporter group signals of Fuc, which are characteristic for these types of linkages. The presence or absence of Man(α 1-3) and of the Fuc residues are sources of heterogeneity as presented by dotted lines in Fig. 18a. Also the oligomannose units without Fuc demonstrate some variability as shown in Fig. 18b. Interestingly, in complex type of structures the Man(α 1-3) branch has as terminal structure GalNAc(β 1-4)[Fuc(α 1-3)]GlcNAc.^{65),66)} This is depicted in Fig. 18c.

Bromelain. Pineapple stem Bromelain has only a single *N*-glycan and exhibits no micro-heterogeneity as determined by Ishihara *et al.*⁶⁷⁾ The *N*-linked glycan of this glycoprotein contains Xyl linked to the branching Man of the core structure, furthermore Fuc(α 1-3) is linked to the Asn-bound GlcNAc as given in Fig. 19.

The NMR parameters showed that this Fuc(α 1-3) could easily be recognised, since the structural reporter groups for Fuc(α 1-3) and Fuc(α 1-6) differ significantly and are characteristic for each type. This concerns the chemical shifts of Fuc H-1, H-5 and CH₃, as well as of the neighbouring GlcNAc signals. Xyl(β 1-2) has its own characteristic reporter signals H-1, H-2 and H-5_{ax} and H-5_{eq}.⁶⁸⁾ In a further study we established also the conformation of the glycan.⁶⁹⁾

Glycans derived from hemocyanins. Hemocyanin is the Cu(I)-containing oxygen transporter glycoprotein of most arthropod and mollusc species. The glycans of the hemocyanins are quite interesting, due to their relatively uncommon structures. We reviewed this group of glycans in ref. 70.

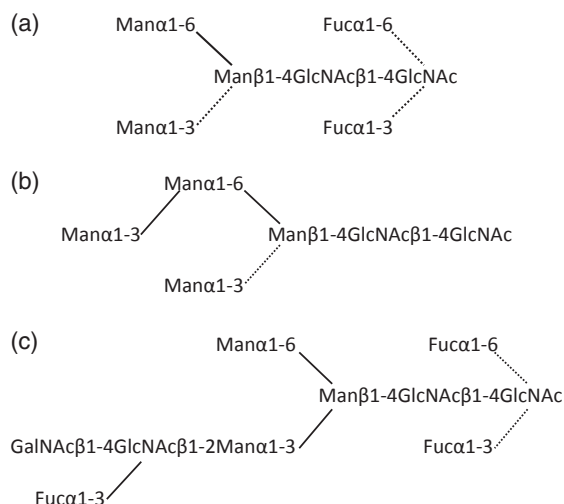


Fig. 18. (a) Difucosyl glycans from honey bee phospholipase A_2 with heterogeneity as depicted by dotted lines. (b) Oligomannose type structures of honey bee phospholipase A_2 without Fuc. (c) Complex type glycan of honey bee phospholipase A_2 with core Fuc residues, occurring only partially as indicated by dotted lines. Thereby a significant heterogeneity is introduced.

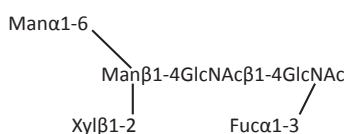


Fig. 19. The Glycan structure of pine apple stem bromelain.

3-O-methyl sugars in hemocyanin. In a mass spectrometric study we could demonstrate that hemocyanin from *Helix pomatia* contains 3-O-methylgalactose in addition to Fuc, Xyl, GalNAc and GlcNAc.

However, in hemocyanin from *Lymnea stagnalis* the occurrence of 3-O-methylmannose was established.⁷¹⁾

Hemocyanin of the mollusc *Lymnea stagnalis*. Hemocyanin-glycan of *Lymnea stagnalis* has a core structure with Xyl(β 1-2) linked to the branching Man and 3-O-Me-Man at the non-reducing termini, but without fucosylation as shown in Fig. 20a.⁷²⁾ Alternatively, the Man residues can be extended to a mono- or di-antenna with 3-OMe-Gal(β 1-3)GalNAc(β 1-4)GlcNAc(β 1-2). In turn GalNAc can be substituted with 3-OMe-Gal(β 1-3) or with Fuc(α 1-2)Gal(β 1-3). In Fig. 20b an example of such a structure is presented. Further extensions of 3-OMe-Man or 3/4-OMe-Gal have not been observed, suggesting that here methylation could act as a kind of stop signal.⁷³⁾

Hemocyanin of *Helix Pomatia*. In a further and detailed investigation of hemocyanin from *Helix*

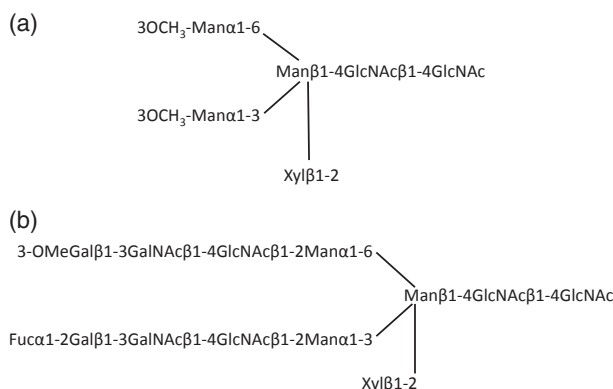


Fig. 20. (a) 3-O-MeMan in glycans of hemocyanin from *Lymnea stagnalis*. (b) Extended antennae in the glycans of hemocyanin from *Lymnea stagnalis*.

Pomatia we could establish the localization of 3-O-methylgalactose. The *N*-glycans obtained by PNGase F degradation have in common the pentasaccharide core largely substituted with Xyl(β 1-2) at the branching mannose residue and for the greater part with Fuc(α 1-6) at the Asn-linked GlcNAc. Structures lacking Xyl, miss also Fuc. Four novel antennae and in total 21 novel mono-antennary and di-antennary *N*-linked glycans were identified. Most structures contain 3-OMe-Gal β and occasionally 4-OMe-Gal β . The antennae are linked to one or both of the core Man residues.⁷⁴⁾

One of the most complex structures is presented in Fig. 21.

Hemocyanin of *Panulirus interruptus*. It should be noted that not all hemocyanins are xylosylated, for example hemocyanin of the spiny lobster *Panulirus interruptus* contains mainly neutral *N*-glycans like oligomannose structures. The pentasaccharide core can also be extended with GlcNAc(β 1-2) and to Man-4 sulfate can be attached at position 6, as shown by a dotted line in Fig. 22.^{75);76)}

General features of O-glycoproteins and the corresponding $^1\text{H-NMR}$ data. The *O*-glycans belonging to the mucin type of chain are attached to the protein through GalNAc α (1-*O*) to the hydroxyl groups of Ser or Thr. These glycans have in common GalNAc as starting point that may be extended at C-3 and/or C-6 by Gal, GalNAc or GlcNAc and at C-6 also by Neu5Ac/5Gc. The substitutions give rise to 8 different types of core structure. We analysed the *O*-glycans as alditols, obtained after alkaline borohydride reduction of the *O*-glycoprotein. The combination of the chemical shift of H-2 and of H-5 of the GalNAc-ol is an excellent monitor of all

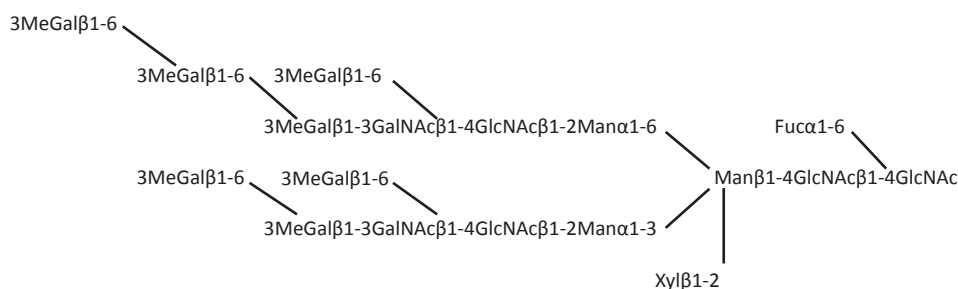


Fig. 21. One the most complicated glycan structures of hemocyanin from *Helix Pomatia*.

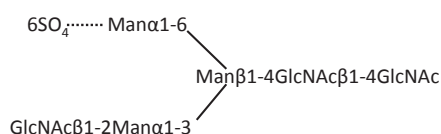


Fig. 22. Acidic glycan of hemocyanin from *Panulirus inter-ruptus*.

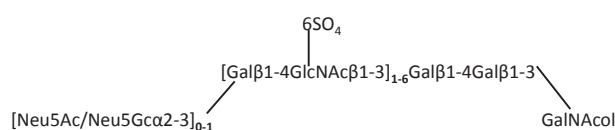


Fig. 24. Acidic *O*-glycans derived from porcine *zona pellucida* glycoproteins. The repeating lactosamine units in the compounds are shown.

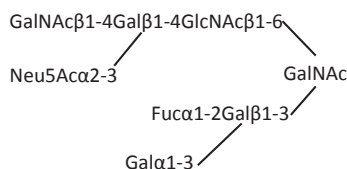


Fig. 23. Structure of a glycan from cervical mucus containing the Sd^a and bloodgroup A epitopes.

core types. In a comprehensive review, we summarized the chemical shifts of a large series of *O*-glycans.²²⁾

Also the structure of a number of functionally important *O*-glycans, other than of the mucin type, has been identified. In these cases we characterized the structures by ¹H NMR. However, the collection of NMR data is too limited to indicate useful reference ranges of chemical shifts.

Bloodgroup activities in cervical mucus glycoproteins from Bonnet monkey (*Maccaca radiata*). The mucus glycoproteins from pooled midcycle cervical epithelial secretion of the Bonnet monkey contain multiple bloodgroup activities. Therefore, the structure of the *O*-linked glycans was analysed. After treatment of the purified glycoprotein fraction with alkaline borohydride the sialyl-oligosaccharide alditols were isolated and fractionated. Interestingly, the glycans contain structure elements corresponding to either A, B or H activities, which may occur in combination with the Sd^a epitope, as shown in Fig. 23.⁷⁷⁾ An obvious explanation would be that this heterogeneity of blood group determinants stems from the pooling of the mucus samples.

N, *O*-glycoproteins

Poly-*N*-acetyllactosamine-containing *O*-glycans of porcine *zona pellucida* glycoproteins. As mentioned before in a glycoprotein *N*-glycans can occur besides *O*-glycans. For the analysis of the glycan structures often the *N*-glycans are detached by PNGase followed by cleavage of the *O*-glycans from the purified *O*-glycoprotein. A summary of our procedures for *N*, *O*-glycoprotein-analysis has been presented in ref. 78.

The porcine *zona pellucida* ZP3 family of glycoproteins was isolated from porcine oocytes. In particular, the acidic *O*-glycans of these *N*-, *O*-glycoproteins are interesting because of the occurrence of repeating units of Gal(β1-4)[6SO₄]GlcNAc. For 32 *O*-glycans the structure was determined. Most of the main structures belong to a family of compounds, summarized in a composite formula Fig. 24.^{79),80)}

Typical for this series of compounds are the chemical shifts of Gal H-1 and Gal H-4 in the various lactosamine units as compiled in Tables 1–5 of ref. 80. However, in addition, several minor structures are present.

The structure of the neutral *N*-glycans have been studied by Kobata *et al.*⁸¹⁾

Poly-*N*-acetyllactosamine-containing *O*-glycans of equine chorionic gonadotropin (eCG). Equine chorionic gonadotropin belongs to the class of heterodimeric (α and β subunits) glycoproteins. In many respects the hormone behaves very similar

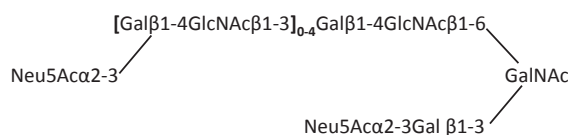


Fig. 25. Structures of *O*-glycans derived from equine chorionic gonadotropin showing various numbers of repeating lactosamine units.

as other gonadotropins, but interestingly, it has from all glycoprotein hormones the highest carbohydrate content. For analysis *N*- and *O*-glycans were split off. After isolation and purification, the released glycans were identified by $^1\text{H-NMR}$ and FAB MS. The *N*-glycosylation of the β -subunit exhibits glycans of considerable heterogeneity. They comprise mono-, di-, tri- and tri'-antennary structures. In part of the compounds the Asn linked GlcNAc is substituted with Fuc(α 1-6). Neu5Ac(α 2-6) is the main terminal residue in the antennae, most pronounced in the diantennary compounds. Overall, a significant portion of Neu5Ac is 4-*O*-acetylated. Interestingly, partially 4-*O*-acetylated Neu5Ac(α 2-3), occurs only in tri- and tri'-antennary compounds.

The *O*-glycans are mostly present in the β -subunit and they consist predominantly of tri-, tetra-, penta- and hexa-saccharides of core type 2. A small part of these glycans contain repeating *N*-acetyl-lactosamine units.^{82),83)} These structures are summarized in Fig. 25. It can be anticipated that part of the Neu5Ac residues in *O*-glycans is 4-*O*-acetylated, in analogy to the *N*-linked glycans. However, due to the alkaline borohydride treatment this information gets lost.

Recombinant glycoproteins. The demand for human glycoproteins for diagnostic and therapeutic purposes is still increasing. This has stimulated the production of recombinant glycoproteins since the natural glycoproteins are hardly accessible in large quantities. However, expression in heterologous cells leads to the glycosylation pattern of the corresponding cell type. Comparison of the structures of these glycans with those of the human system is essential to gain insight in the changes in properties and compatibility with the human immune system. The introduction of *N*-glycolylneuraminic acid or the Galili epitope Gal(α 1-3)Gal(β 1-3/4)GlcNAc-R, are obvious examples of unfavourable epitopes that should be avoided or eliminated to prevent unwanted reactions. These epitopes can be recognized by their $^1\text{H NMR}$ data.²³⁾ Nevertheless, some commercial preparations are available containing to some extent non-human epitopes.

From a significant number of recombinant glycoproteins, we have established the glycan structures.⁸⁴⁾ Two examples will be presented. The first one we studied was γ -interferon expressed in CHO cells. A regular *N*-linked diantennary glycan was found, partially substituted with Fuc(α 1-6) linked at the Asn bound GlcNAc and partially capped at both antennae with (α 2-3) linked Neu5Ac.⁸⁵⁾ The exclusive occurrence of (α 2-3) linked Neu5Ac is typical for CHO cells, since α 2,6-sialyltransferase is absent in these cells. A more complex glycoprotein is human FSH expressed in CHO cells. We determined the glycan structures mainly by $^1\text{H NMR}$ spectroscopy.⁸⁶⁾ The *N*-linked glycans were identified as monosialylated diantennary, disialylated diantennary (main constituent), disialylated tri-antennary, trisialylated tri-antennary, trisialylated tri'-antennary and tetra-antennary compounds. A small percentage of these chains contains additional *N*-acetyllactosamine residues. Sialic acid is exclusively (α 2-3) linked. Comparison of recombinant and pituitary FSH, shows as significant differences the occurrence in pituitary FSH of (α 2-6) linked sialic acid and bisecting GlcNAc. Importantly, the recombinant product can be pharmaceutically applied and is perfectly compatible with the human system.

C-linked α -mannopyranose in urinary RNase II. The existence in glycoproteins of a C-linkage between carbohydrate and protein was discovered in human urinary RNase II. This enzyme contains a α -mannopyranosyl residue that is C-linked to the C-2 position of the indole nucleus of tryptophan. This was discovered in the glycopeptide PheThr(α -Man p)TrpAlaGlnTrp, derived from RNase II. The C-linkage was proved by ^1H - and $^{13}\text{C-NMR}$ spectroscopy, in conjunction with mass spectrometry as depicted in Fig. 26.^{4),5),87)} As the most simple consensus sequence for C-mannosylation TrpXxxXxxTrp has been established, wherein the first Trp is C-mannosylated.⁸⁸⁾ C-mannosylation occurs in many proteins and is widely distributed in eukaryotic organisms and is accomplished by C-mannosyltransferase⁸⁹⁾ with dolicholphosphate-mannose as donor.⁹⁰⁾ This form of glycosylation is not occurring in bacteria or in yeasts, probably due to the absence of a C-mannosyltransferase. A program has been developed for the prediction of potential C-Man sites.⁹¹⁾ Interestingly, in glycopeptides derived from RNase II the conformation of the Man p ring at the NMR time-scale can only be described by an ensemble of conformations. A main contribution to this ensemble stems from the $^1\text{C}_4$ chair. A further

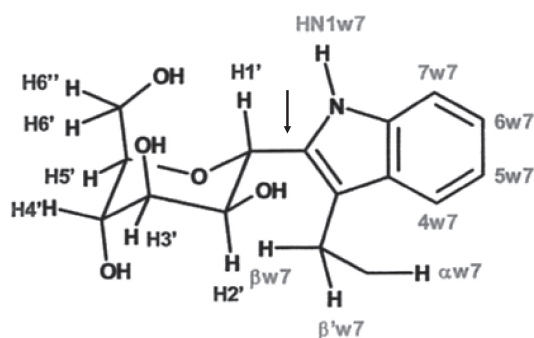


Fig. 26. α -Mannose in the 1C_4 conformation C-linked to the C-2 of tryptophan. The shown conformation of the α Man $_p$ -residue contributes to the ensemble of conformations in the native Glycoprotein. Rotation around the C-C-linkage as indicated by an arrow, yields a conformation contributing to the conformational ensemble in glycopeptides and in denatured glycoproteins.

source of flexibility stems from rotation around Trp C²-Man C-1 bond. The orientation of Man around the C-linkage in native RNase 2 is different from that in the glycopeptide and in the denatured protein.^{5),92)} Apparently, the 3-D structure of the native protein affects the conformational features of the C-Man residue. This influence cannot be observed in glycopeptides (Beccati, D. *et al.* unpublished). The biological function of C-mannosylation requires further investigation.

Glycosaminoglycans. Glycosaminoglycans (GAG's) constitute an important class of linear glycan chains *O*-linked to Ser of the protein backbone of proteoglycans. Compounds like heparan, heparan sulfate, chondroitin sulfate, dermatan sulfate and keratan sulfate belong to this class. The functional and biological properties exhibit a large diversity. From a structural point of view the compounds are rather complex despite of the fact that they are only built up from a linkage region consisting of Gal(β 1-3)Gal(β 1-4)Xyl(β 1-O)Ser as shown in Fig. 27. The core can be extended with repeating disaccharides with GlcA or IduA and with GalNAc or GlcNAc as constituents. The repeating units can be partially and/or differently be sulfated. The resulting large diversity in structure stems not only from the differences in constituents and glycosidic linkages, but also from a diversity in chain lengths in conjunction with the distribution of sulfate groups over the chain. In this way, unique stretches occur in the structure that may be responsible for specific biological effects. For

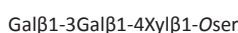


Fig. 27. Linkage region in Glycosaminoglycans.

example, in heparin is a pentasaccharide unit carrier of the anticoagulant properties. In general, the fact that fragments may show a clearly defined activity in bioassays is a complication in defining the biological function in overall terms of the parent structures, because these macromolecules could harbour more than one activity. A further aspect concerns the influence of the neighbouring residues in the polymer on the biological activity of a region in the chain. For the NMR studies we started with the analysis of fragments of the linkage region. Thanks to the availability of synthetic compounds, the structural reporter group signals could excellently be determined.⁹³⁾ As a next step we unravelled by ${}^1\text{H}$ NMR spectroscopy a number of structures in the extended linkage regions of several GAG's, obtained by enzymatic degradation of the macromolecules. At the reducing end the compounds are still linked to Ser and due to the specificity of the used enzymes at the non-reducing end Δ_4 -GlcA is present. After defining unsubstituted hexasaccharides, the extensions with the disaccharides substituted at various positions with sulfate groups are defined by the usual individual reporter group signals and by sulfate-induced downfield chemical shifts. Characteristic signals belong to Δ_4 -GlcA. Interestingly, a novel structural element was discovered in chondroitin sulfate from swarm rat chondrosarcoma containing a 4SO₄GalNAc residue in the extended core region as shown in Fig. 28.⁹⁴⁾

Subsequently, various series of compounds obtained from diverse sources and representing different structures were investigated.⁹⁵⁾⁻⁹⁹⁾ The occurrence of various sulfate groups affects the conformational behaviour of these structures. Discussion of the effects on the conformation is beyond the scope of this review.

Functions. The wide spread decoration of proteins with covalently linked carbohydrates gives rise to fundamental questions about the function of the glycans. First of all, these glycans have a significant influence on the outer sphere of the biomacromolecule and thereby on its contact with the environment. Secondly, biological functions of glycans are often described in global terms like protein folding, protein trafficking, cell signalling,

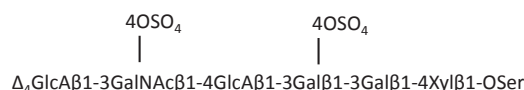


Fig. 28. Extended linkage region as occurring in chondroitin 4-sulfate in swarm rat chondrosarcoma. Novel structural element is 4SO₄Gal at that position.

normal and abnormal cell growth, life-time of cells and glycoproteins, targets for interaction *e.g.*, with pathogens, receptors, antibodies and lectins. For gaining insight in the nature of these functions translation of the problem to the molecular level is essential. This requires detailed knowledge of the three-dimensional structure of the glycoprotein. It should be noted that also a (partially) disordered state is a form of three-dimensional structure, which might be biologically significant, an aspect that has to be taken into account for more than the glycan moiety of glycoproteins, only.^{100)–102)} In most of the discussions on glycoproteins it is neglected that local occurrence of protein random coil structures might be of importance for the function. Next, the molecular details of the interaction of glycoproteins with complementary partner-molecules like receptors and the mutual influence of the respective structures are needed as well. In particular, for membrane bound glycoproteins the effect neighbouring molecules needs consideration. The intermolecular interaction with complementary compounds may concern proteins, carbohydrates, glycoproteins or glycolipids. Carbohydrate-carbohydrate interaction can be of homo- or hetero-type *e.g.*, in cell-cell interactions.¹⁰³⁾ Since the carbohydrate-carbohydrate interactions are much weaker than carbohydrate-protein interactions, polyvalent presentation is needed to be effective.¹⁰⁴⁾ This interaction may require involvement of (part) of the glycan structure, (part) of the protein or comprising (parts) of both entities. A number of techniques including SPR, NMR and X-ray crystallography are available to study aspects of the interaction of the glycan.^{105),106)} Although for biological functions the attention is mainly focused on intermolecular interactions, also the intramolecular interactions should be taken into account. The latter interaction may affect the protein-folding and the three-dimensional structure of the glycoprotein. For example, glycosylation can lead to masking of peptide epitopes as an important feature. The intramolecular interactions may be essential for the presentation and dynamics of the glycans and/or protein-chain. In a *N*-glycoprotein, the interaction of glycans with the protein starts with the Asn-bound GlcNAc. Screening of the protein crystallographic database shows that in X-ray structures this GlcNAc can be oriented in different ways. The orientation can be for example O5-edge, solvent exposed, with α - or with β -face towards the protein. In fact, the GlcNAc-protein interaction has consequences for the conformation and presentation of the outer part of the glycan

structure. An interesting example is provided by the two *N*-glycans of the free α -chain of human chorionic gonadotropin. The *N*-glycan at Asn78 is resistant towards PNGase F, unless the α -chain has been denatured. This observation is in agreement with the ordered structure of the part of the glycan close to the protein chain. The effect is noticeable in the whole glycan and will thereby affect the interaction with complementary molecules. In contrast, the other *N*-glycan at Asn52 can be directly split off by PNGase F, due to the absence of interaction between the Asn-bound GlcNAc and the protein.^{107),108)}

In general, glycoproteins are dynamic entities and may just like other macromolecules consist of an ensemble of conformations. The intrinsic flexibility of glycans enlarges the number of conformations of glycoproteins considerably in comparison to the parent protein. This structural property is relevant in exerting biological functions.

Concluding remarks. This review is centred on our development of high resolution NMR spectroscopy as an important new tool for the structure determination of glycans. As illustrated, we could establish a novel route to identify unambiguously numerous *N*- and *O*-linked glycans of glycoproteins and discovered C-linked mannose as a new type of linkage. A large collection of characteristic chemical shifts was obtained. The structural reporter group concept is robust and can be widely applied for the identification of carbohydrate chains stemming from various sources. The structural reporter group signals are rather insensitive to alterations in the structure remote from the locus involved. Interestingly, the derived principles of glycan identification on guidance of the structural reporter group concept, turned out to be perfectly applicable to many glycans like occurring in glycolipids, glycosaminoglycans, proteoglycans and oligosaccharides derived from homo- as well as hetero-polysaccharides. The greater part of the NMR parameters we collected is summarised in refs. 19–24. The analysis of glycans, starting from simple compounds up to highly complex structures yielded insight into metabolic pathways comprising biosynthesis and catabolic processes. Also for the diagnosis of congenital disorders of carbohydrate metabolism, it is relevant to be able to determine unambiguously the structure of metabolites *e.g.*, in urine or serum. Many fundamental results were obtained that allowed further studies regarding the biological and physical function of glycans in glycoproteins. The studies reveal the impressive variety of glycan structures in nature. For cell-cell

interactions also glycolipids have an essential role. Glycoproteins and glycolipids have many epitopes in common. For each of the glycans the presentation to the outside world, the imbedding of the glycoconjugates in the cell membrane, as well as the mobility and dynamics have to be considered to come to full understanding of the mode of action.

Important challenges remain the functioning of glycans in embryogenesis, cell differentiation, normal and malignant growth and senescence. Advances in analysing the human proteome and the proteome-adaptation in cell programming, show the need for the study of the dynamic aspects of the glycome and the glycoproteome to gain insight in the actual biological functioning of glycoproteins. The unravelling of the message encoded by those structures remains thereby an intriguing problem.

Acknowledgement

The author is deeply indebted to all former members of the Department Bio-organic Chemistry and to many colleagues from all over the world for invaluable contributions to the studies described in this review. In particular, the fruitful collaboration with colleagues in the University Lille, France and in several universities and institutes in Japan should be emphasized.

References

- Montreuil, J., Vliegthart, J.F.G. and Schachter, H. (1995) Glycoproteins I. New comprehensive biochemistry, vol. 29a, Elsevier, Amsterdam.
- Montreuil, J., Vliegthart, J.F.G. and Schachter, H. (1997) Glycoproteins II. New comprehensive biochemistry, vol. 29b, Elsevier, Amsterdam.
- Varki, A., Cummings, R.D., Esko, J.D., Freeze, H., Stanley, P., Bertozzi, C.R. and Eizler, M.E. (2009) Essentials of Glycobiology. Cold Spring Harbour Press, New York.
- Hofsteenge, J., Müller, D.R., De Beer, T., Löffler, A., Richter, W.J. and Vliegthart, J.F.G. (1994) New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry* **33**, 13524–13530.
- De Beer, T., Vliegthart, J.F.G., Löffler, A. and Hofsteenge, J. (1995) The hexopyranosyl residue that is C-linked to the side-chain of tryptophan in human RNase Us is α -mannopyranose. *Biochemistry* **34**, 11785–11789.
- Oman, T.J., Boettcher, J.M., Wang, H., Okalibe, X.N. and Van der Donk, W.A. (2011) Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat. Chem. Biol.* **7**, 78–80.
- Dorland, L., Haverkamp, J., Schut, B.L., Vliegthart, J.F.G., Spik, G., Strecker, G., Fournet, B. and Montreuil, J. (1977) The structure of the asialo-carbohydrate units of human serotransferrin as proven by 360 MHz proton magnetic resonance spectroscopy. *FEBS Lett.* **77**, 15–20.
- Kamerling, J.P., Gerwig, G.J., Vliegthart, J.F.G. and Clamp, J.R. (1975) Characterization by Gas-Liquid-Chromatography-Mass Spectroscopy and Proton Magnetic Resonance Spectroscopy of Permethysilyl Methyl glycosides obtained in the Methanolysis of Glycoproteins and Glycopeptides. *Biochem. J.* **151**, 491–495.
- Gerwig, G.J., Kamerling, J.P. and Vliegthart, J.F.G. (1978) Determination of D and L configuration of neutral monosaccharides by high resolution capillary GLC. *Carbohydr. Res.* **62**, 349–357.
- Gerwig, G.J., Kamerling, J.P. and Vliegthart, J.F.G. (1979) Determination of the absolute configuration of monosaccharides in complex carbohydrates by capillary GLC. *Carbohydr. Res.* **77**, 1–7.
- Hocking, H.G., Gerwig, G., Duterte, S., Violette, A., Favreau, Ph., Stöcklin, R., Kamerling, J.P. and Boelens, R. (2013) Structure of the O-Glycosylated Conopeptide CeTx from *Conus comsors*. *Chem. Eur. J.* **19**, 870–879.
- Ciacanu, I. and Kerek, F. (1984) A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **131**, 209–219.
- Kamerling, J.P. and Vliegthart, J.F.G. (1989) Carbohydrates in Clinical Biochemistry—Principles, Methods, Applications (eds. Curtius, H.Chr. and Roth, M.) Vol. 1, Mass spectrometry (ed. Lawson, A.M.). W. de Gruyter, Berlin, pp. 175–263.
- Kamerling, J.P., Heerma, W., Vliegthart, J.F.G., Green, B.N., Lewis, A.S., Strecker, G. and Spik, G. (1983) Fast Atom Bombardment Mass spectrometry of carbohydrate chains derived from Glycoproteins. *Biomed. Mass Spectrom.* **10**, 420–425.
- Yang, Y., Barendrecht, A., Kamerling, J.P. and Heck, A. (2013) Analyzing protein micro-heterogeneity in chicken ovalbumin by high resolution native mass spectrometry exposes qualitatively and semi-quantitatively 59 proteoforms. *Anal. Chem.* **85**, 12037–12045.
- Kamerling, J.P. and Vliegthart, J.F.G. (1983) Gas liquid chromatography and mass spectrometry of sialic acids. *Cell Biol. Monogr. Vol. 10*, Sialic Acids (ed. Schauer, R.). Springer Verlag, New York, pp. 95–125.
- Vliegthart, J.F.G., Dorland, L., Van Halbeek, H. and Haverkamp, J. (1983) NMR Spectroscopy of sialic acids. *Cell Biol. Monogr. Vol. 10*, Sialic Acids (ed. Schauer, R.). Springer Verlag, New York, pp. 127–172.
- Schauer, R. (2000) Achievements and challenges of sialic acid research. *Glycoconj. J.* **17**, 485–499.
- Vliegthart, J.F.G., Van Halbeek, H. and Dorland, L. (1980) High resolution ^1H NMR spectroscopy in the structure analysis of carbohydrates

- derived from glycoproteins. In 27th IUPAC Int. Congr. Pure and Appl. Chem. (ed. Varmavuori) Pergamon, Oxford, pp. 253–262.
- 20) Vliegenthart, J.F.G., Van Halbeek, H. and Dorland, L. (1981) The applicability of 500-MHz high resolution $^1\text{H-NMR}$ spectroscopy for the structure determination of carbohydrates derived from glycoproteins. *Pure Appl. Chem.* **53**, 45–77.
- 21) Vliegenthart, J.F.G., Dorland, L. and Van Halbeek, H. (1983) High resolution $^1\text{H-NMR}$ spectroscopy as a tool in the structural analysis of carbohydrates related to glycoproteins. *Adv. Carbohydr. Chem. Biochem.* **41**, 209–374.
- 22) Kamerling, J.P. and Vliegenthart, J.F.G. (1992) High resolution $^1\text{H-NMR}$ spectroscopy of oligosaccharide-alditols released from mucin-type *O*-glycoproteins. *Biol. Magn. Reson.* **10**, 1–194.
- 23) Vliegenthart, J.F.G. and Kamerling, J.P. (2007) $^1\text{H-NMR}$ structural-reporter-group concept in carbohydrate analysis. In *Comprehensive Glycoscience* (eds. Kamerling, J.P. et al.). Elsevier, Amsterdam, Vol. **2**, pp. 1–68.
- 24) Leeftang, B.R. and Vliegenthart, J.F.G. (2012) Glycoprotein Analysis: Using nuclear magnetic resonance. *Encyclopedia Analytical Chemistry*, Wiley Online Library, DOI: 10.1002/9780470027318.
- 25) Lis, H. and Sharon, N. (1978) Soybean agglutinin — a plant glycoprotein, structure of the carbohydrate unit. *J. Biol. Chem.* **253**, 3468–3476.
- 26) Dorland, L., Van Halbeek, H., Vliegenthart, J.F.G., Lis, H. and Sharon, N. (1981) Primary structure of the carbohydrate chain of soybean agglutinin, a reinvestigation by high resolution $^1\text{H-NMR}$ spectroscopy. *J. Biol. Chem.* **256**, 7708–7711.
- 27) Neeser, J.R., Del Vedovo, S., Mutsaers, J.H.G.M. and Vliegenthart, J.F.G. (1985) Structural analysis of the carbohydrate chains of legume storage proteins by 500 MHz $^1\text{H-NMR}$ spectroscopy. *Glycoconj. J.* **2**, 355–364.
- 28) Van Halbeek, H., Dorland, L., Vliegenthart, J.F.G., Spik, G., Cheron, A. and Montreuil, J. (1981) Structure determination of two oligomannoside-type glycopeptides obtained from bovine lactotransferrin by 500 MHz $^1\text{H-NMR}$ spectroscopy. *BBA* **675**, 293–296.
- 29) Kozutsumi, Y., Nakao, Y., Teramura, T., Kawasaki, T., Mutsaers, J.H.G.M., Van Halbeek, H. and Vliegenthart, J.F.G. (1986) Structures of oligomannoside chains of α -mannosidase from porcine kidney. *J. Biochem.* **99**, 1253–1265.
- 30) Michalski, J.-C., Montreuil, J., Strecker, G., Van Halbeek, H., Dorland, L., Vliegenthart, J.F.G. and Farriaux, J.-P. (1983) Primary structure of three mannosyl-glycoasparagines isolated from the urine of two patients with Gaucher's disease (Infantile form). *Eur. J. Biochem.* **132**, 375–381.
- 31) De Waard, P., Kamerling, J.P., Van Halbeek, H., Vliegenthart, J.F.G. and Broertjes, J.J.S. (1987) Characterization of *N*-linked gluco-oligomannose type of carbohydrate chains of glycoproteins from the ovary of the starfish *Asteria rubens* (L.). *Eur. J. Biochem.* **168**, 679–685.
- 32) Ogawa, T., Nukada, T. and Kitajima, T. (1983) Synthesis of a linear hexahexosyl unit of a high-mannose type of glycan chain of a glycoprotein. *Carbohydr. Res.* **123**, C12–C15.
- 33) Ronin, C., Van Halbeek, H., Mutsaers, J.H.G.M. and Vliegenthart, J.F.G. (1987) Isolation and $^1\text{H-NMR}$ spectroscopic identification of the glucose-containing lipid-linked precursor oligosaccharide of *N*-linked carbohydrate chains. *Glycoconj. J.* **4**, 247–254.
- 34) De Praeter, C.M., Gerwig, G.J., Bause, E., Nuytinck, L.K., Vliegenthart, J.F.G., Breuer, W., Kamerling, J.P., Espeel, M.F., Martin, J.-J.R., De Paepe, A.M., Chun Chan, N.W., Dacremont, D.A. and Van Coster, R.N. (2000) A novel disorder caused by defective biosynthesis of *N*-linked oligosaccharides due to Glucosidase I deficiency. *Am. J. Hum. Genet.* **66**, 1744–1756.
- 35) Fournet, B., Montreuil, J., Strecker, G., Dorland, L., Haverkamp, J., Vliegenthart, J.F.G., Binette, P. and Schmid, K. (1978) Determination of the primary structures of 16 asialo-carbohydrate units derived from human plasma α 1-acid glycoprotein by 360-MHz $^1\text{H-NMR}$ spectroscopy and permethylation analysis. *Biochemistry* **17**, 5206–5214.
- 36) Schmid, K., Binette, P., Dorland, L., Vliegenthart, J.F.G., Fournet, B. and Montreuil, J. (1979) The primary structure of the asialo-carbohydrate units of the first glycosylation site of human plasma α 1-acid glycoprotein. *BBA* **581**, 356–359.
- 37) Dorland, L., Haverkamp, J., Vliegenthart, J.F.G., Fournet, B., Strecker, G., Spik, G., Montreuil, J., Schmid, K. and Binette, P. (1978) Determination by 360 MHz $^1\text{H-NMR}$ spectroscopy of the type of branching in complex asparagine-linked glycan chains of glycoproteins. *FEBS Lett.* **89**, 149–152.
- 38) Fournet, B., Strecker, G., Spik, G., Montreuil, J., Schmid, K., Binette, P., Dorland, L., Haverkamp, J., Schut, B.L. and Vliegenthart, J.F.G. (1979) Structure of 10 glycopeptides from α 1-acid glycoprotein. *Glycoconj. Res. Proc. 4th International Glycoconj. Symp.* **1**, 149–156. Acad. Press, New York.
- 39) Montreuil, J. and Vliegenthart, J.F.G. (1979) Primary Structure and Conformation of Glycans *N*-Glycosidically linked to Peptide Chains. *Glycoconj. Res. Proc. 4th International Glycoconj. Symp.* **1**, 35–78. Acad. Press, New York.
- 40) Van Halbeek, H., Dorland, L., Vliegenthart, J.F.G., Montreuil, J., Fournet, B. and Schmid, K. (1981) Characterization of the microheterogeneity in glycoproteins by 500 MHz $^1\text{H-NMR}$ spectroscopy of glycopeptide preparations. Application to a monofucosylated tetra-antennary glycopeptide fraction from human α 1-acid glycoprotein. *J. Biol. Chem.* **256**, 5588–5590.
- 41) Paz Parente, J., Wieruszkeski, G., Strecker, G., Montreuil, J., Fournet, B., Van Halbeek, H., Dorland, L. and Vliegenthart, J.F.G. (1982) A novel type of carbohydrate structure present in

- ovomucoid. *J. Biol. Chem.* **257**, 13173–13176.
- 42) PazParente, J., Strecker, G., Leroy, J., Montreuil, J., Fournet, B., Van Halbeek, H., Dorland, L. and Vliegthart, J.F.G. (1983) Primary structure of a novel N-Glycosidic unit, derived from hen ovomucoid. *FEBS Lett.* **152**, 145–152.
- 43) Bernard, N., Engler, R., Strecker, G., Montreuil, J., Van Halbeek, H. and Vliegthart, J.F.G. (1984) The structure of a trisialyl di-antennary N-type glycopeptide obtained from rat plasma hemopexin. *Glycoconj. J.* **1**, 123–140.
- 44) Damm, J.B.L., Voshol, H., Hård, K., Kamerling, J.P. and Vliegthart, J.F.G. (1989) Analysis of N-acetyl-4-O-acetylneuraminic-acid-containing N-linked carbohydrate chains released by peptide- N^4 -(N-acetyl- β -glucosaminyl)asparagine amidase F. Application to the structure determination of the carbohydrate chains of equine fibrinogen. *Eur. J. Biochem.* **180**, 101–110.
- 45) De Waard, P., Koorevaar, J.P., Kamerling, J.P. and Vliegthart, J.F.G. (1991) Structure determination by $^1\text{H-NMR}$ spectroscopy of (sulfated) sialylated N-linked carbohydrate chains released from porcine thyroglobulin by peptide- N^4 -(N-acetyl- β -glucosaminyl)asparagine amidase F. *J. Biol. Chem.* **266**, 4237–4243.
- 46) Buko, A.M., Kentzer, E.J., Petros, A., Menon, G., Zuiderweg, E.R.P. and Sarin, V.K. (1991) Characterization of a posttranslational fucosylation in the growth factor of urinary plasminogen activator. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3992–3996.
- 47) Bergwerff, A.A., Thomas-Oates, J.E., Van Oostrom, J., Kamerling, J.P. and Vliegthart, J.F.G. (1992) Human urokinase contains GalNAc(β -4)[Fuc(α 1-3)]GlcNAc(β -2) as a novel terminal element in N-linked carbohydrate chains. *FEBS Lett.* **314**, 389–394.
- 48) Bergwerff, A.A., Van Oostrom, J., Kamerling, J.P. and Vliegthart, J.F.G. (1995) The major N-linked carbohydrate chains from human urokinase. The occurrence of 4-O-sulfated, (α 2-6)-sialylated or (α 1-3)-fucosylated N-acetylgalactosamine(β 1-4)-N-acetylglucosamine elements. *Eur. J. Biochem.* **228**, 1009–1019.
- 49) Voshol, H., Van Zuylen, C.W.E.M., Orberger, G., Vliegthart, J.F.G. and Schachner, M. (1996) Structure of the HNK-1 Carbohydrate Epitope on Bovine Peripheral Myelin Glycoprotein P0. *J. Biol. Chem.* **271**, 22957–22960.
- 50) Gallego, G., Blanco, J., Thijssen-van Zuylen, C.W.E.M., Gotfredsen, H., Voshol, H., Duus, J.Ø., Schachner, M. and Vliegthart, J.F.G. (2001) Epitope Diversity of N-Glycans from Bovine Myelin Glycoprotein P0 Revealed by Mass Spectrometry and Nano Probe Magic Angle Spinning $^1\text{H NMR}$ Spectroscopy. *J. Biol. Chem.* **276**, 30834–30844.
- 51) Tamm, I. and Horsfall, F.L. (1950) Characterization and separation of an inhibitor of viral hemagglutination present in urine. *Proc. Soc. Exp. Biol. Med.* **74**, 108–114.
- 52) Muchmore, A.V. and Decker, J.M. (1985) Uromodulin a unique 85 kilo-dalton immunosuppressive glycoprotein isolated from the urine of pregnant women. *Science* **229**, 479–481.
- 53) Sikri, K.L., Foster, C.L., MacHugh, N. and Marshall, R.D. (1981) Localization of Tamm-Horsfall glycoprotein in the human kidney using immune-fluorescence and immune-microscopical techniques. *J. Anat.* **132**, 597–605.
- 54) Rindler, M.J., Naik, S.S., Li, N., Hoops, T.C. and Peraldi, M.N. (1990) Uromodulin (Tamm Horsfall glycoprotein/uromucoid) is a phosphatidylinositol-linked membrane protein. *J. Biol. Chem.* **265**, 20784–20789.
- 55) Schaeffer, C., Santambrogio, S., Perucca, S., Casari, G. and Rampoldi, L. (2009) Analysis of Uromodulin Polymerization provides new insights into the mechanisms regulating ZP-mediated protein assembly. *Mol. Biol. Cell* **20**, 589–599.
- 56) Hård, K., Van Zadelhoff, G., Moonen, P., Kamerling, J.P. and Vliegthart, J.F.G. (1992) The Asn linked carbohydrate chains of human Tamm Horsfall glycoprotein of one male. Novel sulfated and novel N-acetylgalactosamine-containing N-carbohydrate chains. *Eur. J. Biochem.* **209**, 895–915.
- 57) Van Rooijen, J.J.M., Kamerling, J.P. and Vliegthart, J.F.G. (1998) Sulfated di-, tri- and tetra-antennary oligosaccharides in human Tamm-Horsfall glycoprotein. *Eur. J. Biochem.* **256**, 471–487.
- 58) Van Rooijen, J.J.M., Voskamp, A.F., Kamerling, J.P. and Vliegthart, J.F.G. (1999) Glycosylation sites and site-specific glycosylation in human Tamm-Horsfall glycoprotein. *Glycobiology* **9**, 21–30.
- 59) Van Rooijen, J.J.M., Kamerling, J.P. and Vliegthart, J.F.G. (1998) The abundance of additional N-acetylglucosamine units in N-linked tetra-antennary oligosaccharides of human Tamm-Horsfall glycoprotein is a donor-specific feature. *Glycobiology* **8**, 1065–1075.
- 60) Rohfritsch, Ph.F., Rinnbauer, M., Vliegthart, J.F.G. and Kamerling, J.P. (2004) Donor-specificity in the glycosylation of Tamm-Horsfall glycoprotein: Conservation of the Sd^a determinant in pairs of twins. *Glycobiology* **14**, 365–372.
- 61) Van Rooijen, J.J.M., Hermentin, P., Kamerling, J.P. and Vliegthart, J.F.G. (2001) The patterns of the complex- and oligomannose-type glycans of uromodulin (Tamm-Horsfall glycoprotein) in the course of pregnancy. *Glycoconj. J.* **18**, 539–546.
- 62) Easton, R.L., Patankar, M.S., Clark, G.F., Morris, H.R. and Dell, A. (2000) Pregnancy associated changes in the glycosylation of Tamm-Horsfall glycoprotein. Expression of sialyl Lewis^x sequences on core 2 type O-glycans derived from uromodulin. *J. Biol. Chem.* **275**, 21928–21938.
- 63) Stroop, C.J.M., Weber, W., Gerwig, G.J., Nimtz, M., Kamerling, J.P. and Vliegthart, J.F.G. (2000) Characterization of the carbohydrate chains of the secreted form of the human epidermal growth factor receptor. *Glycobiology*

- 10, 901–917.
- 64) Stroop, C.J.M., Weber, W., Nimtz, M., Gutiérrez Galego, R., Kamerling, J.P. and Vliegenthart, J.F.G. (2000) Fucosylated hybrid-type N-glycans on the secreted epidermal growth factor receptor from swainsonine-treated A431 cells. *Arch. Biochem. Biophys.* **374**, 42–51.
- 65) Staudacher, E., Altmann, F., Glössl, J., März, L., Schachter, H., Kamerling, J.P., Hård, K. and Vliegenthart, J.F.G. (1991) GDP-Fucose: β -N-acetylglucosamine-[Fuc to (Fuc α 1-6GlcNAc)-Asn-peptide]- α 1-3-fucosyl-transferase activity in honeybee (*Apis mellifica*) venom glands. The difucosylation of asparagine-bound N-acetylglucosamine. *Eur. J. Biochem.* **199**, 745–751.
- 66) Kubelka, V., Altmann, F., Staudacher, E., Tretter, V., März, L., Hård, K., Kamerling, J.P. and Vliegenthart, J.F.G. (1993) Primary structures of the N-linked carbohydrate chains from honey bee venom phospholipase A₂. *Eur. J. Biochem.* **213**, 1193–1204.
- 67) Ishihara, H., Takahashi, N., Oguri, S. and Tejima, S. (1979) Complete structure of the carbohydrate moiety of stem bromelain. *J. Biol. Chem.* **254**, 10715–10719.
- 68) Van Kuik, J.A., Hoffmann, R.A., Mutsaers, J.H.G.M., Van Halbeek, H., Kamerling, J.P. and Vliegenthart, J.F.G. (1986) A 500 MHz ¹H-NMR study on the N-linked carbohydrate chain on Bromelain. *Glycoconj. J.* **3**, 27–34.
- 69) Bouwstra, J.B., Spoelstra, E.C., De Waard, P., Leeflang, B.R., Kamerling, J.P. and Vliegenthart, J.F.G. (1990) Conformational studies on the N-linked carbohydrate chain of bromelain. *Eur. J. Biochem.* **190**, 113–122.
- 70) Kamerling, J.P. and Vliegenthart, J.F.G. (1997) Hemocyanins in Glycoproteins II, *New Comprehensive Biochemistry* **29^b** (eds. Montreuil, J., Vliegenthart, J.F.G. and Schachter, H.). Elsevier, Amsterdam, pp. 123–142.
- 71) Hall, R.L., Wood, E.J., Kamerling, J.P., Gerwig, G.J. and Vliegenthart, J.F.G. (1977) 3-O-methyl sugars as constituents of glycoproteins. Identification of 3-O-methylgalactose and 3-O-methyl mannose in pulmonate gastropod haemocyanine. *Biochem. J.* **165**, 173–176.
- 72) Van Kuik, A.J., Sijbesma, R., Kamerling, J.P., Vliegenthart, J.F.G. and Wood, E.J. (1986) Primary structure of a low-molecular mass N-linked oligosaccharide of *Lymnea stagnalis*; 3-O-methyl-D-mannose as a constituent of the xylose-containing core structure in an animal glycoprotein. *Eur. J. Biochem.* **160**, 621–625.
- 73) Van Kuik, J.A., Sybesma, R., Kamerling, J.P., Vliegenthart, J.F.G. and Wood, E.J. (1987) Primary structure determination of seven novel N-linked carbohydrate chains derived from hemocyanin of *Lymnea stagnalis*. *Eur. J. Biochem.* **169**, 399–411.
- 74) Lommerse, J.P.M., Thomas-Oates, J.E., Gielens, C., Préaux, G., Kamerling, J.P. and Vliegenthart, J.F.G. (1997) Primary structure of 21 novel monoantennary and diantennary N-linked carbohydrate chains from α _D-Hemocyanin of *Helix pomatia*. *Eur. J. Biochem.* **249**, 195–222.
- 75) Van Kuik, A.J., Van Halbeek, H., Kamerling, J.P. and Vliegenthart, J.F.G. (1986) Primary structure of the neutral carbohydrate chains of hemocyanin from *Panulirus interruptus*. *Eur. J. Biochem.* **159**, 297–301.
- 76) Van Kuik, A.J., Breg, J., Kamerling, J.P. and Vliegenthart, J.F.G. (1987) Primary structure of the acidic carbohydrate chain of hemocyanin from *Panulirus interruptus*. *FEBS Lett.* **221**, 150–154.
- 77) Nasir-Ud-Din, Jeanloz, R.W., Lamblin, G., Roussel, Ph., Van Halbeek, H., Mutsaers, J.H.G.M. and Vliegenthart, J.F.G. (1986) Structure of sialyl-oligosaccharides isolated from Bonnet monkey (*Macaca radiata*) cervical mucus glycoproteins exhibiting multiple blood group activities. *J. Biol. Chem.* **261**, 1992–1997.
- 78) Gerwig, G. and Vliegenthart, J.F.G. (2000) Analysis of glycoprotein derived glycopeptides in Proteomics in Functional Genomics EXS **88** (eds. Jollès, P. and Jörnvall, H.). pp. 159–186.
- 79) Hokke, C.H., Damm, J.B.L., Kamerling, J.P. and Vliegenthart, J.F.G. (1993) Structure of three acidic O-linked carbohydrate chains of porcine *zona pellucida* glycoproteins. *FEBS Lett.* **329**, 29–34.
- 80) Hokke, C.H., Damm, J.B.L., Penninkhof, B., Aitken, R.J., Kamerling, J.P. and Vliegenthart, J.F.G. (1994) Structure of the O-linked carbohydrate chains of porcine *zona pellucida* glycoproteins. *Eur. J. Biochem.* **221**, 491–512.
- 81) Mori, E., Takasaki, S., Hedrick, J.L., Wardrip, J.L., Mori, T. and Kobata, A. (1991) Neutral Oligosaccharide Structures Linked to Asparagines of Porcine Zona Pellucida Glycoproteins. *Biochemistry* **30**, 2078–2087.
- 82) Damm, J.B.L., Hård, K., Kamerling, J.P., Van Dedem, G.W.K. and Vliegenthart, J.F.G. (1990) Structure determination of the major N- and O-linked carbohydrate chains of the β subunit of equine chorionic gonadotropin. *Eur. J. Biochem.* **189**, 175–183.
- 83) Hokke, C.H., Roosenboom, M.J.H., Thomas-Oates, J.E., Kamerling, J.P. and Vliegenthart, J.F.G. (1994) Structure determination of the disialylated poly-(N-acetyllactosamine)-containing O-linked carbohydrate chains of equine chorionic gonadotropin. *Glycoconj. J.* **11**, 35–41.
- 84) Kamerling, J.P., Hård, K. and Vliegenthart, J.F.G. (1990) Structure analysis of the carbohydrate chains of native and recombinant — DNA glycoproteins. *In* From Clone to Clinic (eds. Crommelin, D.J.A. and Schellekens, H.). Kluwer Academic Publishers, Dordrecht, pp. 295–304.
- 85) Mutsaers, J.H.G.M., Kamerling, J.P., Devos, R., Guisez, Y., Fiers, W. and Vliegenthart, J.F.G. (1986) Structural studies on the carbohydrate chains of recombinant human γ -interferon. *Eur. J. Biochem.* **156**, 651–654.
- 86) Hård, K., Mekking, A., Damm, J.B.L., Kamerling,

- J.P., De Boer, W., Wijnands, R.A. and Vliegthart, J.F.G. (1990) Isolation and structure determination of the intact sialylated N-linked carbohydrate chains of recombinant human Follitropin expressed in Chinese Hamster Ovary cells. *Eur. J. Biochem.* **193**, 263–271.
- 87) Hofsteenge, J., Löffler, A., Müller, D.R., Richter, W.J., De Beer, T. and Vliegthart, J.F.G. (1996) Protein C-glycosylation. *Techn. Prot. Chem.* **7**, 163–171.
- 88) Krieg, J., Hartmann, S., Vicentini, A., Gläser, W., Hess, D. and Hofsteenge, J. (1998) Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-x-x-Trp. *Mol. Biol. Cell* **9**, 301–309.
- 89) Buettner, F.F.R., Ashikov, A., Tiemann, B., Lehle, L. and Bakker, H. (2013) *C. elegans* DPY is a C-mannosyltransferase glycosylating thrombospondin repeats. *Mol. Cell* **50**, 295–302.
- 90) Doucey, M.-A., Hess, D., Cacan, R. and Hofsteenge, J. (1998) Protein C-mannosylation is enzyme catalysed and uses dolichol-phosphate-mannose as a precursor. *Mol. Biol. Cell* **9**, 291–300.
- 91) Julenius, K. (2007) NetCGlyc 1.0: Prediction of mammalian C-sites. *Glycobiology* **17**, 868–876.
- 92) Vliegthart, J.F.G. and Casset, F. (1998) Novel forms of protein glycosylation. *Curr. Opin. Struct. Biol.* **8**, 567–571.
- 93) Van Halbeek, H., Dorland, L., Veldink, G.A., Vliegthart, J.F.G., Garegg, P.J., Norberg, Th. and Lindberg, B. (1982) A 500 MHz proton-magnetic-resonance study of several fragments of the carbohydrate-protein linkage region commonly occurring in proteoglycans. *Eur. J. Biochem.* **127**, 1–6.
- 94) Sugahara, K., Yamashina, I., De Waard, P., Van Halbeek, H. and Vliegthart, J.F.G. (1988) Structural studies on sulfated glycopeptides from the carbohydrate-protein linkage region of chondroitin 4-sulfate proteoglycans of swarm rat chondrosarcoma. *J. Biol. Chem.* **263**, 10168–10174.
- 95) Sugahara, K., Ohi, Y., Harada, T., De Waard, P. and Vliegthart, J.F.G. (1992) Structural studies on sulfated glycopeptides from the carbohydrate-protein linkage region of chondroitin 6-sulfate proteoglycans of shark cartilage. I Six compounds containing 0 or 1 sulfate and/or phosphate residue. *J. Biol. Chem.* **267**, 6027–6035.
- 96) De Waard, P., Vliegthart, J.F.G., Harada, T. and Sugahara, K. (1992) Structural studies on sulfated glycopeptides from the carbohydrate-protein linkage region of chondroitin 6-sulfate proteoglycans of shark cartilage. II Seven compounds containing 2 or 3 sulfate residues. *J. Biol. Chem.* **267**, 6036–6043.
- 97) Sugahara, K., Masuda, M., Harada, T., Yamashina, I., De Waard, P. and Vliegthart, J.F.G. (1991) Structural studies on sulfated oligosaccharides derived from the carbohydrate-protein linkage region of chondroitin sulfate proteoglycans of whale cartilage. *Eur. J. Biochem.* **202**, 805–811.
- 98) Sugahara, K., Tsuda, K., Yoshida, K., Yamada, S., De Beer, T. and Vliegthart, J.F.G. (1995) Structure determination of the octa- and deca-saccharide sequences isolated from the carbohydrate-protein linkage of porcine intestinal heparin. *J. Biol. Chem.* **270**, 22914–22923.
- 99) De Beer, T., Inui, A., Tsuda, H., Sugahara, K. and Vliegthart, J.F.G. (1996) Polydispersity in sulfation profile of oligosaccharide alditols isolated from the protein-linkage region and the repeating disaccharide region of chondroitin 4-sulfate of bovine nasal septal cartilage. *Eur. J. Biochem.* **240**, 789–797.
- 100) Ferreon, A.C.M., Ferreon, J.C., Wright, P.E. and Deniz, A.A. (2013) Modulation of allostery by protein intrinsic disorder. *Nature* **498**, 390–394.
- 101) Bah, A., Vernon, R.M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, Ch., Sonenberg, N., Kay, L.E. and Forman-Kay, J.D. (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, 106–109.
- 102) Babu, M.M. (2016) The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **44**, 1185–1200.
- 103) Eggens, I., Fendersen, B., Toyokuni, T., Dean, B., Stroud, M. and Hakomori, S. (1989) Specific interaction between Le^x and Le^x determinants. *J. Biol. Chem.* **264**, 9476–9484.
- 104) Mamman, M., Choi, S.-K. and Whitesides, G. (1998) Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew. Chem. Int. Ed. Engl.* **37**, 2754–2794.
- 105) Haseley, S.R., Vermeer, H.J., Kamerling, J.P. and Vliegthart, J.F.G. (2001) Carbohydrate self-recognition mediates marine sponge cellular adhesion. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 19–24.
- 106) Santos, J.L., Carvalho de Souza, A., Canada, F.J., Martin-Santamaria, S., Kamerling, J.P. and Jimenez-Barbero, J. (2009) Assessing carbohydrate-carbohydrate interactions by NMR spectroscopy: The trisaccharide epitope from the marine sponge *Microciona prolifera*. *ChemBioChem* **10**, 511–519.
- 107) Van Zuylen, C.W.E.M., De Beer, T., Rademaker, G.J., Haverkamp, J., Thomas-Oates, J.E., Hård, K., Kamerling, J.P. and Vliegthart, J.F.G. (1995) Site-specific and complete enzymatic deglycosylation of the native human chorionic gonadotropin α subunit. *Eur. J. Biochem.* **231**, 754–760.
- 108) De Beer, T., Van Zuylen, C.W.E.M., Leeflang, B.R., Hård, K., Boelens, R., Kaptein, R., Kamerling, J.P. and Vliegthart, J.F.G. (1996) NMR studies of the free α subunit of human chorionic gonadotropin; structural influences of N-glycosylation and the β subunit on the conformation of the α subunit. *Eur. J. Biochem.* **241**, 229–242.

(Received Nov. 5, 2016; accepted Dec. 9, 2016)

Profile

Johannes Frederik Gerardus (Hans) Vliegenthart, was born in The Netherlands, in 1936. He studied Chemistry at Utrecht University and obtained his Ph.D. in 1967. Subsequently, he was offered a research position at Utrecht University to create a new research group focused on Bio-organic Chemistry. As specific project, he selected determination of the structure of carbohydrates by mass spectrometry and NMR spectroscopy. These techniques were under development offering new approaches to this field. The application of mass spectrometry enabled the determination of structural details of oligosaccharides and the identification of a large number of members of the sialic acid family. A breakthrough was realized by the introduction of high-resolution $^1\text{H-NMR}$ spectroscopy. The resolving power of the high-field instruments (220–900 MHz) allowed the identification of numerous glycans, ranging from oligosaccharides to very complex compounds. For this work he developed the ‘structural reported group concept’ that made the unambiguous analysis of glycans feasible. In parallel programs, organic synthesis of oligosaccharides was performed and the development of synthetic carbohydrate-based vaccines initiated. New methods were explored to study carbohydrate-carbohydrate interactions. For example, after structure determination and synthesis of the glycan structures involved in the cell-cell interaction of the sponge cells *Microciona prolifera*, the molecular basis of the interaction could be established.

He was author or co-author of about 750 research publications and co-editor of 7 books.

Under his supervision 90 Ph.D. theses were prepared. He was appointed Professor at Utrecht University in 1975, Head of Department in 1980 and honorary Professor in 2003. He was Dean of the Faculty of chemistry from 1985–1989 and from 2000–2003. He was founder of the Bijvoet Center for Biomolecular Research in 1988, its Scientific Director till 2000 and chairman of the International Scientific Board of the Bijvoet Center since 2006.

He acted as Chairman/Member of various national and international scientific bodies, including the International Scientific Advisory Committee of the Science Frontier Program of the Riken Institute, Wako, Japan. He organized several international and European conferences.

He is Foreign Member of the Royal Swedish Academy of Sciences (1987), Member of the Royal Netherlands Academy of Arts and Sciences (1990), Honorary Member of the American Society for Biochemistry and Molecular Biology (1989), of the Netherlands Society for Glycobiology (2006), and Fellow of IUPAC (1998).

He is Honorary Doctor at University Debrecen, Hungary (1992), at University Lille, France in 1993 and at University Stockholm, Sweden in 1997. He received the Hilditch Memorial Lecture Award in 1978, the Shield of the Medical Faculty of the University of Tokyo, Japan (1988), Louis Pasteur Medal of the University Lille, France (1993), the Claude S. Hudson Award from the ACS (1994), Bijvoet Medal of Utrecht University in 2000, Medal of Rome University, Tor Vergata, Italy (2003), Silver Medal of Utrecht University in 2003. He was distinguished as Knight in the Order of the Lion of the Netherlands (1998).

