Eur. J. Biochem. 221, 491–512 (1994)

Structure of the O-linked carbohydrate chains of porcine zona pellucida glycoproteins

Cornelis H. HOKKE[1], Jan B. L. DAMM[2], Bea PENNINKHOF[1], R. John AITKEN[3], Johannis P. KAMERLING[1] and Johannes F. G. VLIEGENTHART[1]

[1] Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, Utrecht, The Netherlands
[2] Organon International BV, Oss, The Netherlands
[3] Medical Research Center, Reproductive Biology Unit, Edinburgh, Scotland

(Received October 5/December 12, 1993) – EJB 931507/5

The N-linked carbohydrate chains of porcine zona pellucida glycoproteins were released by digestion with peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase F and subsequently separated from the O-glycoprotein by gel-permeation chromatography on Bio-Gel P-100. The O-linked carbohydrate chains were released from the O-glycoprotein by alkaline borohydride treatment. Fractionation of the extremely heterogeneous mixture of O-linked oligosaccharide alditols was achieved by a combination of chromatographic techniques comprising gel-permeation chromatography on Bio-Gel P-4 and P-6, anion-exchange FPLC on Mono Q, and high-pH anion-exchange chromatography on CarboPac PA-1. The primary structures of 32 O-glycans were determined by one- and two-dimensional ^{1}H-NMR spectroscopy. The major part of the analyzed compounds contain a combination of the structural elements Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol, Galβ1-4(6SO_{4-})GlcNAc, and α2-3-linked Neu5Gc or Neu5Ac. This series of compounds has the following structure, where $n = 0$ to >6: [Neu5Gc/Acα2-3]$_{0-1}$[Galβ1-4(6SO_{4-})GlcNAcβ1-3]$_n$Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol.

In addition, smaller compounds were identified in which the Galβ1-3GalNAc-ol core is substituted by Neu5Gc/Ac α2-6-linked to GalNAc-ol and/or Neu5Gc/Ac α2-3-linked to Gal. Furthermore, oligosaccharides were obtained in which the distribution of 6-*O*-sulfated GlcNAc residues differs from that in the above-mentioned general structure, and a small portion of the oligosaccharides has the GlcNAcβ1-3GalNAc-ol core structure. Analysis of the endo-β-galactosidase digests of pools of N- and O-glycans indicated that the two types of oligosaccharides contain qualitatively similar poly(*N*-acetyllactosamine) chains. In the case of the N-linked carbohydrate chains, multiple branching of the core structures occurs, resulting in an even larger heterogeneity than observed for the O-linked carbohydrate chains.

The zona pellucida (ZP) is the glycoprotein matrix surrounding the mammalian oocyte during the final stages of its development. It mediates several critical steps in the fertilization process, including the species-specific binding of spermatozoa to the ovum, induction of the acrosome reaction, prevention of polyspermy and the physical protection of the growing embryo during the preinplantation stages of development [1, 2].

As shown for different species, the ZP consists of families of glycoproteins differing in molecular mass and containing both N- and O-linked carbohydrate chains [2–6]. Although the molecular basis of the functions of the ZP is still poorly understood, it has become clear during the past few years that the carbohydrate part of the ZP glycoproteins plays an important role in the sperm recognition event. In the case of murine ZP, it has been shown that a specific population of the O-linked oligosaccharides possesses sperm-receptor activity [7, 8]. For porcine ZP (pZP), which shares cross-reactive antigens with human ZP [9], contradictory results have been reported concerning the involvement of the carbohydrate chains in sperm binding. One study showed that the mixture of neutral N-linked carbohydrate chains has sperm-receptor activity [10], whereas another presented evidence for the mixture of O-linked glycans to inhibit sperm-oocyte binding [11].

Since porcine oocytes are relatively easy to obtain in sufficient quantities, most studies concerning the structure of

Correspondence to J. F. G. Vliegenthart, Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, P.O. Box 80.075, NL-3508 TB, Utrecht, The Netherlands
Fax: +31 30 540980.

Abbreviations. FID, free induction decay; HOHAHA, homonuclear Hartmann-Hahn; HPAEC, high-pH anion-exchange chromatography; MLEV, composite pulse devised by M. Levitt; NAc, *N*-acetyl; NGc, *N*-glycolyl; Neu5Ac, *N*-acetylneuraminic acid; Neu5Gc, *N*-glycolylneuraminic acid; Neu5Gc/Ac, mixture of *N*-glycolylneuraminic acid and *N*-acetylneuraminic acid; PAD, pulsed amperometric detection; PNGase-F, peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase F; pZP, porcine zona pellucida; Sia, sialic acid; WEFT, water-eliminated Fourier transform; ZP, zona pellucida.

Enzymes. Peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase F (EC 3.5.1.52); endo-β-galactosidase (EC 3.2.1.103).

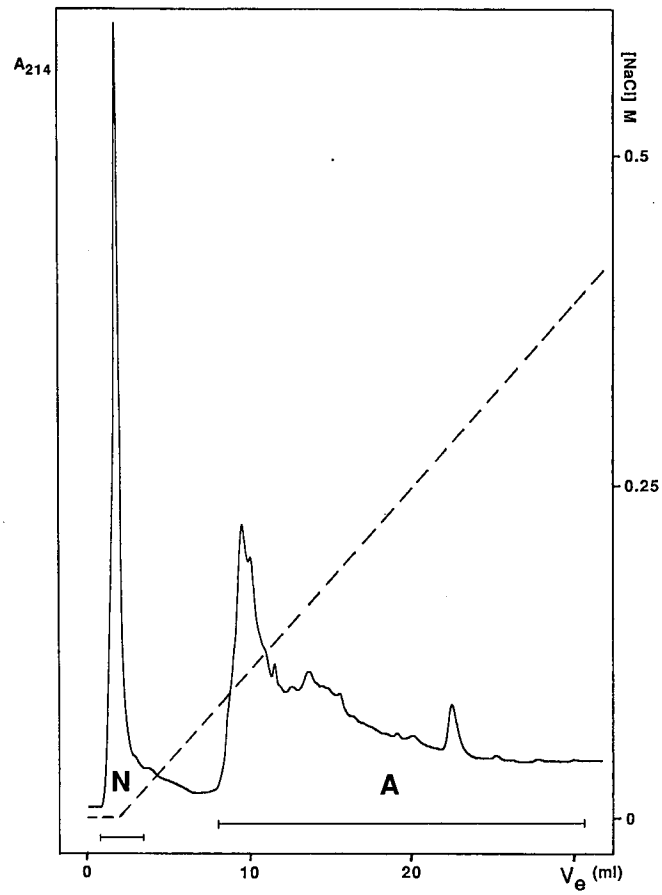


Fig. 1. Fractionation pattern at 214 nm on a FPLC HR 5/5 Mono Q column of the O-linked carbohydrate chains derived from de-*N*-glycosylated pZP glycoproteins by alkaline borohydride treatment. The column was eluted with a linear concentration gradient of 0–500 mM NaCl in H_2O as indicated (-----) at a flow rate of 2 ml/min.

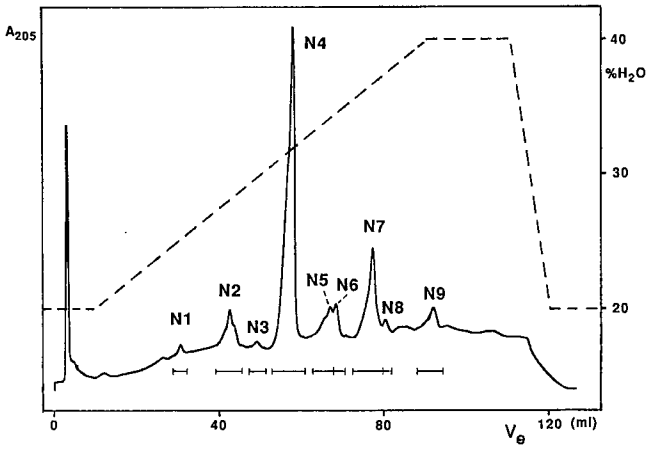


Fig. 2. Fractionation pattern at 205 nm on a HPLC Lichrosorb-NH_2 column (25×0.46 cm) of FPLC fraction N, containing neutral O-linked oligosaccharide alditols obtained from pZP glycoproteins. The elution was carried out at a flow rate of 2 ml/min using a gradient of 20–40% (by vol.) H_2O in acetonitrile as indicated (-----).

ZP-derived carbohydrate chains have been carried out using pZP. The pZP is composed of three major glycoprotein families (pZP1–pZP3) and contains both N- and O-linked carbohydrate chains [3–5, 11, 12]. The pZP3 family accounts for approximately 78% of the total mass of the pZP [3] and possesses the boar sperm receptor activity [13]. The structures of the neutral N-linked carbohydrate chains of unfractionated pZP glycoproteins [14] and purified pZP3 [10] have been established, and recently the structures of the neutral O-linked oligosaccharide alditols of pZP glycoproteins have been reported [15]. Furthermore, some structures and structural elements of the acidic N-glycans of pZP3 have been described [16]. In addition, general aspects of the pZP glycoproteins have been reported, suggesting the presence of sialylated and/or sulfated *N*-acetyllactosamine repeats in both N- and O-linked carbohydrate chains giving rise to extreme structural heterogeneity [4, 5, 11, 16].

In order to clarify the role of the carbohydrate chains in the biological functioning of pZP, detailed knowledge of the structures of the N- and O-linked oligosaccharides is a prerequisite. In this study, we have purified the O-linked glycans from pZP glycoproteins by a sequence of chromatographic techniques and determined their structures by one- and two-dimensional ^{1}H-NMR spectroscopy. Furthermore, pZP-derived pools of both N- and O-linked carbohydrate chains were digested with endo-β-galactosidase and the digestion products were purified and analyzed by ^{1}H-NMR spectroscopy.

MATERIALS AND METHODS

Materials

Peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase F (PNGase-F) from *Flavobacterium meningosepticum* and endo-β-galactosidase from *Bacteroides fragilis* were obtained from Boehringer Mannheim.

Isolation of zona pellucida glycoproteins

Zonae pellucidae were obtained by homogenizing frozen-thawed porcine ovaries in a commercial meat grinder with copious amounts of ice-cold saline. This homogenate was sieved through two nylon screens of pore sizes 500 µm and 210 µm to remove debris. The zonae were finally isolated from the filtrate on a 75-µm nylon screen. This crude zona preparation was purified by centrifugation in a discontinuous Percoll gradient (40%, 20%, 10% bottom→top) for 30 min (2000 *g*, 25°C). The ovae were collected from the 10–20% interface and were washed free from Percoll using NaCl/P_i (Na_2HPO_4, 1.15 g/l; KH_2PO_4, 0.2 g/l; NaCl, 8 g/l; KCl, 0.2 g/l; pH 7.4) and gently homogenized using a small glass homogenizer. Zonae were resuspended in NaCl/P_i, heat solubilized at 78°C for 20 min and finally ultracentrifuged (100000 *g*, 90 min, 4°C). The clear supernatant was dialyzed against several changes of 0.1 M NH_4HCO_3 and lyophilized.

SDS/PAGE of the heat-solubilized zonae preparation gave rise to two diffuse bands in the molecular mass ranges of 90–85 kDa and 66–50 kDa. The higher band corresponds with ZP1. The lower band represents a combination of ZP2, ZP3α and ZP3β, yielding a diffuse pattern due to the extreme heterogeneity of the carbohydrate moieties of these glycoproteins.

Table 1. ^{1}H-chemical shifts of structural-reporter-group protons of the constituent monosaccharides of neutral O-linked oligosaccharide alditols derived from pZP. Chemical shifts are given at 22 °C and were measured in ^2H_2O relative to internal acetone (δ 2.225 [36]). For indexing of the monosaccharide residues, see text; n.d., not determined. An asterisk indicates assignments may have to be interchanged. Some values are given with only two decimals because of spectral overlap.

Residue	Reporter group	Chemical shift in					
		N2A	N2B	N4	N6	N7	N9
		ppm					
GalNAc-ol	H2	4.291	4.400	4.401	4.401	4.400	4.400
	H3	4.002	4.049	4.050	4.050	4.049	4.049
	H4	n.d.	3.490	3.493	3.493	3.492	3.493
	H5	4.147	4.186	4.187	4.190	4.186	4.186
	NAc	2.038	2.047	2.047	2.047	2.047	2.046
Gal^3	H1	–	4.462	4.461	4.462	4.460	4.460
	H4	–	4.124	4.127	4.127	4.125	4.125
GlcNAc^3	H1	4.631	4.662	4.683	4.686	4.680	4.680
	H6	4.021	3.895	3.952	3.960	3.953*	3.95
	NAc	2.083	2.044	2.041	2.042	2.038	2.035
Gal^4	H1	4.454	–	4.481	4.553	4.467	4.465
	H4	3.924	–	3.926	4.185	4.160	4.157
GlcNAc^i	H1	–	–	–	–	–	4.698
	H6	–	–	–	–	–	3.95
	NAc	–	–	–	–	–	2.035
Gal^i	H1	–	–	–	–	–	4.465
	H4	–	–	–	–	–	4.157
GlcNAc^t	H1	–	–	–	–	4.700	4.698
	H6	–	–	–	–	3.949*	3.95
	NAc	–	–	–	–	2.034	2.035
Gal^t	H1	–	–	–	–	4.479	4.479
	H4	–	–	–	–	3.925	3.924
Galα	H1	–	–	–	5.146	–	–
	H4	–	–	–	4.019	–	–

Liberation of the carbohydrate chains

The N-linked carbohydrate chains were released from pZP glycoproteins essentially as described [17]. Briefly, 60 mg lyophilized heat-solubilized pZP was dissolved in 6 ml 50 mM Tris/HCl, pH 7.5, containing 50 mM EDTA. Subsequently, 1% (by vol.) 2-mercaptoethanol and 1% (mass/vol.) SDS were added and the mixture was kept for 3 min at 80 °C. After cooling down to room temperature, 1 µl Nonidet P-40/mg pZP was added and, after thorough mixing, 1 U PNGase-F/mg pZP was introduced. The mixture was incubated at ambient temperature. After 4 h, a fresh aliquot of 1 U PNGase-F/mg pZP was introduced and the incubation was continued for another 16 h. Completeness of the liberation of the N-linked carbohydrate chains was checked by SDS/PAGE. The incubation mixture was fractionated by gel-permeation chromatography on a Bio-Gel P-100 column (57×1.5 cm, 200–400 mesh, Bio-Rad) using 25 mM NH_4HCO_3 as eluent. Apart from the void-volume fraction which contained the O-glycoprotein–SDS complex, all carbohydrate-containing fractions (orcinol/H_2SO_4) were pooled and lyophilized. Remaining detergents were removed from the carbohydrate pool by affinity chromatography on an ExtractiGel-D column (8×1 cm, Pierce), using 25 mM NH_4HCO_3 as eluent. The run-through fraction was collected and desalted on a Bio-Gel P-2 column (18×1 cm, 200–400 mesh, Bio-Rad) eluted with water.

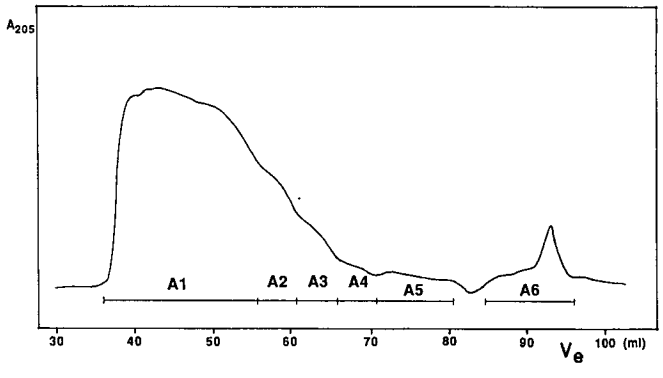


Fig. 3. Elution profile at 205 nm on a Bio-Gel P-4 column (150×1.15 cm) of FPLC fraction A, containing O-linked anionic oligosaccharide alditols obtained from pZP glycoproteins. The column was eluted at 7 ml/h with 100 mM NH_4HCO_3.

In order to release the O-linked carbohydrate chains, the N-deglycosylated pZP glycoproteins eluting in the void-volume peak of the P-100 column were treated with alkaline borohydride as described previously [18]. The reaction products were purified via ExtractiGel-D as described for the N-linked carbohydrate chains.

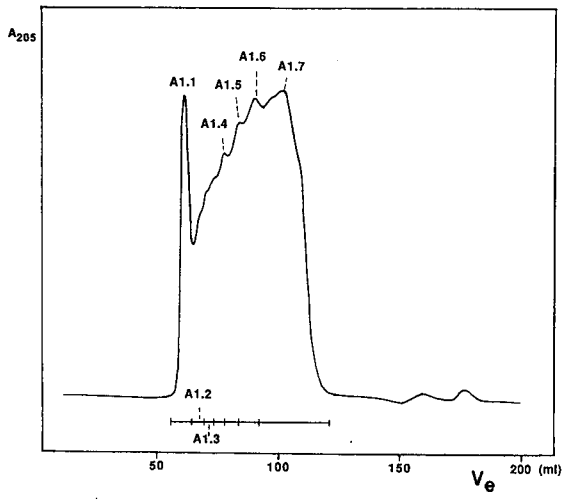


Fig. 4. Elution profile at 205 nm on a Bio-Gel P-6 column (135×2.2 cm) of Bio-Gel P-4 fraction A1. The column was eluted at 23 ml/h with 100 mM NH_4HCO_3.

FPLC fractionation

The pools of enzymically and chemically released oligosaccharides as well as Bio-Gel P-4 and Bio-Gel P-6 fractions (see below) were fractionated on a Mono Q HR 5/5 anion-exchange column using a Pharmacia FPLC system [18]. The column was eluted at 1 ml/min with a concentration gradient of NaCl in water, as indicated in the figures. The eluent was monitored at 214 nm, and collected fractions were lyophilized, desalted on a Bio-Gel P-2 column (20×1 cm, 200–400 mesh, Bio-Rad), and lyophilized again.

Gel chromatography

The anionic O-linked oligosaccharides and the endo-β-galactosidase digestion products were separated on a Bio-Gel P-4 column (150×1.15 cm, 200–400 mesh, Bio-Rad), eluted with 100 mM NH_4HCO_3 at a flow rate of 7 ml/h. The eluent was monitored at 206 nm and fractions of 1.2 ml were collected. The fractions were pooled as indicated in the figures, and lyophilized.

Subfractionation of Bio-Gel P-4 fraction A1 was carried out on a Bio-Gel P-6 column (135×2.2 cm, 200–400 mesh, Bio-Rad), using 100 mM NH_4HCO_3 as eluent. The flow rate was 23 ml/h and the absorbance was monitored at 206 nm. Fractions of 5.8 ml were collected, pooled as indicated in the figure, and lyophilized.

Bio-Gel P-4 and P-6 fractions, as well as the neutral FPLC fraction N were applied to a small column of Dowex 50W-X8, H^+ form (6×0.5 cm, 100–200 mesh, Fluka). The column was eluted with 6 ml 0.01 M formic acid, and the eluate was lyophilized.

HPLC fractionation

Fractionation of oligosaccharide mixtures by HPLC on a Lichrosorb-NH_2 10-μm column (25×0.46 cm, Chrompack) was carried out using a Spectroflow 400 HPLC system (ABI analytical, Kratos Division). The column was eluted at 2 ml/min and the eluent was monitored at 205 nm or 206 nm. Mixtures of neutral oligosaccharides were fractionated using a concentration gradient of H_2O in acetonitrile, whereas anionic oligosaccharides were separated using a gradient of 15 mM KH_2PO_4/K_2HPO_4, pH 7.0 in acetonitrile, as indicated in the figures. Acetonitrile was evaporated from collected fractions by a stream of nitrogen prior to lyophilization. The fractions obtained upon using phosphate buffer were desalted on a Bio-Gel P-2 column (45×1 cm, 200–400 mesh, Bio-Rad), and lyophilized again.

High-pH anion-exchange chromatography

Further fractionations of some Mono Q and Lichrosorb-NH_2 fractions were carried out by high-pH anion-exchange chromatography (HPAEC) with pulsed amperometric detection (PAD) on a CarboPac PA-1 pellicular anion-exchange column (25×0.9 cm, Dionex) as described [19]. The column was eluted with a concentration gradient of NaOAc in 0.1 M NaOH, as indicated in the figures, at a flow rate of 4 ml/h. Collected fractions were immediately neutralized by addition of 2 M HCl and lyophilized. The fractions were desalted on a column of Bio-Gel P-2 (45×1 cm, 200–400 mesh, Bio-Rad) eluted with 5 mM NH_4HCO_3, and lyophilized again.

Monosaccharide analysis

Monosaccharide analysis was carried out by gas chromatography on a capillary CP-Sil 5 WCOT fused silica column (25 m×0.34 mm internal diameter, Chrompack) using a Hewlett Packard 5890 GC station. The trimethylsilylated methyl glycosides were prepared by methanolysis, *N*-(re)acetylation, and trimethylsilylation as reported [20].

Endo-β-galactosidase digestion

A portion of the PNGase-F released oligosaccharides (D_N; 1 mg) and a portion of the oligosaccharide alditols obtained by alkaline borohydride treatment (D_O; 1 mg) were each dissolved in 50 μl 50 mM NaOAc, pH 5.8. Then, 200 mU endo-β-galactosidase in 200 μl 50 mM NaOAc, pH 5.8, containing bovine serum albumin (0.2 mg/ml) and 0.05% (mass/vol.) sodium azide, were added to the solution of D_N, and 100 mU endo-β-galactosidase in 100 μl of the same buffer were added to the solution of D_O. Each mixture was incubated for 40 h at 37°C and after centrifugation (12000×*g*), the supernatants were subjected to Bio-Gel P-4 chromatography.

^{1}H-NMR spectroscopy

Prior to ^{1}H-NMR spectroscopic analysis, samples were exchanged twice in 99.9% ^2H_2O. Finally, samples were dissolved in 99.96% ^2H_2O (MSD Isotopes). ^{1}H-NMR spectra were recorded at 500 MHz on a Bruker AMX-500 or at 600 MHz on a Bruker AMX-600 spectrometer at a probe temperature of 22°C, unless indicated otherwise. Chemical shifts are expressed in ppm by reference to internal acetone (δ 2.225). Typically, one-dimensional spectra were recorded with a spectral width of 5000 Hz at 500 MHz or 6000 Hz at 600 MHz, collecting 128–2500 free induction decays (FIDs) of 8 K or 16 K complex data points. Suppression of the residual water signal was achieved by applying the WEFT pulse sequence as described [19]. The resolution of the one-dimensional spectra was enhanced by Lorentzian-to-Gaussian transformation and the final spectra were baseline corrected with a polynomal function when necessary.

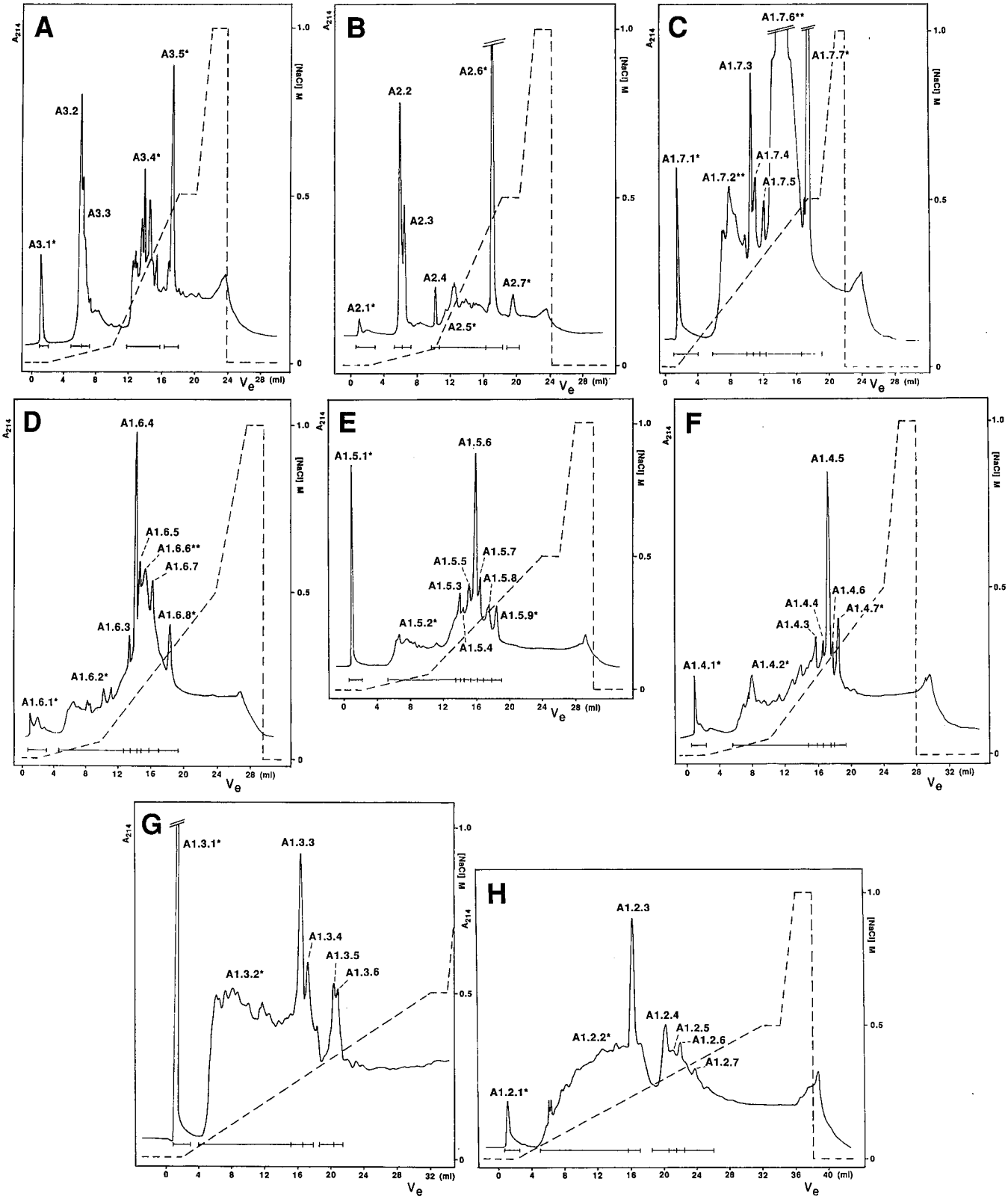


Fig. 5. Fractionation patterns at 214 nm on a FPLC HR 5/5 Mono Q column of Bio-Gel P-4 fraction A3 and A2, and Bio-Gel P-6 fractions A1.7–A1.2. Elutions were carried out at a flow rate of 1 ml/min with gradients of NaCl in H_2O as indicated (–––––). Fractions marked with * did not contain carbohydrate material, and fractions marked with ** contained mainly non-carbohydrate contaminants. (A) Fraction A3; (B) fraction A2; (C) fraction A1.7; (D) fraction A1.6; (E) fraction A1.5; (F) fraction A1.4; (G) fraction A1.3; (H) fraction A1.2.

Table 2. ^{1}H-chemical shifts of structural-reporter-group protons of the constituent monosaccharides of sulfated or sialylated O-linked oligosaccharide alditols derived from pZP, isolated from the Bio-Gel P-4 fractions A2 and A3. Chemical shifts are given at 22°C and were measured in ^2H_2O relative to internal acetone (δ 2.225 [36]). For indexing of the monosaccharide residues, see text; n.d., not determined. Some values are given with only two decimals because of spectral overlap. The GlcNAc^3 H1 signal given here for A2.2.2 (δ 4.676) differs from that reported for reference compound 81 in [22] at δ 4.698 so the latter was apparently not correctly assigned; it was obtained from the ^{1}H-NMR spectrum of a complex mixture.

Residue	Reporter group	Chemical shift in							
		A3.2A	A3.2B	A3.3.5	A2.2.2	A2.2.6	A2.2.7	A2.3	A2.4
		ppm							
GalNAc-ol	H2	4.392	4.392	4.401	4.399	4.400	4.387	4.401	4.384
	H3	4.074	4.074	4.049	4.049	4.049	4.047	4.049	4.073
	H4	3.492	3.492	3.490	3.489	3.491	3.532	3.490	3.525
	H5	4.192	4.192	4.195	4.188	4.189	4.239	4.187	4.249
	H6′	n.d.	n.d.	n.d.	n.d.	n.d.	3.474	n.d.	3.472
	NAc	2.045	2.045	2.047	2.047	2.047	2.046	2.047	2.044
Gal^3	H1	4.548	4.548	4.460	4.461	4.461	4.461	4.460	4.549
	H3	4.125	4.135	n.d.	n.d.	n.d.	n.d.	n.d.	4.13
	H4	3.928	3.934	4.154	4.128	4.129	4.123	4.126	3.932
GlcNAc^3	H1	–	–	4.699	4.676	4.677	4.690	4.679	–
	H6	–	–	4.391	3.959	3.95	3.952	3.950	–
	H6′	–	–	4.317	n.d.	n.d.	n.d.	n.d.	–
	NAc	–	–	2.039	2.037	2.038	2.041	2.038	–
Gal^4	H1	–	–	4.526	4.560	4.563	4.481	4.470	–
	H3	–	–	n.d.	4.118	4.131	n.d.	n.d.	–
	H4	–	–	3.926	3.956	3.962	3.926	4.188	–
GlcNAc^t	H1	–	–	–	–	–	–	4.716	–
	H6	–	–	–	–	–	–	4.395	–
	H6′	–	–	–	–	–	–	4.314	–
	NAc	–	–	–	–	–	–	2.032	–
Gal^t	H1	–	–	–	–	–	–	4.523	–
	H4	–	–	–	–	–	–	3.926	–
Neu5Ac^3	H3a	1.802	–	–	1.799	–	–	–	–
	H3e	2.772	–	–	2.757	–	–	–	–
	NAc	2.032	–	–	2.030	–	–	–	–
Neu5Gc^3	H3a	–	1.818	–	–	1.816	–	–	1.819
	H3e	–	2.789	–	–	2.776	–	–	2.792
	NGc	–	4.121	–	–	4.120	–	–	4.120
Neu5Gc^6	H3a	–	–	–	–	–	1.709	–	1.712
	H3e	–	–	–	–	–	2.741	–	2.741
	NGc	–	–	–	–	–	4.121	–	4.120

For the two-dimensional homonuclear Hartmann-Hahn (HOHAHA) spectra [21], a MLEV-17 mixing sequence of 100–120 ms was used. The 90° pulse width was adjusted to about 26 µs and the spectral width was 3000–3500 Hz (for 500-MHz spectra) or 3600–4200 Hz (for 600-MHz spectra) in both dimensions. The HO^{2}H signal was presaturated for 1 s during the relaxation delay. In total 296–512 spectra of 2048 data points with 56–128 scans per t_1 value were recorded. Two-dimensional NMR data were processed on a VAXstation 3100, or on Silicon Graphics Iris Indigo or 4D/35 stations, using TRITON software (Bijvoet Center, Department of NMR Spectroscopy, Utrecht University). The time-domain data were multiplied with a phase-shifted sine bell. After Fourier transformation, the resulting data set of 1024×2048 points was baseline corrected in both frequency domains with a fourth-order polynomal fit when necessary.

RESULTS

Monosaccharide analysis of pZP revealed the presence of Fuc, Gal, GalNAc, GlcNAc, Man and Neu5Ac in a molar ratio of 1.1:8.3:1.1:10.8:3.0:1.1. It should be noted that during the release of monosaccharides followed by *N*-reacetylation, Neu5Gc is converted to Neu5Ac.

The N-linked carbohydrate chains were split off enzymically with PNGase-F, and the released oligosaccharides were separated from the remaining O-glycoprotein by gel-permeation chromatography on Bio-Gel P-100 (not shown). The Bio-Gel P-100 void volume fraction was lyophilized and subsequently treated with alkaline borohydride. The pool of released O-linked carbohydrate chains was subjected to FPLC on Mono Q, giving rise to two fractions, denoted N and A (Fig. 1). Fraction N, eluting in the void volume, contains the neutral oligosaccharide alditols, whereas fraction A,

Table 3. ^{1}H-chemical shifts of structural-reporter-group protons of the constituent monosaccharides of sialylated and/or sulfated O-linked oligosaccharide alditols derived from pZP, isolated from the Bio-Gel P-6 fractions A1.6 and A1.7. Chemical shifts are given at 22°C and were measured in ^2H_2O relative to internal acetone (δ 2.225 [36]). For indexing of the monosaccharide residues, see text. A1.7.3 is a mixture of compounds containing different types of sialic acids, see text. Some values are given with only two decimals because of spectral overlap; n.d., not determined. Asterisks indicate that assignments may have to be interchanged within one column.

Residue	Reporter group	Chemical shift in								
		A1.7.3	A1.7.4A	A1.7.4B	A1.7.5	A1.6.4A	A1.6.4B	A1.6.5	A1.6.7A	A1.6.7B
		ppm								
GalNAc-ol	H2	4.383	4.399	4.399	4.400	4.400	4.400	4.400	4.400	4.400
	H3	4.071	4.048	4.048	4.048	4.049	4.049	4.049	4.048	4.048
	H4	3.522	3.489	3.489	3.489	3.488	3.488	3.487	3.487	3.487
	H5	4.25	4.188	4.188	4.188	4.189	4.189	4.189	4.192	4.192
	NAc	2.043	2.047	2.047	2.046	2.047	2.047	2.047	2.047	2.047
Gal^3	H1	4.544	4.459	4.459	4.459	4.459	4.459	4.459	4.459	4.459
	H3	4.12–4.13	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
	H4	3.93	4.126	4.126	4.125	4.127	4.127	4.126	4.155	4.155
GlcNAc^3	H1	–	4.677	4.677	4.677	4.677	4.677	4.678	4.692	4.692
	H6	–	n.d.	n.d.	3.949	3.956	3.956	3.950	4.35–4.40	4.35–4.40
	H6′	–	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4.28–4.35	4.28–4.35
	NAc	–	2.037	2.037	2.037	2.037	2.037	2.038	2.033	2.033
Gal^4	H1	–	4.469	4.469	4.468	4.469	4.469	4.468	4.511	4.511
	H4	–	4.193	4.193	4.186	4.193	4.193	4.186	4.188	4.188
GlcNAc^i	H1	–	–	–	4.711	4.703*	4.703*	4.708*	4.706	4.706
	H6	–	–	–	4.376	4.381	4.381	4.375	4.35–4.40	4.35–4.40
	H6′	–	–	–	4.328	4.330	4.330	4.332	4.28–4.35	4.28–4.35
	NAc	–	–	–	2.028**	2.028**	2.028**	2.028**	2.033	2.033
Gal^i	H1	–	–	–	4.507	4.506	4.506	4.506	4.511	4.511
	H4	–	–	–	4.186	4.193	4.193	4.186	4.188	4.188
GlcNAc^{ii}	H1	–	–	–	–	–	–	4.708*	–	–
	H6	–	–	–	–	–	–	4.375	–	–
	H6′	–	–	–	–	–	–	4.332	–	–
	NAc	–	–	–	–	–	–	2.033**	–	–
Gal^{ii}	H1	–	–	–	–	–	–	4.513	–	–
	H4	–	–	–	–	–	–	4.186	–	–
GlcNAc^t	H1	–	4.707	4.707	4.711	4.707*	4.707*	4.716*	4.706	4.706
	H6	–	4.403	4.403	4.397	4.398	4.398	4.394	4.35–4.40	4.35–4.40
	H6′	–	4.314	4.314	4.297	4.292	4.292	4.294	4.28–4.35	4.28–4.35
	NAc	–	2.029	2.029	2.037**	2.033**	2.033**	2.038**	2.033	2.033
Gal^t	H1	–	4.601	4.601	4.528	4.608	4.606	4.530	4.608	4.608
	H3	–	4.12–4.13	4.12–4.13	n.d.	4.12–4.13	4.12–4.13	n.d.	4.12–4.13	4.12–4.13
	H4	–	3.969	3.963	3.923	3.969	3.963	3.923	3.969	3.962
Neu5Ac^3	H3a	1.802	–	1.804	–	–	1.802	–	–	1.803
	H3e	2.774	–	2.746	–	–	2.747	–	–	2.750
	NAc	2.031	–	2.029	–	–	2.028	–	–	2.028
Neu5Ac^6	H3a	1.693	–	–	–	–	–	–	–	–
	H3e	2.724	–	–	–	–	–	–	–	–
	NAc	2.031	–	–	–	–	–	–	–	–
Neu5Gc^3	H3a	1.819	1.819	–	–	1.820	–	–	1.819	–
	H3e	2.792	2.767	–	–	2.768	–	–	2.768	–
	NGc	4.120	4.116	–	–	4.119	–	–	4.116	–
Neu5Gc^6	H3a	1.712	–	–	–	–	–	–	–	–
	H3e	2.740	–	–	–	–	–	–	–	–
	NGc	4.120	–	–	–	–	–	–	–	–

Table 4. ^{1}H-chemical shifts of structural-reporter-group protons of the constituent monosaccharides of sulfated or sialylated O-linked oligosaccharide alditols derived from pZP, isolated from the Bio-Gel P-6 fractions A1.4 and A1.5. Chemical shifts are given at 22°C and were measured in ^2H_2O relative to internal acetone (δ 2.225 [36]). For indexing of the monosaccharide residues, see text. It was an arbitrary choice that GlcNAc^i of A1.4.3A and A1.4.3B is not 6-*O*-sulfated (see text). Some values are given with only two decimals because of spectral overlap; n.d., not determined. Asterisks indicate that assignments may have to be interchanged within one column.

Residue	Reporter group	Chemical shift in								
		A1.5.3A	A1.5.3B	A1.5.6A	A1.5.6B	A1.5.7	A1.4.3A	A1.4.3B	A1.4.5A	A1.4.5B
		ppm								
GalNAc-ol	H2	4.399	4.399	4.399	4.399	4.400	4.400	4.400	4.400	4.400
	H3	4.049	4.049	4.048	4.048	4.049	4.048	4.048	4.049	4.049
	H4	3.488	3.488	3.489	3.489	3.488	3.489	3.489	3.488	3.488
	H5	4.19	4.19	4.188	4.188	4.19	4.19	4.19	4.19	4.19
	NAc	2.046	2.046	2.046	2.046	2.047	2.046	2.046	2.047	2.047
Gal^3	H1	4.459	4.459	4.459	4.459	4.459	4.460	4.460	4.459	4.459
	H4	4.125	4.125	4.126	4.126	4.126	4.125	4.125	4.126	4.126
GlcNAc^3	H1	4.678	4.678	4.677	4.677	4.677	4.677	4.677	4.677	4.677
	H6	n.d.	n.d.	3.95	3.95	3.949	n.d.	n.d.	n.d.	n.d.
	NAc	2.038	2.038	2.037	2.037	2.037	2.037	2.037	2.037	2.037
Gal^4	H1	4.466	4.466	4.468	4.468	4.468	4.465	4.465	4.468	4.468
	H4	4.158	4.158	4.189	4.189	4.186	4.156	4.156	4.188	4.188
GlcNAc^i	H1	4.707*	4.707*	4.706	4.706	4.707*	4.705	4.705	4.705	4.705
	H6	n.d.	n.d.	4.35–4.40	4.35–4.40	4.35–4.40	n.d.	n.d.	4.35–4.40	4.35–4.40
	H6′	n.d.	n.d.	4.28–4.35	4.28–4.35	4.28–4.35	n.d.	n.d.	4.28–4.35	4.28–4.35
	NAc	2.028**	2.028**	2.028**	2.028**	2.028**	2.032**	2.032**	2.027**	2.027**
Gal^i	H1	4.466	4.466	4.506	4.506	4.507	4.465	4.465	4.51	4.51
	H4	4.190	4.190	4.189	4.189	4.186	4.186	4.186	4.188	4.188
GlcNAc^{ii}	H1	4.707*	4.707*	4.706	4.706	4.707*[a]	4.705[a]	4.705[a]	4.705[a]	4.705[a]
	H6	4.35–4.40	4.35–4.40	4.35–4.40	4.35–4.40	4.35–4.40[a]	4.35–4.40[a]	4.35–4.40[a]	4.35–4.40[a]	4.35–4.40[a]
	H6′	4.28–4.35	4.28–4.35	4.28–4.35	4.28–4.35	4.28–4.35[a]	4.28–4.35[a]	4.28–4.35[a]	4.28–4.35[a]	4.28–4.35[a]
	NAc	2.028**	2.028**	2.032**	2.032**	2.033**[b]	2.028/32**	2.028/32**	2.032**[b]	2.032**[b]
Gal^{ii}	H1	4.506	4.506	4.510	4.510	4.513[a]	4.506/4.510	4.506/4.510	4.51[a]	4.51[a]
	H4	4.190	4.190	4.189	4.189	4.186[a]	4.186[a]	4.186[a]	4.188[a]	4.188[a]
GlcNAc^t	H1	4.712*	4.712*	4.706	4.706	4.716*	4.705	4.705	4.705	4.705
	H6	4.35–4.40	4.35–4.40	4.35–4.40	4.35–4.40	4.35–4.40	4.35–4.40	4.35–4.40	4.35–4.40	4.35–4.40
	H6′	4.28–4.35	4.28–4.35	4.28–4.35	4.28–4.35	4.28–4.35	4.28–4.35	4.28–4.35	4.28–4.35	4.28–4.35
	NAc	2.033**	2.033**	2.032**	2.032**	2.037**	2.032**	2.032**	2.032**	2.032**
Gal^t	H1	4.605	4.605	4.607	4.607	4.531	4.607	4.607	4.609	4.609
	H3	4.12–4.13	4.12–4.13	4.12–4.13	4.12–4.13	n.d.	4.12–4.13	4.12–4.13	4.12–4.13	4.12–4.13
	H4	3.968	3.963	3.968	3.963	3.922	3.97	3.96	3.968	3.963
Neu5Ac^3	H3a	–	1.804	–	1.804	–	–	1.804	–	1.806
	H3e	–	2.743	–	2.747	–	–	2.747	–	2.747
	NAc	–	2.028	–	2.027	–	–	2.028	–	2.027
Neu5Gc^3	H3a	1.819	–	1.820	–	–	1.819	–	1.820	–
	H3e	2.767	–	2.768	–	–	2.768	–	2.769	–
	NGc	4.116	–	4.115	–	–	4.115	–	4.115	–

[a] Signal stemming from two protons.
[b] Signal stemming from two NAc groups.

eluting as a very broad, tailing peak in the 50–500 mM NaCl region, contains the anionic oligosaccharide alditols.

Neutral O-linked oligosaccharides

Analysis of fraction N by ^{1}H-NMR spectroscopy (not shown) indicated that non-carbohydrate material, presumably remaining from the alkaline borohydride treatment, was present in the oligosaccharide mixture. Therefore, fraction N was applied to a small column of Dowex 50W-X8 (H^+ form), which removed most of the contaminants. Subsequently, fraction N was further fractionated by HPLC on Lichrosorb-NH_2, and nine subfractions, denoted N1–N9, were collected (Fig. 2). Because of the low amount of material or the presence of a complex mixture of compounds in some fractions, structure determination by ^{1}H-NMR spectroscopy could only be carried out for the fractions N2, N4, N6, N7 and N9 (see Table 6). The ^{1}H-NMR data are compiled in Table 1. For convenience, the chemical shift values of the H1 and H4 resonances of Gal residues as constituents of various struc-

tural elements are summarized in Table 5. These Gal signals are excellent structural reporters for the sequence determination of *N*-acetyllactosamine units.

The ^{1}H-NMR spectrum of fraction N2 shows the presence of a mixture of two trisaccharide alditols, denoted N2A and N2B, in a molar ratio of 3:2. The structural-reporter-group data match those of reference compounds 17 and 16, respectively, in [22].

N2A Galβ1-4GlcNAcβ1-3GalNAc-ol

N2B GlcNAcβ1-3Galβ1-3GalNAc-ol

The structural-reporter-group signals in the ^{1}H-NMR spectrum of the main subfraction N4 match those of compound 12 in [22], proving the presence of the following tetrasaccharide alditol.

N4 Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol

The ^{1}H-NMR spectrum of fraction N6 provides conclusive evidence for the occurrence of a pentasaccharide alditol with a terminal α1-3-linked Gal residue.

N6 Galα1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol

The Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol element is deduced from the relevant structural-reporter-group signals by comparison with the spectrum of N4. The presence of the terminal Galα1-3Galβ element is evident from the typical Galα H1 and H4 signals at δ 5.146 ($^3J_{1,2}$ 3.9 Hz) and δ 4.019, respectively, in combination with the Gal^4 H1 and H4 signals at δ 4.553 and δ 4.185, respectively (cf. compound 1-5 in [23]). The downfield shift for Gal^4 H1 ($\Delta\delta$ +0.072), when going from N4 to N6 is identical to that when going from 1-3 to 1-5 in [23].

The ^{1}H-NMR spectrum of fraction N7 shows the presence of a hexasaccharide alditol with the following structure.

N7 Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol

The presence of the additional Gal$^t\beta$1-4GlcNAc^t unit as compared to compound N4 is inferred from the GlcNAc^t H1 (δ 4.700) and H6 (δ 3.949) resonances, together with the H1 (δ 4.479) and H4 (δ 3.925) signals of Gal^t, which are characteristic for such a terminal element in a poly(*N*-acetyllactosamine) sequence (cf. compounds 8 and 9 in [24]). The positions of the Gal^4 H1 (δ 4.467) and H4 (δ 4.160) signals indicate that the additional *N*-acetyllactosamine unit is β1-3-linked to Gal^4 (cf. compounds B, C and D in [25] and compounds 8 and 9 in [24]; see also Table 5).

^{1}H-NMR spectroscopy shows that fraction N9 contains an octasaccharide alditol, which can be considered as an extension of N7 with one Galβ1-4GlcNAc unit.

N9 Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol

The presence of the additional Galβ1-4GlcNAc group as compared to N7 is indicated by the doubled intensity of the signals at δ 4.465 (H1 of Gal^4 and Gal^i) and δ 4.157 (H4 of Gal^4 and Gal^i), and of the H1 signal at δ 4.698 (GlcNAc^t and GlcNAc^i; cf. compounds 8 and 9 in [24]; see also Table 5).

Anionic O-linked oligosaccharides

Fraction A (Fig. 1), containing the anionic O-glycans, was subjected to Bio-Gel P-4 gel chromatography, giving rise to one broad, tailing peak. The collected fractions were pooled into six subfractions, denoted A6–A1, as indicated in Fig. 3. Fractions A6–A4 did not contain carbohydrate material (orcinol/H_2SO_4). Fraction A1, containing the bulk of the material, was further fractionated on Bio-Gel P-6, and although the peak separation was again unsatisfactory, a distinct V_o peak was present. The collected fractions were pooled as indicated in Fig. 4, yielding seven subfractions, denoted A1.7–A1.1. ^{1}H-NMR spectroscopy (results not shown) revealed in each of the fractions A3, A2 and A1.7–A1.1 the presence of a complex mixture of sialylated and/or sulfated oligosaccharide alditols. The sialic acids can be recognized from the characteristic H3a and H3e signals, whereas the presence of the sulfate groups at C6 of GlcNAc residues can be concluded from the typical H6 and H6′ signals at δ 4.35–4.40 and δ 4.28–4.35, respectively. Furthermore, in each fraction, the occurrence of peptides and other non-carbohydrate contaminants was demonstrated. Therefore, each Bio-Gel P-4 and P-6 subfraction was passed over a small Dowex H^+ column, which removed part of this contaminating material (checked by ^{1}H-NMR spectroscopy).

The Bio-Gel P-4 fractions A3 and A2, and the Bio-Gel P-6 fractions A1.7–A1.2 were subfractionated on Mono Q. The collected fractions were pooled as indicated in Fig. 5A–H, and analyzed by ^{1}H-NMR spectroscopy. Fractions not discussed below did not contain enough material for structural analysis by ^{1}H-NMR spectroscopy or contained only non-carbohydrate material, or the fractions contained too complex mixtures of compounds (see figures). The structure determinations will be discussed below in an order which is convenient for explanation of the ^{1}H-NMR data. Relevant ^{1}H-NMR data are compiled in Tables 2–5. The structures are summarized in Table 6. In fractions which contain mixtures of compounds differing only in the type of sialic acid α2-3-linked to Galβ1-4R, the molar ratio of Neu5Gcα2-3Galβ1-4R/Neu5Acα2-3Galβ1-4R is approximately 2:1. The relevant figures can be used to make a rough estimation of the relative amounts of the various compounds by using the relative peak areas in the chromatograms since an accurate quantification can not be given.

Sialylated O-linked oligosaccharides

The ^{1}H-NMR spectrum of fraction A3.2 shows that it contains a mixture of two differently sialylated trisaccharide alditols A3.2A and A3.2B, in the molar ratio 5:3.

Table 5. Summary of ^{1}H-NMR chemical shifts of the H1 and H4 signals of Gal residues as constituents of various elements of poly(*N*-acetyllactosamine) chains. See also Tables 1–4 and 7. Values in parentheses are calculated based on the chemical shift values observed for the other elements.

Element	Chemical shift for	
	Gal H1	Gal H4
	ppm	
Galβ1-4GlcNAcβ1-	4.48	3.92–3.93
Galβ1-4GlcNAcβ1- \| 6SO_4-	4.52–4.53	3.92–3.93
-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-	4.47	4.16
-4GlcNAcβ1-3Galβ1-4GlcNAcβ1- \| 6SO_4-	4.47	4.19
-4GlcNAcβ1-3Galβ1-4GlcNAcβ1- \| 6SO_4-	(4.51)	(4.16)
-4GlcNAcβ1-3Galβ1-4GlcNAcβ1- \| 6SO_4-	4.51	4.19
Neu5Gc/Acα2-3Galβ1-4GlcNAcβ1-	4.56	3.96–3.97
Neu5Gc/Acα2-3Galβ1-4GlcNAcβ1- \| 6SO_4-	4.60–4.61	3.96–3.97
(-4)GlcNAcβ1-3Galβ1-3GalNAc-ol	4.46	4.12–4.13
(-4)GlcNAcβ1-3Galβ1-3GalNAc-ol \| 6SO_4-	4.46	4.15–4.16

A3.2A Neu5Acα2-3(3)Galβ1-3GalNAc-ol

A3.2B Neu5Gcα2-3(3)Galβ1-3GalNAc-ol

The individual sets of structural-reporter-group signals of A3.2A and A3.2B match those of the reference compounds 78 and 78A in [22], respectively.

The ^{1}H-NMR spectrum of Mono Q fraction A2.2 shows that it contains a complex mixture of sialylated oligosaccharides. Therefore, fraction A2.2 was further fractionated by HPAEC-PAD on CarboPac PA-1, and 10 subfractions were collected as shown in Fig. 6A. Owing to the low amount of material in most subfractions, structure determination by ^{1}H-NMR spectroscopy could only be carried out for the fractions A2.2.2, A2.2.6 and A2.2.7. The ^{1}H-NMR spectrum of fraction A2.2.2 demonstrates the presence of the following sialylated pentasaccharide alditol (cf. compound 81 in [22]).

A2.2.2

Neu5Acα2-3(3)Galβ1-4GlcNAcβ1-3(3)Galβ1-3GalNAc-ol

From the ^{1}H-NMR spectrum of fraction A2.2.6, the structure of the Neu5Gc-containing analogue of A2.2.2 can be deduced.

A2.2.6

Neu5Gcα2-3(3)Galβ1-4GlcNAcβ1-3(3)Galβ1-3GalNAc-ol

The presence of the α2-3-linked Neu5Gc residue in A2.2.6 is concluded from the set of typical Neu5Gc H3a, H3e and NGc signals at δ 1.816, δ 2.776 and δ 4.120, respectively (Table 2; cf. compound A4.5 in [26]). Compared with the ^{1}H-NMR data of A2.2.2 (Table 2), shift increments occur for Gal^4 H4 ($\Delta\delta$+0.006) and H3 ($\Delta\delta$+0.013), which are similar to those observed going from A3.2A to A3.2B (Table 2). The chromatographic behaviour of A2.2.6 as compared to A2.2.2 is in accordance with the general effect of introducing a NGc group instead of a NAc group on the retention time on CarboPac PA-1 under strong alkaline conditions [27]. The ^{1}H-NMR spectrum of fraction A2.2.7 indicates the presence of the following Neu5Gc-containing pentasaccharide alditol.

A2.2.7

Neu5Gcα2-6(6)

Galβ1-4(3)GlcNAcβ1-3(3)Galβ1-3GalNAc-ol

The ^{1}H-NMR data of A2.2.7 are similar to those of its Neu5Ac-containing analogue (compound 87 in [22], but now typical Neu5Gc instead of Neu5Ac signals are present (H3a, δ 1.709; H3e, δ 2.741; NGc, δ 4.121; Table 2; cf. compound 9A in [22]). Going from compound 87 in [22] to A2.2.7, small shift effects on the GalNAc-ol signals are observed (cf. compounds 9 and 9A in [22]).

From the ^{1}H-NMR spectrum of fraction A2.4, the structure of the following disialylated tetrasaccharide alditol is deduced.

A2.4

Neu5Gcα2-6(6)

Neu5Gcα2-3(3)Galβ1-3GalNAc-ol

The substitution pattern of the Galβ1-3GalNAc-ol core structure is found by comparison of the NMR data of an analogous oligosaccharide containing two Neu5Ac instead of two Neu5Gc residues (cf. compound 85 in [22]). As for A2.2.7, small shift effects on the GalNAc-ol signals compared to the Neu5Ac-containing analogue were observed. The presence of two Neu5Gc residues in A2.4 is reflected by the typical structural reporters (Table 2) of the Neu5Gc residue which is α2-6-linked to GalNAc-ol (H3a, δ 1.712; H3e, δ 2.741; NGc, δ 4.120; cf. compound 83A in [22]) and of the Neu5Gc residue which is α2-3-linked to Gal (H3a, δ 1.819; H3e, δ 2.792; NGc, δ 4.120; cf. compound 78A in [22]).

The ^{1}H-NMR spectrum of fraction A1.7.3 indicates the presence of the above-mentioned compound A2.4, but additional Neu5Ac signals characteristic for the Neu5Acα2-3Gal element (H3a, δ 1.802; H3e, δ 2.774) and for the Neu5Acα2-6GalNAc-ol element (H3a, δ 1.693; H3e, δ 2.724) are observed (Table 3; cf. compound 85 in [22]). These data indicate that the following mixture of disialylated tetrasaccharide alditols is present.

A1.7.3

Neu5Gc/Acα2-6(6)

Neu5Gc/Acα2-3(3)Galβ1-3GalNAc-ol

It is not possible to conclude from the NMR data of A1.7.3 which combinations of sialic acids occur exactly, but the

Table 6. O-linked carbohydrate chains obtained from pZP by alkaline borohydride treatment of the PNGase-F N-deglycosylated protein.

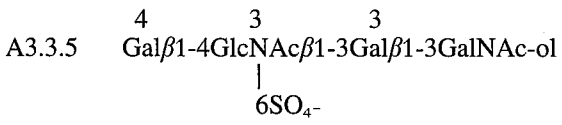
Code	Structure
N2A	Galβ1-4GlcNAcβ1-3GalNAc-ol
N2B	GlcNAcβ1-3Galβ1-3GalNAc-ol
N4	Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol
N6	Galα1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol
N7	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol
N9	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol
A3.2A	Neu5Acα2-3|Galβ1-3GalNAc-ol
A3.2B	Neu5Gcα2-3|
A3.3.5	Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | $6SO_4^-$
A2.2.2	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol
A2.2.6	Neu5Gcα2-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol Neu5Gcα2-6 \
A2.2.7	Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol
A2.3	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | $6SO_4^-$ Neu5Gcα2-6 \
A2.4	Neu5Gcα2-3Galβ1-3GalNAc-ol Neu5Gc/Acα2-6 \
A1.7.3	Neu5Gc/Acα2-3Galβ1-3GalNAc-ol
A1.7.4A	Neu5Gcα2-3|Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol |
A1.7.4B	Neu5Acα2-3| $6SO_4^-$
A1.7.5	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | | $6SO_4^-$ $6SO_4^-$
A1.6.4A	Neu5Gcα2-3|Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | | $6SO_4^-$ $6SO_4^-$
A1.6.4B	Neu5Acα2-3|
A1.6.5	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | | | $6SO_4^-$ $6SO_4^-$ $6SO_4^-$
A1.6.7A	Neu5Gcα2-3|Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | | | $6SO_4^-$ $6SO_4^-$ $6SO_4^-$
A1.6.7B	Neu5Acα2-3|
A1.5.3A	Neu5Gcα2-3|Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | | $6SO_4^-$ $6SO_4^-$
A1.5.3B	Neu5Acα2-3|
A1.5.6A	Neu5Gcα2-3|Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | | | $6SO_4^-$ $6SO_4^-$ $6SO_4^-$
A1.5.6B	Neu5Acα2-3|
A1.5.7	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | | | | $6SO_4^-$ $6SO_4^-$ $6SO_4^-$ $6SO_4^-$
A1.4.3A	Neu5Gcα2-3|Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | | $6SO_4^-$ $2\times6SO_4^-$
A1.4.3B	Neu5Acα2-3|
A1.4.5A	Neu5Gcα2-3|Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol | | | | $6SO_4^-$ $6SO_4^-$ $6SO_4^-$ $6SO_4^-$
A1.4.5B	Neu5Acα2-3|

approximate molar ratios are 1:2 for Neu5Gc[3]/Neu5Ac[3] and 2:1 for Neu5Gc[6]/Neu5Ac[6].

Based on the above-mentioned structural data and the reported behaviour of Neu5Gc- and Neu5Ac-containing oligosaccharide alditols on CarboPac PA-1 [27], it is suggested that the first group of peaks in Fig. 6A (A2.2.1–A2.2.4) results from Neu5Ac-containing alditols, whereas the second group (A2.2.5–A2.2.10) represents Neu5Gc-containing alditols.

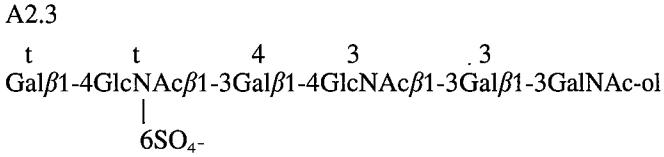
Sulfated O-linked oligosaccharides

The ^{1}H-NMR spectrum of fraction A3.3 demonstrates considerable heterogeneity. Therefore, fraction A3.3 was further fractionated by HPAEC-PAD on CarboPac PA-1, yielding nine subfractions, denoted A3.3.1−A3.3.9 (Fig. 6B). Because of the low amounts of material, structure determination by ^{1}H-NMR spectroscopy could only be carried out for the major fraction A3.3.5, which contains the following tetrasaccharide alditol.

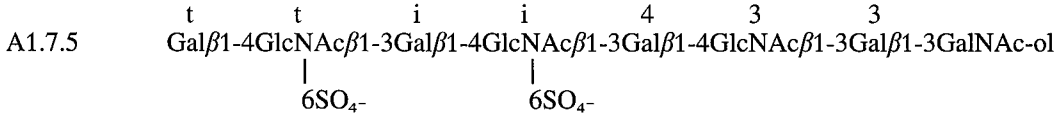


The ^{1}H-NMR spectrum of A3.3.5 has been interpreted on the basis of the ^{1}H-NMR data of the non-sulfated analogue N4 (Table 1). The presence of the sulfate group at C6 of the GlcNAc^3 residue in A3.3.5 is deduced from the characteristic position of the GlcNAc^3 H6 and H6′ signals at δ 4.391 and δ 4.317, respectively (Table 2; cf. compound A2.3 in [28], and [23, 29]). Furthermore, as compared to N4, the sulfate group affects the position of the Gal^4 H1 signal at δ 4.526, and induces typical shift increments for GlcNAc^3 H1 (δ 4.699, $\Delta\delta$ +0.016) and Gal^3 H4 (δ 4.154, $\Delta\delta$ +0.027; see Table 5).

As discussed in an earlier report [28], the structural-reporter-group regions of the ^{1}H-NMR spectrum of fraction A2.3 (see Table 2) indicate the presence of the following sulfated analogue of N7.

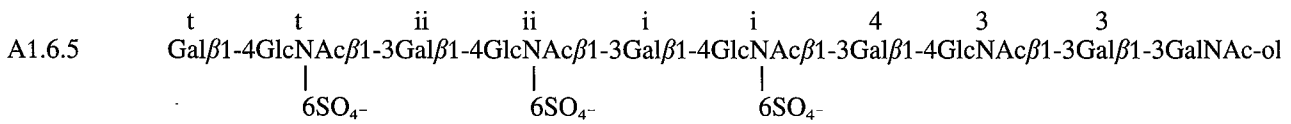


The ^{1}H-NMR spectrum of fraction A1.7.5 (Fig. 7A; Table 3) revealed the presence of the following disulfated octasaccharide alditol.



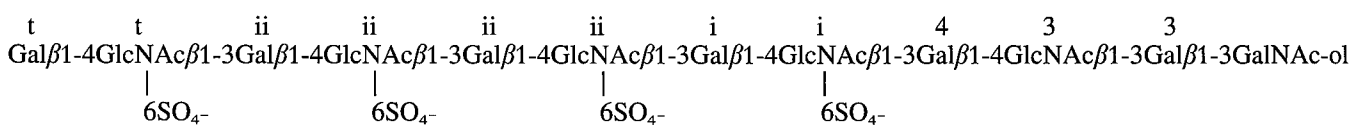
The presence of the terminal Gal^{t}β1-4(6SO$_4$-)GlcNAc^t unit and of the Gal^{4}β1-4GlcNAc^{3}β1-3Gal^{3}β1-3GalNAc-ol element are readily deduced by comparison with the NMR data of compound A2.3 (Table 2). However, the ^{1}H-NMR spectrum now contains extra signals from the internal Gal^{i}β1-4(6SO$_4$-)GlcNAc^i element at δ 4.507 (Gal^i H1), δ 4.376 (GlcNAc^i H6) and δ 4.328 (GlcNAc^i H6′). Furthermore, the intensities of the signals at δ 4.186 (Gal^4 and Gal^i H4) and δ 4.711 (GlcNAc^i and GlcNAc^t H1) have increased twofold. The downfield shift of Gal^i H1 ($\Delta\delta$ +0.042) compared to the Gal^i H1 signal in the non-sulfated analogue N9 (Table 1) is in agreement with the effect on the H1 signal of Gal^t ($\Delta\delta$ +0.044) when compounds N7 (Table 1) and A2.3 are compared (see also Table 5). As is evident from the two-dimensional HOHAHA spectrum of A1.7.5 (Fig. 8A), Gal^i H1 at δ 4.507 is interconnected with the corresponding H4 signal at δ 4.186. The overlapping GlcNAc^t and GlcNAc^i H1 signals at δ 4.711 show cross-peaks with the two sets of H6/H6′ signals at δ 4.397/δ 4.297 and δ 4.376/δ 4.328, respectively.

As shown by its ^{1}H-NMR spectrum (see Table 3), fraction A1.6.5 contains the following trisulfated decasaccharide alditol, being a linear extension of A1.7.5 with a sulfated *N*-acetyllactosamine unit.



As compared to the ^{1}H-NMR spectrum of A1.7.5 (see Table 3), additional signals are observed for GlcNAc^{ii} and Gal^{ii} H1 at δ 4.708 and δ 4.513, respectively. Furthermore, signals with increased intensities occur, stemming from H6 (δ 4.375) and H6′ (δ 4.332) of GlcNAc^i and GlcNAc^{ii}, and from H4 (δ 4.186) of Gal^4, Gal^i and Gal^{ii}. The various Gal H1 signals were assigned by correlating them to the corresponding H4 signals at δ 4.126 (for Gal^3 in a GlcNAcβ1-3Galβ1-3GalNAc-ol element), δ 4.186 [for Gal^4, Gal^i and Gal^{ii} in a (6SO$_4$-)GlcNAcβ1-3Galβ1-4 element] and δ 3.923 [for Gal^t in a Galβ1-4(6SO$_4$-)GlcNAc element], using two-dimensional HOHAHA spectroscopy (see Table 5).

The ^{1}H-NMR spectrum of fraction A1.5.7 (see Table 4) shows the occurrence of the following tetrasulfated dodecasaccharide alditol, being an extension of A1.6.5 with a sulfated *N*-acetyllactosamine unit.

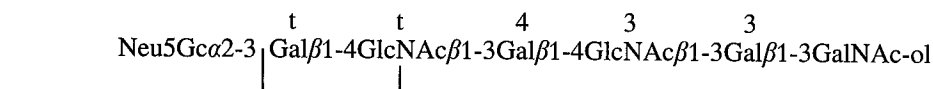


The presence of the additional Galβ1-4(6SO$_4$-)GlcNAc unit compared to compound A1.6.5 is inferred from the increased intensity of the GlcNAc^{ii}/GlcNAc^i H1, H6 and H6′ signals, and of the Gal^{ii} H1 and Gal^{ii}/Gal^i H4 signals, in the same way as described for A1.6.5 when compared to A1.7.5.

Sialylated sulfated oligosaccharides

As discussed previously in [28], the ^{1}H-NMR spectrum of fraction A1.7.4 (see Table 3) shows the presence of a mixture of two sialylated analogues of compound A2.3.

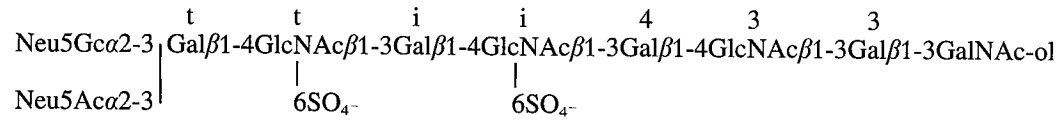
A1.7.4A



A1.7.4B

The ^{1}H-NMR spectrum of fraction A1.6.4 (Fig. 7B; Table 3) indicates the presence of the following two nonasaccharide alditols, being sialylated analogues of compound A1.7.5.

A1.6.4A

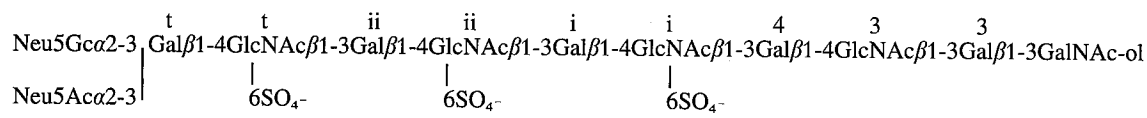


A1.6.4B

In a similar way as discussed for A1.7.4 and A2.3 [28], the extension of A1.7.5 with α2-3-linked Neu5Gc/Ac, yielding A1.6.4A/B, gives rise to the typical structural-reporter-group data for Gal^t H1 and H4 and for Neu5Gc/Ac H3a, H3e and NGc/NAc. The correlation of the diagnostic H1 and H4 signals of Gal^t, Gal^i and Gal^4 (see Table 5), and of the H1, H6 and H6′ signals of GlcNAc^t, GlcNAc^i and GlcNAc^3 was generated by a two-dimensional HOHAHA measurement (Fig. 8B).

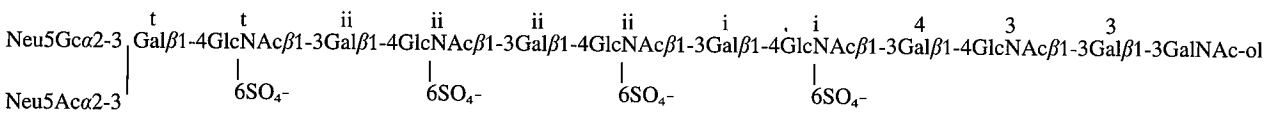
The ^{1}H-NMR spectrum of fraction A1.5.6 (see Table 4) indicates the presence of the sialylated trisulfated undecasaccharide alditols A1.5.6A and A1.5.6B, being extensions of A1.6.5 with α2-3-linked Neu5Gc/Ac. The interpretation of the ^{1}H-NMR data follows the same reasoning as presented for the A1.7.4/A2.3 and A1.6.4/A1.7.5 pairs. Analogously, the ^{1}H-NMR data of fraction A1.4.5 (Table 4) are in accordance with the presence of the sialylated tetrasulfated tridecasaccharide alditols A1.4.5A and A1.4.5B, being sialylated extensions of A1.5.7.

A1.5.6A



A1.5.6B

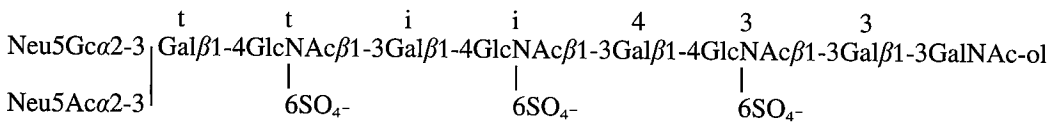
A1.4.5A



A1.4.5B

The ^{1}H-NMR spectrum of fraction A1.6.7 (Fig. 9A; Table 3) indicates the presence of the following monosialylated trisulfated compounds, in which all GlcNAc residues bear a 6-*O*-sulfate group.

A1.6.7A

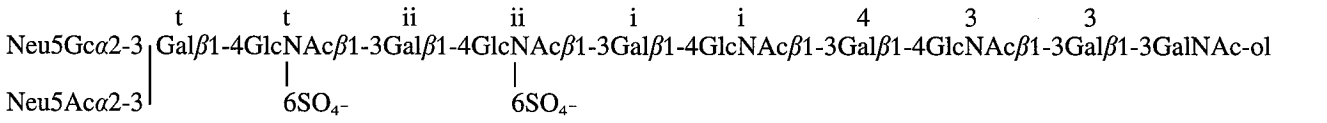


A1.6.7B

The structure of these compounds is deduced by comparing the ^{1}H-NMR spectral data of A1.6.7 with those of fraction A1.6.4, showing downfield shift effects for Gal^3 H4 (δ 4.155; $\Delta\delta$ +0.028), GlcNAc^3 H1 (δ 4.692; $\Delta\delta$ +0.015) and Gal^4 H1 (δ 4.511; $\Delta\delta$ +0.042). These shift increments are characteristic for sulfation of GlcNAc^3 (cf. compounds N4 in Table 1 and A3.3.5 in Table 2; see also Table 5). The interconnection of the Gal^3 H1 and H4, and of the Gal^4/Gal^i H1 and H4 signals is demonstrated by a two-dimensional HOHAHA measurement (Fig. 10A).

^{1}H-NMR analysis of fraction A1.5.3 (Fig. 9B; Table 4) points to the occurrence of the following undecasaccharide alditols with the exceptional feature that GlcNAc^i is not 6-*O*-sulfated. The presence of two sulfated GlcNAc residues instead of three, like in A1.5.6A/B, is supported by the relatively early elution position of fraction A1.5.3 on Mono Q, as compared to A1.5.6.

A1.5.3A

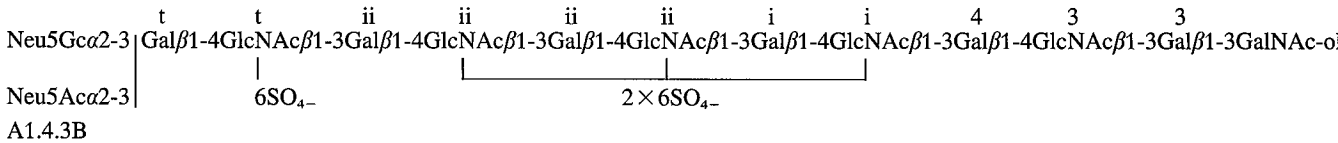


A1.5.3B

The positions of the sulfated and non-sulfated GlcNAc residues in the poly(*N*-acetyllactosamine) sequence of A1.5.3A/B are determined using the two-dimensional HOHAHA connectivities between the diagnostic Gal H1 and H4 signals (Fig. 10B; see Table 5). The Gal^{i} H1 signal at δ 4.605 is connected to H4 at δ 3.968 (for A1.5.3A) or δ 3.963 (for A1.5.3B), indicating the presence of a Neu5Gc/Acα2-3Galβ1-4(6SO$_4$-)GlcNAc element. The Gal^{ii} H1 signal at δ 4.506 shows a cross-peak with the H4 signal at δ 4.190, which points to a Gal residue which is sandwiched between two (6SO$_4$-)GlcNAc residues. One of the overlapping signals at δ 4.466 (Gal^{i} H1) is connected with the corresponding H4 signal at δ 4.190, a combination which is indicative of a (6SO$_4$-)GlcNAcβ1-3Galβ1-4GlcNAc element. The other signal at δ 4.466 (Gal^{4} H1) shows a cross-peak with H4 at δ 4.158. This set of Gal signals demonstrates the occurrence of the GlcNAcβ1-3Galβ1-4GlcNAc sequence. Finally, Gal^{3} H1 at δ 4.459 shows a cross-peak with the H4 signal at δ 4.125. Together, the aforementioned data prove the positions of the (6SO$_4$-)GlcNAc residues in the sequence. The isomeric structure possessing a 6-*O*-sulfated GlcNAc^{i} and a non-sulfated GlcNAc^{ii} residue can be excluded because of the absence of a set of interconnected Gal H1 and H4 signals at δ 4.51 and δ 4.16 (see Table 5).

The ^{1}H-NMR spectrum of fraction A1.4.3 (see Table 4) demonstrates the presence of the following sialylated sulfated tridecasaccharide alditols.

A1.4.3A



A1.4.3B

These carbohydrate chains are identical to A1.4.5A and A1.4.5B, respectively, with the exception that one of the internal GlcNAc residues is not sulfated. This feature is deduced from the presence of a Gal H1 signal overlapping with Gal^{4} H1 at δ 4.465, and a Gal H4 signal at δ 4.156 (see Table 5). Because of the low amount of material, two-dimensional HOHAHA spectroscopy could not be carried out. From the one-dimensional ^{1}H-NMR data, it cannot be concluded which of the internal GlcNAc residues is non-sulfated since it is not clear to which residue each Gal H1 and H4 belongs. In Table 4, the NMR data are assigned to fit the possible structure in which GlcNAc^{i} is nonsulfated.

^{1}H-NMR spectroscopy of the Mono Q subfractions of the Bio-Gel P-6 fractions A1.2 and A1.3 provides only more general structural information. The increasing overlap of the structural-reporter-group signals stemming from the repeating Galβ1-4(6SO$_4$-)GlcNAc units, caused by the increasing length of the poly(*N*-acetyllactosamine) chain, hampers determination of exactly defined structures of the compounds. However, on the basis of the spectral data and the chromatographic behaviour of these compounds, it is suggested that they represent a continuation of the series of oligosaccharide alditols built up from linear sialylated sulfated poly(*N*-acetyllactosamine) chains. Fraction A1.1 was not further fractionated. It is evident from its ^{1}H-NMR spectrum (not shown) that compounds similar to those detected in A1.2 and A1.3, but of larger size, are present. It should be noted that in fraction A1.1 the presence of a small amount of N-glycan-derived alditols was established, indicating that a small portion of the N-linked oligosaccharides was not released by the PNGase-F treatment. A similar observation has been reported previously [11].

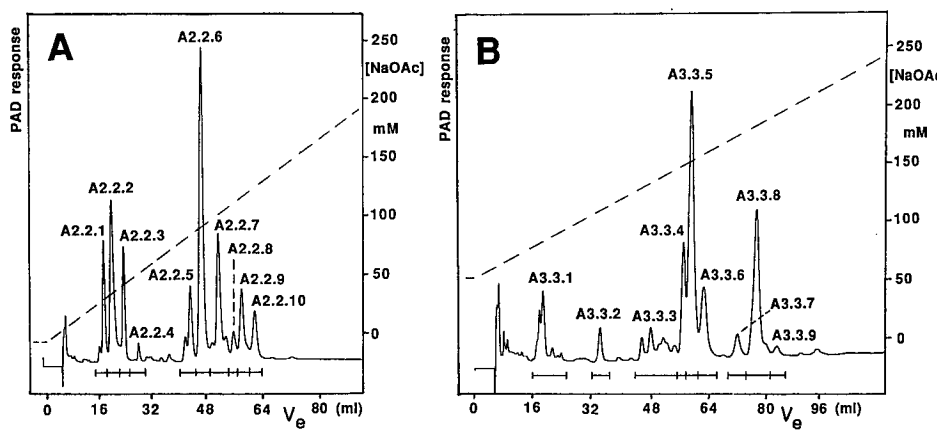


Fig. 6. Elution profiles of Mono Q fractions A2.2 (A) and A3.3 (B) on a CarboPac PA-1 column (25×0.9 cm) using pulsed amperometric detection (PAD). Elutions were carried out at a flow rate of 4 ml/min with a gradient of NaOAc in 0.1 M NaOH as indicated (– – – – –).

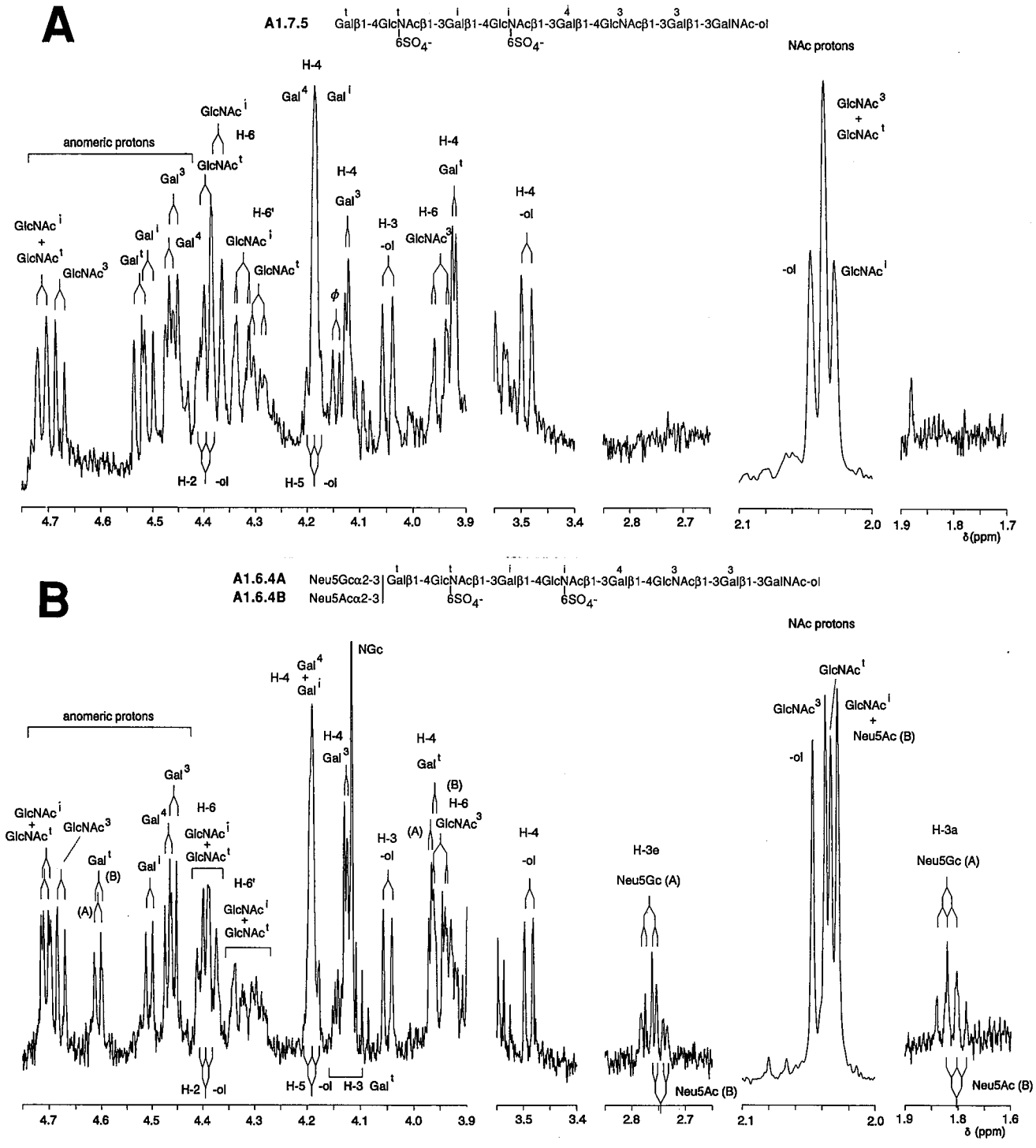


Fig. 7. Structural-reporter-group regions of the resolution-enhanced ^{1}H-NMR spectra of fractions A1.7.5 (A) at 500 MHz and A1.6.4 (B) at 600 MHz of the O-linked oligosaccharide alditols derived from pZP glycoproteins, recorded in ^2H_2O at 22°C. The relative scales of the NAc protons regions differ from those of the rest of the spectra. ϕ denotes non-carbohydrate contamination.

Endo-β-galactosidase digestion of the O-linked carbohydrate chains

A pool of oligosaccharide alditols obtained by alkaline borohydride treatment of PNGase-F-treated pZP was incubated with endo-β-galactosidase, and the digest D_o was fractionated on Bio-Gel P-4 (Fig. 11A). Comparison of the P-4 patterns before (Fig. 3) and after incubation shows that endo-β-galactosidase generates a series of fragments of the O-linked oligosaccharides. The pooled subfractions D_o1-D_o8 were further fractionated by HPLC on Lichrosorb-NH_2. In each case, the collected fractions were pooled as indicated in Fig. 12A–E, and analyzed by ^{1}H-NMR spectroscopy. For fractions not discussed below, the amount of material was too low for structure identification by ^{1}H-NMR spectroscopy, or the material consisted of mixtures that were too complex. Relevant NMR data are given in Table 7. For the inter-

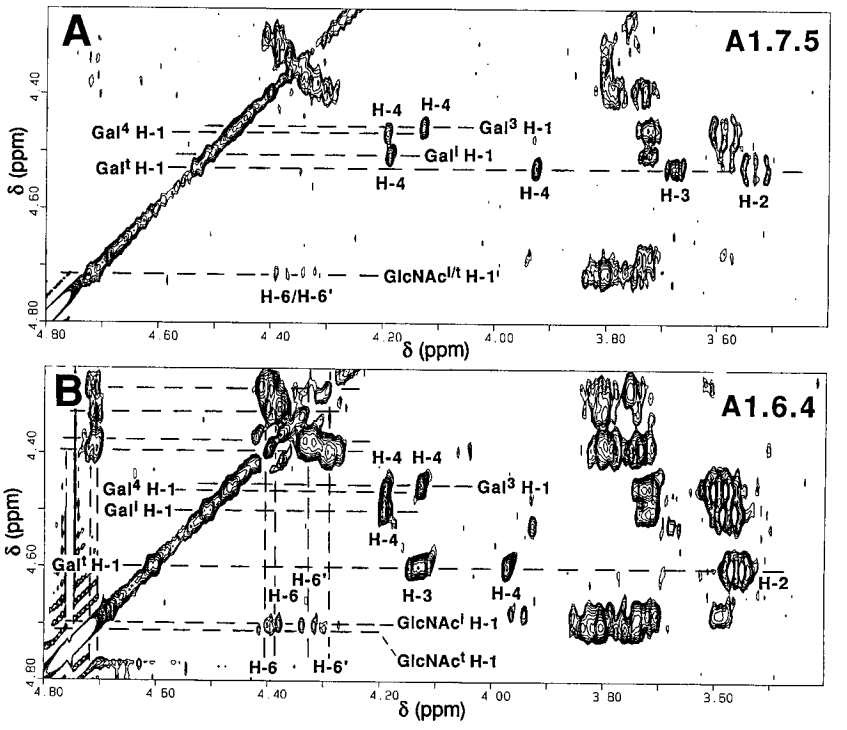


Fig. 8. Relevant regions of the two-dimensional HOHAHA spectra at 500 MHz of fractions A1.7.5 (A) and A1.6.4 (B) of the O-linked oligosaccharide alditols derived from pZP glycoproteins, recorded in ^2H_2O at 22°C. Lines are drawn to show the interconnection between the protons of one residue.

pretation of the results, it should be kept in mind that the action of endo-β-galactosidase on a R-GlcNAcβ1-3Galβ1-4GlcNAc-R′ element results in the formation of fragments which contain a reducing Gal residue. Therefore, all compounds found in the digestion mixture containing a reducing Gal residue can be considered as stemming from the action of endo-β-galactosidase on a poly(*N*-acetyllactosamine) sequence, whereas all alditols can be considered as either intact or endo-β-galactosidase-generated core-element-containing oligosaccharides. An overview of the established structures is given in Table 8.

^{1}H-NMR analysis of the fractions D_o7.3 (cf. compound A in [29]), D_o7.4 [30] (for rectification see [22]), D_o6.3 (cf. compound 3 in [22]), D_o6.4 (cf. compound 16 in [22]), D_o6.6 (cf. compound 12 in [22]) and D_o5.1 (cf. compound B in [29]) shows the occurrence of the following series of carbohydrate chains.

D_o7.3 GlcNAcβ1-3Gal

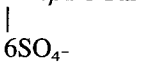
D_o7.4 Galβ1-4GlcNAcβ1-3Gal

D_o6.3 GlcNAcβ1-3GalNAc-ol

D_o6.4 GlcNAcβ1-3Galβ1-3GalNAc-ol

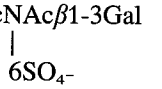
D_o6.6 Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol

D_o5.1



The ^{1}H-NMR spectrum of fraction D_o5.2 indicates the presence of a sulfated trisaccharide.

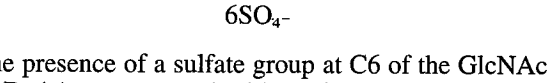
D_o5.2



Comparison of the spectral data of D_o5.2 with those of D_o7.4 shows additional signals that have shifted out of the bulk region at δ 4.397 (GlcNAc H6) and δ 4.313 (GlcNAc H6′), reflecting the presence of a sulfate group at C6 of the GlcNAc residue (cf. compound A2.3; Table 2). This conclusion is corroborated by the specific shifts (considering only the β-anomer of the reducing Gal residue) for GlcNAc H1 (Δδ +0.017), Gal H1 (Δδ +0.044) and Gal(red) H4 (Δδ +0.028) (cf. compounds N7 in Table 1 and A2.3 in Table 2).

The ^{1}H-NMR spectrum of fraction D_o4.1 shows the occurrence of the following sulfated trisaccharide alditol.

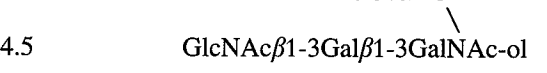
D_o4.1



The presence of a sulfate group at C6 of the GlcNAc residue in D_o4.1 as compared with D_o6.4 results in the appearance of the typical GlcNAc H6 and H6′ signals outside the bulk region (cf. compound D_o5.1). The structural-reporter-group signals of the Galβ1-3GalNAc-ol sequence remain unchanged, except for the Gal H4 signal, which shifted downfield (Δδ +0.031; cf. compounds N7 in Table 1 and A2.3 in Table 2).

The ^{1}H-NMR spectrum of fraction D_o4.5 indicates the presence of a sialylated extension of D_o6.4.

D_o4.5



As compared to the NMR data of D_o6.4, additional Neu5Gc signals are present at δ 1.709 (H3a), δ 2.741 (H3e) and δ 4.121 (NGc) reflecting, in combination with the typical GalNAc-ol structural-reporter-group signals, the occurrence

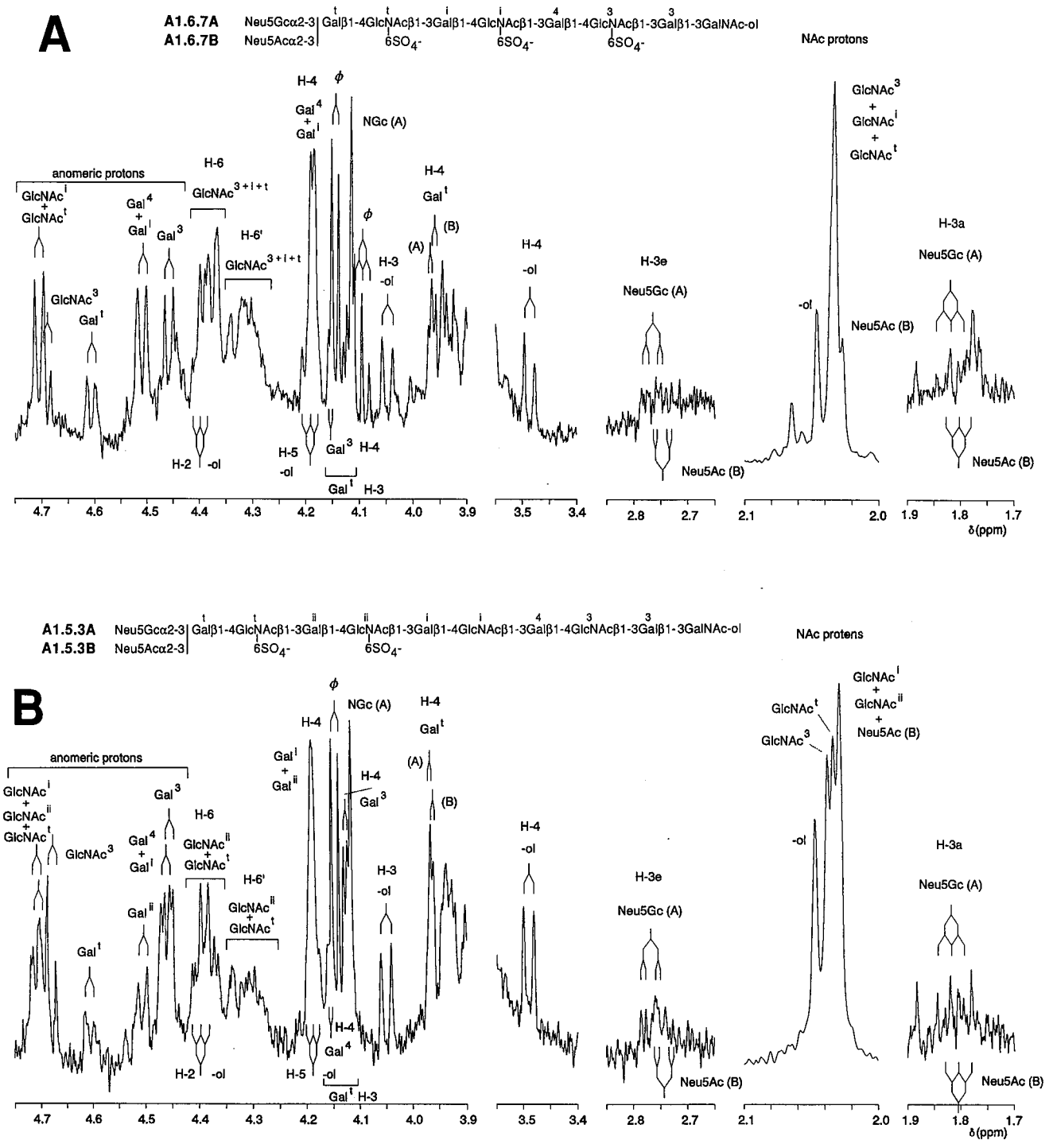


Fig. 9. Structural-reporter-group regions of the resolution-enhanced 500-MHz ^{1}H-NMR spectra of fractions A1.6.7 (A) and A1.5.3 (B) of the O-linked oligosaccharide alditols derived from pZP glycoproteins, recorded in ^2H_2O at 22°C. The relative scales of the NAc protons regions differ from those of the rest of the spectra. ϕ denotes non-carbohydrate contamination.

of a Neu5Gc residue at C6 of GalNAc-ol (cf. compound 9A in [22]).

As shown by their ^{1}H-NMR spectra, fractions $D_o3.2$ and $D_o3.3$ contain the following sialylated sulfated tetrasaccharides, being sialylated extensions of $D_o5.2$.

$D_o3.2$ Neu5Acα2-3Galβ1-4GlcNAcβ1-3Gal
|
$6SO_4$-

$D_o3.3$ Neu5Gcα2-3Galβ1-4GlcNAcβ1-3Gal
|
$6SO_4$-

The structural-reporter-group data of the Neu5Acα2-3Galβ1-4(6SO_4-)GlcNAc element in $D_o3.2$ closely resemble those of the same element in compound A1.7.4B (Table 3). The same holds when comparing the Neu5Gcα2-3Galβ1-4(6SO_4-)GlcNAc elements in $D_o3.3$ and A1.7.4A. The signals of the

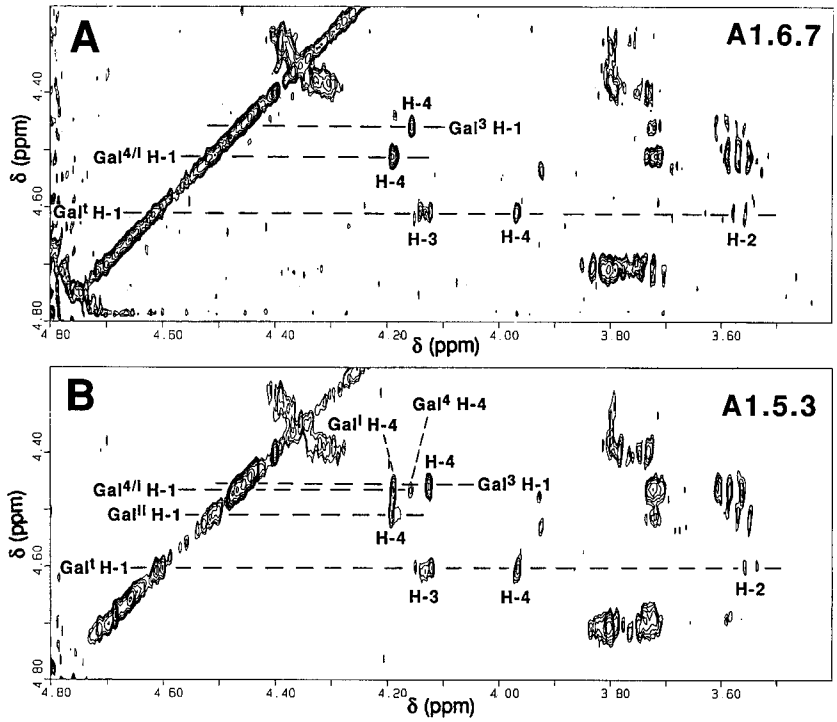


Fig. 10. Relevant regions of the two-dimensional HOHAHA spectra at 500 MHz of fractions A1.6.7 (A) and A1.5.3 (B) of the O-linked oligosaccharide alditols derived from pZP glycoproteins, recorded in ^2H_2O at 22°C. Lines are drawn to show the interconnection between the protons of one residue.

(6SO_4-)GlcNAcβ1-3Gal element in D_O3.3 are assigned by comparing the spectral data of D_O3.2. The structural-reporter-group signals of the Neu5Acα2-3Gal element in D_O3.2 are similar to those in the corresponding alditol in [31] and in the corresponding pyridylamino derivative [16], but the GlcNAc signals differ significantly in both cases.

Endo-β-galactosidase digestion of the N-linked carbohydrate chains

Fractionation of the pool of enzymically released N-linked oligosaccharides on Mono Q (results not shown) yielded a distinct void volume peak, representing the neutral carbohydrate chains, and a broad area of poorly resolved peaks in the 10−500 mM NaCl region, representing the anionic oligosaccharides. The ^{1}H-NMR spectrum (not shown) of the fraction containing the neutral N-linked carbohydrate chains showed the presence of a mixture of di- to tetraantennary oligosaccharides with or without additional *N*-acetyllactosamine units, which is in accordance with studies presented in [10, 14]. ^{1}H-NMR spectroscopic investigation (not shown) of fractions of the anionic carbohydrate chains revealed the presence of an extremely heterogeneous mixture of N-linked oligosaccharides, containing similar sulfated poly(*N*-acetyllactosamine) chains in the branches as present in the O-linked carbohydrate chains.

For comparison of the N- and O-linked carbohydrate chains with respect to their poly(*N*-acetyllactosamine) parts, a pool of the N-linked oligosaccharides was incubated with endo-β-galactosidase. The digest D_N was fractionated on Bio-Gel P-4, and the collected fractions were pooled as indicated in Fig. 11B, yielding seven subfractions, denoted D_N1-D_N7. Comparison of the peak patterns of D_N and D_O suggests that in both cases similar distinct fragments are present. The fractions D_N7, D_N6 and D_N3 elute at the same positions as D_O7, D_O5 and D_O3, respectively, the main fractions of digest D_O. In Figs 13A/B and 14, subfractionation of fractions D_N7, D_N6 (Lichrosorb-NH_2) and D_N3 (Mono Q) are presented. ^{1}H-NMR analysis of the subfractions $D_N7.2$, $D_N7.3$, $D_N6.1$, $D_N6.2$ and $D_N3.3$ showed the occurrence of compounds identical to $D_O7.3$, $D_O7.4$, $D_O5.1$, $D_O5.2$ and $D_O3.2/3.3$, respectively (see Table 8). The Bio-Gel P-4 fractions D_N5, D_N4 and D_N2 contain, according to their ^{1}H-NMR spectra (not shown), complex mixtures of N-glycans. In fraction D_N5 mainly neutral diantennary oligosaccharides occur, whereas fraction D_N4 contains sialylated and/or sulfated diantennary N-glycans. Fraction D_N2 contains an extremely complex mixture of sialylated and/or sulfated tri- or tetraantennary oligosaccharides. No indications for the presence of repeating (sulfated) *N*-acetyllactosamine units were found in fractions D_N5, D_N4 or D_N2, indicating that the endo-β-galactosidase digestion was complete. Because of the complexity of the above-mentioned mixtures, in combination with the low amounts of material, no further subfractionation and structural analysis of D_N5, D_N4 and D_N2 by ^{1}H-NMR spectroscopy was carried out.

DISCUSSION

In this study, the N- and O-linked carbohydrate chains were sequentially released from pZP, and the pool of O-glycans was fractionated using various chromatographic techniques. The structures of 32 O-linked carbohydrate chains were determined using ^{1}H-NMR spectroscopy (see Table 6). The major part of the O-glycans forms a series of (sialylated) sulfated oligosaccharides with a linear poly(*N*-acetyllactos-

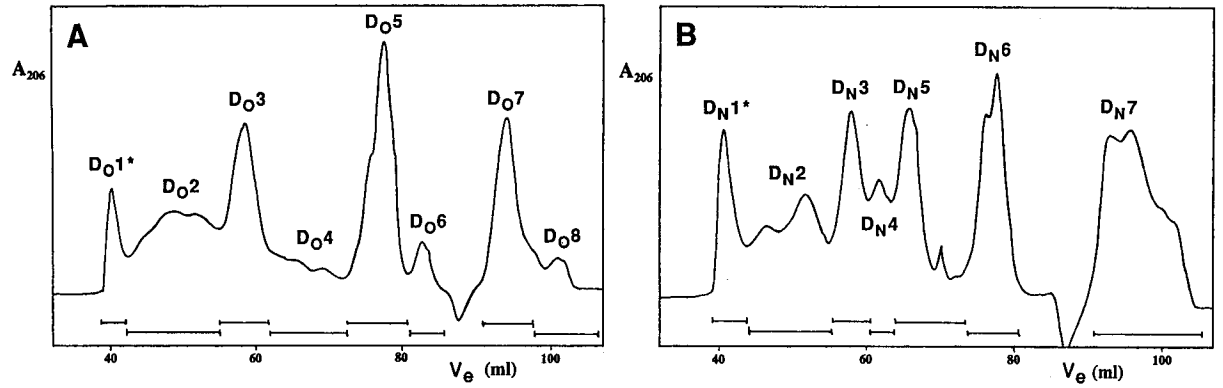


Fig. 11. Elution profiles at 206 nm on a Bio-Gel P-4 column (150×1.15 cm) of the endo-β-galactosidase digests D_O (A) and D_N (B) of the pZP glycoproteins-derived O-linked and N-linked carbohydrate chains, respectively. The column was eluted at 7 ml/h with 100 mM NH_4HCO_3.

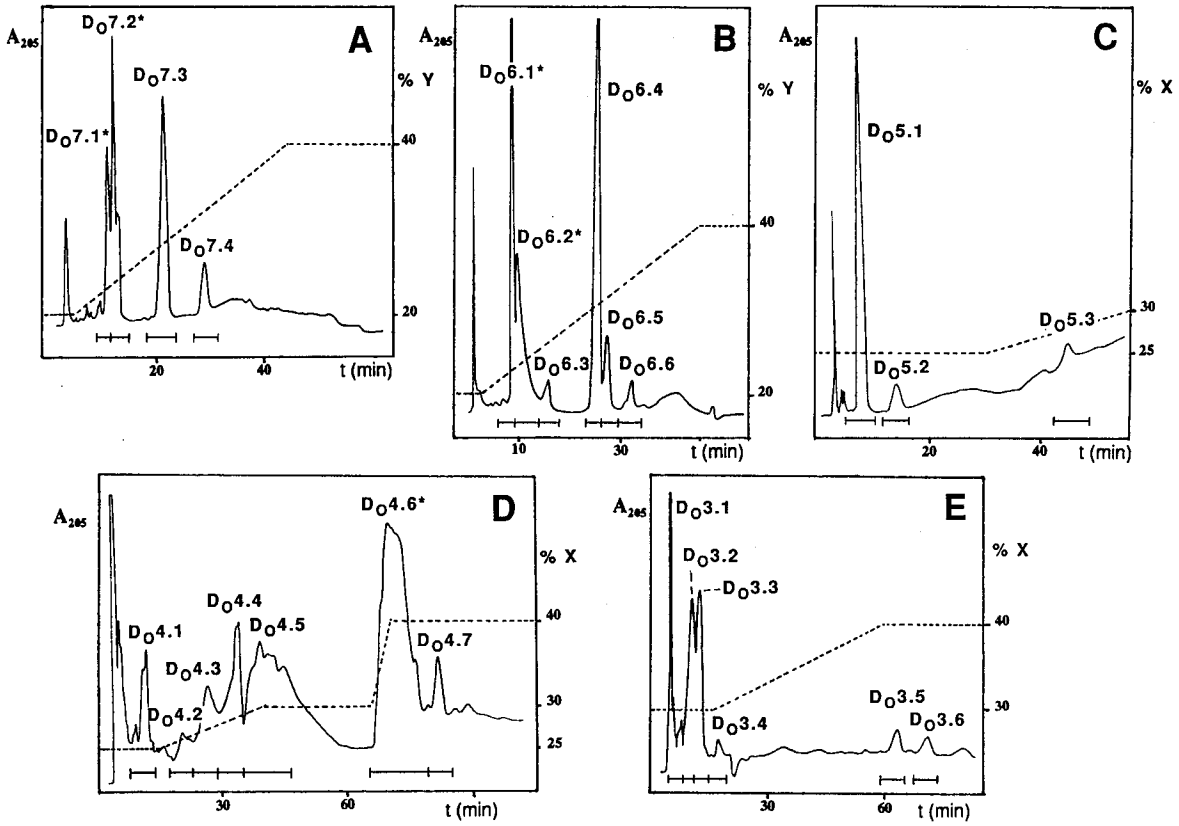


Fig. 12. Fractionation patterns at 205 nm on a HPLC Lichrosorb-NH_2 column (25×0.46 cm) of the Bio-Gel P-4 fractions D_O3–D_O7 of the endo-β-galactosidase digest of the pool of O-linked carbohydrate chains derived from pZP glycoproteins. Elutions were carried out at a flow rate of 2 ml/min using a gradient of 15 mM K_2HPO_4/KH_2PO_4, pH 7.0 (X) or H_2O (Y) in acetonitrile as indicated (------). Fractions marked with * did not contain carbohydrate material. (A) Fraction D_O7; (B) fraction D_O6; (C) fraction D_O5; (D) fraction D_O4; (E) fraction D_O3.

amine) backbone. The core structure is almost exclusively Galβ1-3GalNAc-ol, but a small portion of the identified oligosaccharides contains the GlcNAcβ1-3GalNAc-ol core type. The Galβ1-3GalNAc-ol core is extended at the Gal residue with none to more than seven repetitive Galβ1-4GlcNAcβ1-3 units which in most cases have a sulfate group linked to C6 of the GlcNAc residue. Sialic acid, present as Neu5Gc or Neu5Ac in an overall molar ratio of approximately 2:1, is either α2-3-linked to Gal or α2-6-linked to GalNAc-ol. The α2-6-linked Neu5Gc/Ac residues are not present when the Galβ1-3GalNAc-ol core is extended with more than one *N*-acetyllactosamine unit. Terminal Gal residues predominantly occur in the shorter carbohydrate chains. The longer chains seem to be preferably terminated with Neu5Gc or Neu5Ac. Interestingly, the sialic acids attached to the poly(*N*-acetyllactosamine) chains are exclusively α2-

Table 7. ^{1}H-chemical shifts of structural-reporter-group protons of endo-β-galactosidase digestion products of the O-linked pZP carbohydrate chains. Chemical shifts are given at 22°C and were measured in ^2H_2O relative to internal acetone (δ 2.225 [36]). α and β refer to the anomeric configuration of Gal(red); n.d., not determined.

Residue	Reporter group	Chemical shift in										
		D_o3.2	D_o3.3	D_o5.1	D_o5.2	D_o7.3	D_o7.4	D_o4.1	D_o4.5	D_o6.3	D_o6.4	D_o6.6
		ppm										
GalNAc-ol	H2	–	–	–	–	–	–	4.401	4.386	4.288	4.401	4.400
	H3	–	–	–	–	–	–	4.050	4.047	3.997	4.050	4.051
	H4	–	–	–	–	–	–	3.490	3.534	3.539	3.493	3.493
	H5	–	–	–	–	–	–	4.195	4.241	4.146	4.188	4.187
	NAc	–	–	–	–	–	–	2.047	2.046	2.037	2.048	2.047
Gal(red)	H1α	5.224	5.225	5.226	5.226	5.224	5.225	–	–	–	–	–
	H1β	4.560	4.560	4.562	4.560	4.560	4.560	–	–	–	–	–
	H4α	4.236	4.237	4.232	4.232	4.198	4.203	–	–	–	–	–
	H4β	4.178	4.180	4.176	4.175	4.142	4.147	–	–	–	–	–
	H5α	4.084	4.086	4.085	4.085	4.086	4.086	–	–	–		
Gal^3	H1	–	–	–	–	–	–	4.461	4.461	–	4.462	4.462
	H4	–	–	–	–	–	–	4.155	4.123	–	4.124	4.128
GlcNAc^3	H1α	4.745	4.746	4.726	4.753	4.717	4.739	4.671	4.669	4.605	4.662	4.683
	H1β	4.724	4.725	4.705	4.734	4.695	4.717					
	H6	4.405	4.407	4.336	4.397	3.898	3.957	4.330	n.d.	3.950	3.895	3.952
	H6'α	4.315	4.317	4.229	4.313	n.d.	n.d.	4.223	n.d.	n.d.	n.d.	n.d.
	H6'β	4.315	4.317	4.219	4.313	n.d.	n.d.					
	NAc	2.032	2.032	2.038	2.035	2.039	2.036	2.042	2.046	2.085	2.044	2.041
Gal^4	H1α	4.605	4.609	–	4.529	–	4.480	–	–	–	–	4.479
	H1β	4.601	4.603	–	4.524	–	4.480					
	H3	4.124	4.136	–	n.d.	–	n.d.	–	–	–	–	n.d.
	H4	3.964	3.970	–	3.926	–	3.928	–	–	–	–	3.926
Neu5Ac^3	H3a	1.803	–	–	–	–	–	–	–	–	–	–
	H3e	2.749	–	–	–	–	–	–	–	–	–	–
	NAc	2.028	–	–	–	–	–	–	–	–	–	–
Neu5Gc^3	H3a	–	1.819	–	–	–	–	–	–	–	–	–
	H3e	–	2.769	–	–	–	–	–	–	–	–	–
	NGc	–	4.117	–	–	–	–	–	–	–	–	–
Neu5Gc^6	H3a	–	–	–	–	–	–	–	1.709	–	–	–
	H3e	–	–	–	–	–	–	–	2.741	–	–	–
	NGc	–	–	–	–	–	–	–	4.121	–	–	–

3-linked. Concerning the structures of the neutral O-glycans, the present results are in close agreement with a recent study on these compounds [15].

^{1}H-NMR analyses of mixtures of the anionic N-linked carbohydrate chains of pZP suggest that these contain similar structural elements as found for the O-linked carbohydrate chains, like repeating *N*-acetyllactosamine units, sulfate groups, and α2-3/6-linked Neu5Gc and Neu5Ac residues (see also [10, 16]). It is obvious that combination of variably sialylated/sulfated poly(*N*-acetyllactosamine) chains attached to di- to tetraantennary N-linked oligosaccharides causes extreme heterogeneity. This leads to difficulties in purifying and subsequently analyzing the structure of the intact N-glycans of pZP.

In Table 8, an overview is given of the fragments obtained by endo-β-galactosidase digestion of pools of N-linked and pools of O-linked carbohydrate chains of pZP. Comparison of the two digests shows that the isolated core structures, which may or may not have been affected by the action of the enzyme, form a complex mixture in the case of the N-glycans, whereas in the case of the O-glycans only a few core types are found. All fragments obtained from the O-linked carbohydrate chains which do not contain a core element were also found in the digest of the N-linked carbohydrate chains. Therefore, it is concluded that there are no qualitative differences in the poly(*N*-acetyllactosamine) part of the N- and the O-linked oligosaccharides. On the basis of this conclusion and the observed differences for the receptor activity of the pZP N- and O-glycans, it is tempting to suggest that the carbohydrate epitope recognized by the sperm is located at least partially in the core region of the oligosaccharides. In this context, it is interesting to note that endo-β-galactosidase-digested pZP3 retains its ligand activity [13]. In [16] only two branch fragments were reported to be present in the endo-β-galactosidase digest of the N-glycans of pZP3, namely ($6SO_4$-)GlcNAcβ1-3Gal and Siaα2-3Galβ1-4($6SO_4$-)GlcNAcβ1-3Gal. In addition, we found minor amounts of GlcNAcβ1-3Gal, Galβ1-4GlcNAcβ1-3Gal and

Table 8. Fragments obtained by endo-β-galactosidase digestion of the pools of N- or O-linked carbohydrate chains derived from pZP. In the left column reducing fragments obtained from both N- and O-linked carbohydrate chains are given, whereas in the right column oligosaccharide alditols obtained from the O-linked carbohydrate chains are listed. The fractions $D_O7.4/D_O6.5$ and $D_O5.3/D_O6.6$ contain identical fragments because of peak overlap in the Bio-Gel P-4 fractionation.

Code	Structure	Code	Structure
$D_O7.3$ $D_N7.2$	GlcNAcβ1-3Gal	$D_O6.3$	GlcNAcβ1-3GalNAc-ol
$D_O7.4/D_O6.5$ $D_N7.3$	Galβ1-4GlcNAcβ1-3Gal	$D_O6.4$	GlcNAcβ1-3Galβ1-3GalNAc-ol
$D_O5.1$ $D_N6.1$	GlcNAcβ1-3Gal | $6SO_4^-$	$D_O5.3/D_O6.6$	Galβ1-4GlcNAcβ1-3Galβ1-3GalNAc-ol
$D_O5.2$ $D_N6.2$	Galβ1-4GlcNAcβ1-3Gal | $6SO_4^-$	$D_O4.1$	GlcNAcβ1-3Galβ1-3GalNAc-ol | $6SO_4^-$
$D_O3.2$ $D_N3.3A$	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Gal | $6SO_4^-$		Neu5Gcα2-6 \
		$D_O4.5$	GlcNAcβ1-3Galβ1-3GalNAc-ol
$D_O3.3$ $D_N3.3B$	Neu5Gcα2-3Galβ1-4GlcNAcβ1-3Gal | $6SO_4^-$		

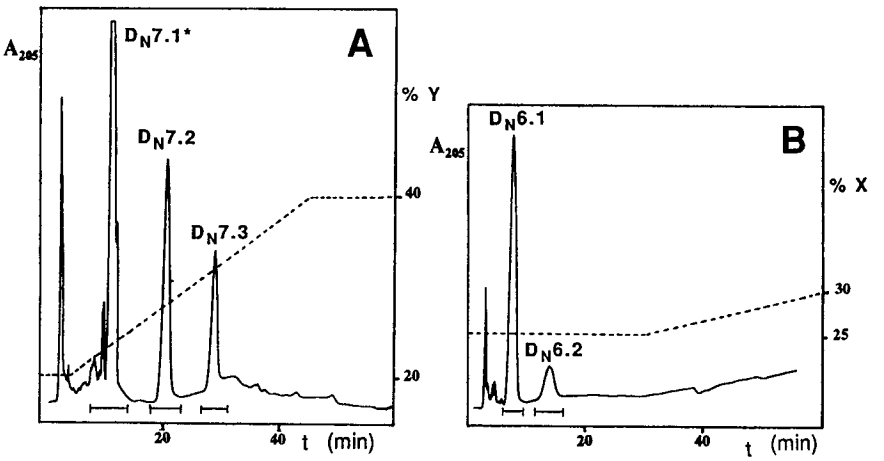


Fig. 13. Fractionation patterns at 205 nm on a HPLC Lichrosorb-NH_2 column (25×0.46 cm) of the Bio-Gel P-4 fractions D_N6 (B) and D_N7 (A) of the endo-β-galactosidase digest of the pool of N-linked carbohydrate chains derived from pZP glycoproteins. Elutions were carried out at a flow rate of 2 ml/min using a gradient of 15 mM K_2HPO_4/KH_2PO_4, pH 7.0 (X) or H_2O (Y) in acetonitrile as indicated. Fractions marked with * did not contain carbohydrate material.

Galβ1-4($6SO_4^-$)GlcNAcβ1-3Gal. The finding of these five different fragments implies that the poly(*N*-acetyllactosamine) chains of the N-glycans of pZP are constructed from repeating Galβ1-4GlcNAc units which are in most cases sulfated, and that they terminate with β1-4-linked Gal or with α2-3-linked Neu5Gc/Ac. No fragments containing a disubstituted Gal residue, indicative of a branched poly(*N*-acetyllactosamine) chain, or a sulfated Gal residue were found. As observed for the O-glycans, the sialic acids in the *N*-acetyllactosamine extensions of the N-glycans are exclusively α2-3-linked. The Siaα2-6Galβ1-4($6SO_4^-$)GlcNAcβ1-3Gal fragment was not found. The results of the endo-β-galactosidase digestion of the O-linked carbohydrate chains corroborate the structural data obtained for the intact oligosaccharides.

Recent observations suggest that the formation of Galβ1-3(GlcNAcβ1-6)GalNAc (core 2) is the key reaction to form poly(*N*-acetyllactosamine) extensions in O-glycans [32]. For example, leukosialin from HL-60 cells, which express the core 2 β1,6-GlcNAc transferase, does contain poly(*N*-acetyllactosamine)-type O-glycans, whereas leukosialin from K562 cells, which lack this particular GlcNAc transferase, does not [33]. In addition, it was found for poly(*N*-acetyllactosamine)-type O-linked carbohydrate chains from leukosialin [34] and human skim milk mucins [35] that the extensions are almost exclusively attached to the GlcNAc residue linked to C6 of GalNAc. The present study shows that, in the porcine system, long poly(*N*-acetyllactosamine) chains can be formed when the β1-6-linked GlcNAc residue is not present, since

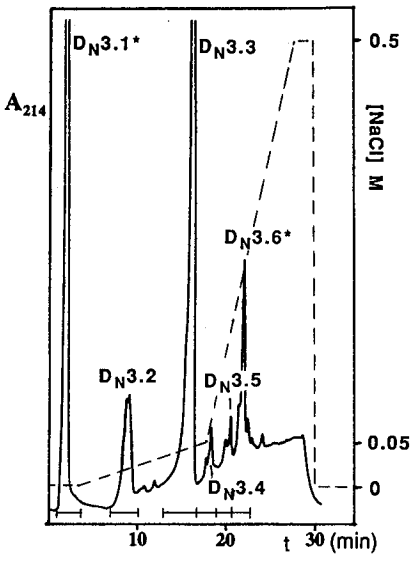


Fig. 14. Fractionation pattern at 214 nm on a FPLC HR 5/5 Mono Q column of Bio-Gel P-4 fraction D_N3 of the endo-β-galactosidase digest of the pool of N-linked carbohydrate chains derived from pZP glycoproteins. The column was eluted with a gradient of NaCl in H_2O as indicated, at a flow rate of 1 ml/min. Fractions marked with * did not contain carbohydrate material.

the poly(*N*-acetyllactosamine) extensions are attached to C3 of Gal in the Galβ1-3GalNAc core type.

The reported study provides a detailed structural analysis of the O-linked carbohydrate chains of pZP and gives additional information on the structure of the N-linked carbohydrate chains. Present studies are aimed at correlating the structure of the pZP carbohydrate chains to their biological activities.

This investigation was supported by the Netherlands Program for Innovation Oriented Carbohydrate Research (IOP-k) with financial aid from the Ministry of Economic Affairs and the Ministry of Agriculture, Nature Management and Fisheries, and Organon International BV (Oss, The Netherlands).

REFERENCES

1. Wassarman, P. M. (1987) *Annu. Rev. Cell. Biol. 3*, 109–142.
2. Wassarman, P. M. (1988) *Annu. Rev. Biochem. 57*, 415–442.
3. Nakano, M., Hatanaka, Y., Sawai, T., Kobayashi, N. & Tobita, T. (1987) *Biochem. Int. 14*, 417–423.
4. Nakano, M., Hatanaka, Y., Kobayashi, N., Noguchi, S., Ishikawa, S. & Tobita, T. (1990) *J. Biochem. (Tokyo) 107*, 144–150.
5. Yurewicz, E. C., Sacco, A. G. & Subramanian, M. G. (1987) *J. Biol. Chem. 262*, 564–571.
6. Araki, Y., Orgebin-Christ, M.-C. & Tulsiani, D. R. (1992) *Biol. Reprod. 46*, 912–919.
7. Florman, H. M. & Wassarman, P. M. (1985) *Cell 41*, 313–324.
8. Bleil, J. D. & Wassarman, P. M. (1988) *Proc. Natl Acad. Sci. USA 85*, 6778–6782.
9. Shivers, C. A. & Dunbar, B. S. (1977) *Science 197*, 1082–1084.
10. Noguchi, S., Hatanaka, Y., Tobita, T. & Nakano, M. (1992) *Eur. J. Biochem. 204*, 1089–1100.
11. Yurewicz, E. C., Pack, B. A. & Sacco, A. G. (1991) *Mol. Reprod. Dev. 30*, 126–134.
12. Yurewicz, E. C., Pack, B. A. & Sacco, A. G. (1992) *Mol. Reprod. Dev. 33*, 182–188.
13. Sacco, A. G., Yurewicz, E. C., Subramanian, M. G. & Matzat, P. D. (1989) *Biol. Reprod. 41*, 523–532.
14. Mori, E., Takasaki, S., Hedrick, J. L., Wardrip, N. J., Mori, T. & Kobata, A. (1991) *Biochemistry 30*, 2078–2087.
15. Hirano, T., Takasaki, S., Hedrick, J. L., Wardrip, N. J., Amano, J. & Kobata, A. (1993) *Eur. J. Biochem. 214*, 763–769.
16. Noguchi, S. & Nakano, M. (1992) *Eur. J. Biochem. 209*, 883–894.
17. Damm, J. B. L., Voshol, H., Hård, K., Kamerling, J. P. & Vliegenthart, J. F. G. (1989) *Eur. J. Biochem. 180*, 101–110.
18. Damm, J. B. L., Kamerling, J. P., Van Dedem, G. W. K. & Vliegenthart, J. F. G. (1987) *Glycoconjugate J. 4*, 129–144.
19. Hård, K., Van Zadelhoff, G., Moonen, P., Kamerling, J. P. & Vliegenthart, J. F. G. (1992) *Eur. J. Biochem. 209*, 895–915.
20. Kamerling, J. P. & Vliegenthart, J. F. G. (1989) in *Clinical biochemistry; principles, methods, applications*, vol. 1, *Mass spectrometry* (Lawson, A. M., ed.) pp. 175–263, Walter de Gruyter, Berlin.
21. Bax, A. & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
22. Kamerling, J. P. & Vliegenthart, J. F. G. (1992) *Biol. Magn. Reson. 10*, 1–194.
23. De Waard, P., Koorevaar, A., Kamerling, J. P. & Vliegenthart, J. F. G. (1991) *J. Biol. Chem. 266*, 4237–4243.
24. Feizi, T., Hounsell, E. F., Alais, J., Veyrieres, A. & David, S. (1992) *Carbohydr. Res. 228*, 289–297.
25. Hokke, C. H., Kamerling, J. P., Van Dedem, G. W. K. & Vliegenthart, J. F. G. (1991) *FEBS Lett. 286*, 18–24.
26. Hokke, C. H., Bergwerff, A. A., Van Dedem, G. W. K., Van Oostrum, J., Kamerling, J. P. & Vliegenthart, J. F. G. (1990) *FEBS Lett. 275*, 9–14.
27. Lloyd, O. K. & Savage, A. (1991) *Glycoconjugate J. 8*, 493–498.
28. Hokke, C. H., Damm, J. B. L., Kamerling, J. P. & Vliegenthart, J. F. G. (1993) *FEBS Lett. 329*, 29–34.
29. Hounsell, E. F., Feeney, J., Scudder, P., Tang, P. W. & Feizi, T. (1986) *Eur. J. Biochem. 157*, 375–384.
30. Herlant-Peers, M.-C., Montreuil, J., Strecker, G., Dorland, L., Van Halbeek, H., Veldink, G. A. & Vliegenthart, J. F. G. (1981) *Eur. J. Biochem. 117*, 291–300.
31. Dickenson, J. M., Huckerby, T. N. & Nieduszynski, I. A. (1992) *Biochem. J. 283*, 267–271.
32. Dennis, J. W. (1993) *Glycobiology 3*, 91–93.
33. Maemura, K. & Fukuda, M. (1992) *J. Biol. Chem. 267*, 24379–24386.
34. Fukuda, M., Carlsson, S. R., Klock, J. C. & Dell, A. (1986) *J. Biol. Chem. 261*, 12796–12806.
35. Hanisch, F.-G., Uhlenbruck, G., Peter-Katalinic, J., Egge, H., Dabrowski, J. & Dabrowski, U. (1989) *J. Biol. Chem. 264*, 872–883.
36. Vliegenthart, J. F. G., Dorland, L. & Van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem. 41*, 209–374.