Eur. J. Biochem. *127*, 7–20 (1982)

Primary-Structure Determination of Fourteen Neutral Oligosaccharides Derived from Bronchial-Mucus Glycoproteins of Patients Suffering from Cystic Fibrosis, Employing 500-MHz ^{1}H-NMR Spectroscopy

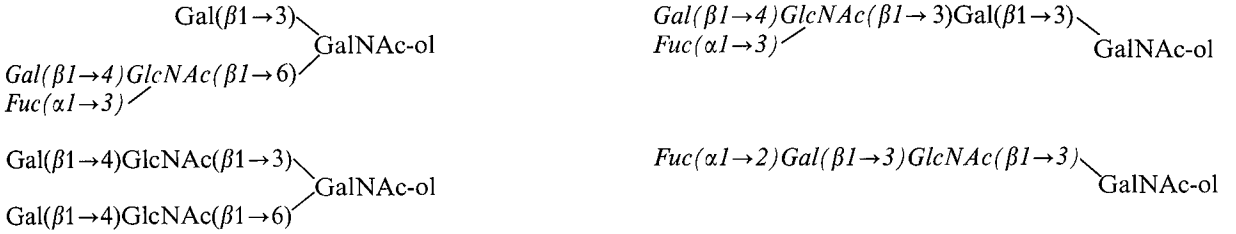
Herman VAN HALBEEK, Lambertus DORLAND, Johannes F. G. VLIEGENTHART, William E. HULL, Geneviève LAMBLIN, Michel LHERMITTE, Arnold BOERSMA, and Philippe ROUSSEL

Department of Bio-Organic Chemistry, University of Utrecht; Bruker Analytische Messtechnik, GmbH, Rheinstetten; and Unité des Protéines, Institut National de la Santé et de la Recherche Médicale 16, Lille

(Received May 7, 1982)

The structure of carbohydrate units of bronchial-mucus glycoproteins obtained from cystic fibrosis patients was investigated by 500-MHz ^{1}H-NMR spectroscopy and methylation analysis. To that purpose, the mucin was subjected to alkaline borohydride degradation. Neutral oligosaccharide-alditols, ranging in size from disaccharides to pentasaccharides, were isolated. Eight compounds could be purified to homogeneity; furthermore, three fractions were obtained consisting mainly of two components. For all 14 compounds the primary structure could be elucidated. 500-MHz ^{1}H-NMR spectroscopy was found to be effective in detecting heterogeneity and to be invaluable for the determination of structures in mixtures of oligosaccharide-alditols.

The structures can be divided into two groups depending on the core disaccharide. One group contains Gal(β1→3)GalNAc-ol as common structural element, the other GlcNAc(β1→3)GalNAc-ol. Both disaccharides were identified as such; the other compounds can be conceived as extensions thereof. The most complex representatives of the two groups are:



The italicized structural elements, comprising the SSEA-1 determinant and the type-1 blood-group-H determinant, are novel sequences in oligosaccharide chains of mucin-type glycoproteins.

Tracheobronchial mucus is a principal component of the mucociliary system which plays an important role in the defense of the respiratory tract. The visco-elastic properties of bronchial mucus, essential for the efficiency of the mucociliary system, are derived from its glycoprotein components which are large macromolecules with a very high carbohydrate content. In cystic fibrosis (CF), overproduction of bronchial-mucus glycoproteins and changes in their composition may alter the visco-elastic properties of the mucus [1] thereby leading to the development of a chronic obstructive lung disease which increases the gravity of affection. Therefore, it is important to determine the primary structure of CF bronchial glycoproteins in order to be able to compare them with normal bronchial glycoproteins and with glycoproteins isolated from other bronchial hypersecretions.

Bronchial glycoproteins have already been shown to be very heterogeneous with regard to acidity and molecular size of their carbohydrate chains [2–6]. The latter chains are glycosidically linked to the hydroxyl groups of serine or threonine residues of the protein core. A similar heterogeneity of carbohydrate chains has been found in several other mucus glycoproteins [7–13].

Here, we report the isolation, purification and primary structure determination of 14 neutral oligosaccharide-alditols from CF mucus glycoproteins. For primary structure analysis, high-resolution ^{1}H-NMR spectroscopy is the method of choice. This technique has proved to be able to provide, in a non-destructive way, the primary structure of the oligosaccharide chains released from a glycoprotein, both for glycopeptides and oligosaccharides of the *N*-glycosidic type as well as for oligosaccharide-alditols derived from *O*-glycosidic chains. For comprehensive reviews, see [14–16]. The NMR method can often cope with the frequently occurring

Dedicated to Prof. Dr J. Kistemaker on the occasion of his 65th birthday.

Abbreviations. Fuc, L-fucose; Gal, D-galactose; GlcNAc, *N*-acetyl-D-glucosamine; GalNAc, *N*-acetyl-D-galactosamine; GalNAc, *N*-acetyl-D-galactosaminitol; Asn, L-asparagine; NMR, nuclear magnetic resonance; CF, cystic fibrosis.

(micro)heterogeneity of the carbohydrate chains [17]. At 500 MHz, 25 nmol of carbohydrate is sufficient for structural analysis [18]. A preliminary account of the results obtained on the employment of 500-MHz ^{1}H-NMR spectroscopy for the structural elucidation of the CF bronchial-mucin oligosaccharide chains has previously been presented [19].

MATERIALS AND METHODS

Isolation and Purification of the Oligosaccharides

A pool of oligosaccharides was used, which was obtained by alkaline borohydride reductive degradation of human acidic bronchial glycoproteins from six patients, with blood-group-0 activity, suffering from cystic fibrosis. Experimental details have been published previously [20]. The oligosaccharides were fractionated by ion-exchange chromatography on Dowex AG1 × 2 according to charge (acidity), and by gel-filtration on Bio-Gel P4 (100 – 200 mesh) according to molecular size. Fraction Ic, being a mixture of small, neutral reduced carbohydrate chains [5], was used for further investigation.

The neutral oligosaccharide fraction Ic (30 mg) was subsequently fractionated on a DAX4 column, as described before [6], resulting in 12 oligosaccharide fractions. Seven major fractions were purified by paper chromatography (Whatman 3MM paper), as earlier described [6], with the solvent system 1-butanol/pyridine/water (6:4:3, v/v) for 32 h. Purified oligosaccharides were eluted from the paper with 0.1 M acetic acid and lyophilized.

To improve the purity of the eluted oligosaccharides for NMR analysis, each fraction was submitted to high-performance liquid chromatography, essentially as described in [21]. The column used was a 10 µ microbondapak-carbohydrate (30 × 0.46 cm internal diameter, Waters Associates).

Analytical Methods

The molar ratios of neutral sugars and *N*-acetylhexosamines in the purified bronchial-mucin oligosaccharide-alditols were determined by gas-liquid chromatography of their trimethylsilyl derivatives, according to the slightly modified [3] procedure of Reinhold [22].

Periodate oxidation of the disaccharides was performed according to Lee and Socca [6,23].

Methylation of the oligosaccharide-alditols was carried out according to Hakomori [24]. Partially *O*-methylated alditol acetates and *N*-methylated *N*-acetylhexosaminitol acetates were prepared according to Björndal et al. [25] and identified by gas-liquid chromatography/mass spectrometry [26].

Prior to ^{1}H-NMR spectroscopic analysis, the oligosaccharide-alditol fractions were repeatedly treated with ^2H_2O at room temperature, with intermediate lyophilization. Finally, the samples were redissolved in 0.4 ml ^2H_2O (99.96 atom% ^{2}H, Aldrich, Milwaukee, WI, USA).

500-MHz ^{1}H-NMR spectroscopy was performed on a Bruker WM-500 spectrometer (Rheinstetten, FRG) operating in the Fourier-transform mode and equipped with a Bruker Aspect-2000 computer. The spectra were obtained using a 70° pulse width, and taken up in 16 K data points with an acquisition time of 3.28 s and a spectral width of 2500 Hz. Resolution enhancement was achieved by Lorentzian to Gaussian transformation [27] from quadrature phase detection, followed by employment of a 32 K complex Fourier transformation. In general, a few hundreds of acquisitions for each sample were accumulated. The indicated probe temperature was 300 K and was kept constant within 0.1 K.

The chemical shifts are expressed in ppm downfield from internal sodium 4,4-dimethyl-4-silapentane-1-sulfonate (DSS) but were actually measured by reference to internal (≈ 1 µl) acetone (δ = 2.225 ppm) with an accuracy of 0.002 ppm.

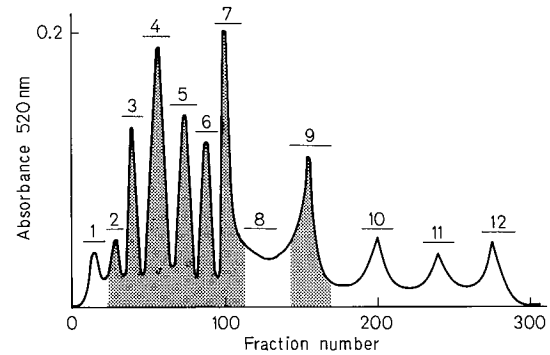


Fig. 1. *Elution profile of the fractionation of neutral bronchial oligosaccharide-alditols Ic (30 mg) on an anion-exchange chromatography DAX4 column (0.9 × 30 cm).* Elution was with 0.2 M ammonium borate buffer, pH 8.0; 2-ml fractions were collected [6]. The fractions indicated by bars were pooled, lyophilized and used for structural investigations

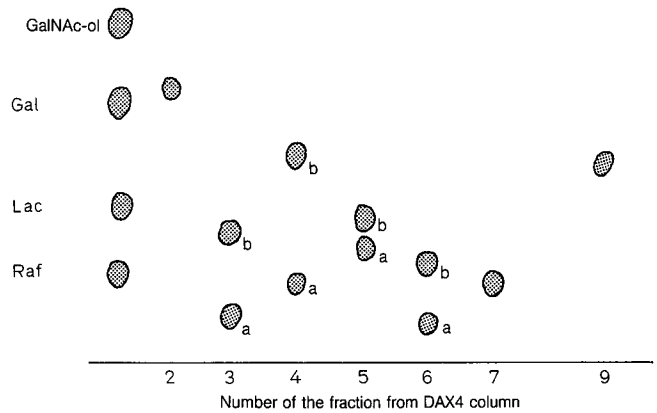


Fig. 2. *Paper chromatography of neutral bronchial oligosaccharide-alditol fractions, eluted from the anion-exchanger (DAX4), on Whatman 3MM in 1-butanol/pyridine/water, 6:4:3 (v/v), for 32 h.* Standards used: Raf (raffinose), Lac (lactose), Gal, and GalNAc-ol

RESULTS

Purification and Carbohydrate Composition of the Oligosaccharide-Alditols

Fraction Ic was separated into 12 oligosaccharide-alditol fractions, denoted 1 – 12, by chromatography over a DAX4 anion-exchanger (see Fig. 1). Further purification by preparative paper chromatography (see Fig. 2) yielded for some fractions two subfractions, designated a and b. The amounts of material of fractions 1, 8, 10, 11 and 12 obtained after preparative paper chromatography, were too low to permit structural determination.

To remove contaminants, the remaining fractions were submitted to high-performance liquid chromatography, with detection at 206 nm. The eluent mixture, acetonitrile/water, was removed by evaporation under reduced pressure. The

Table 1. *Characteristics of the major, neutral oligosaccharide-alditol fractions obtained from cystic fibrosis bronchial mucin*
The molar carbohydrate composition of the oligosaccharide-alditols was calculated on the basis of one residue of GalNAc-ol/molecule

Oligosaccharide-alditol reaction	Yield	$R_{GalNAc-ol}$	Molar composition				Estimated molecular size
			Fuc	Gal	GlcNAc	GalNAc-ol	
	µg		mol/mol oligosaccharide				
2	196	0.80	0	0	0.8	1.0	disaccharide
3a	319	0.12	1.2	2.2	1.2	1.0	pentasaccharide
3b	449	0.38	0	1.2	0.9	1.0	trisaccharide
4a	880	0.20	0	1.2	1.6	1.0	tetrasaccharide
4b	310	0.60	0	1.0	0	1.0	disaccharide
5a	300	0.34	0	1.1	0.8	1.0	trisaccharide
5b	700	0.41	0	1.2	0.8	1.0	trisaccharide
6a	454	0.10	0.2	1.8	1.6	1.0	pentasaccharide
6b	423	0.28	1.0	1.1	0.8	1.0	tetrasaccharide
7	1530	0.21	0	2.0	0.8	1.0	tetrasaccharide
9	158	0.59	1.0	1.1	0	1.0	trisaccharide

Table 2. *Molar ratios of partially* O*-methylated monosaccharide-alditol acetates present after methylation analysis of nine neutral bronchial oligosaccharide-alditol fractions*
2,3,4-Me_3-Fuc = 1,5-di-*O*-acetyl-2,3,4-tri-*O*-methyl-fucitol; 1,4,5,6-Me_4-GalNAcNMe = 3-*O*-acetyl-1,4,5,6-tetra-*O*-methyl-*N*-acetyl-*N*-methyl-galactosaminitol, etc. The molar ratios were determined on the basis of one residue of 1,4,5,6-Me_4-GalNAcNMe or 1,4,5-Me_3-GalNAcNMe

Oligosaccharide-alditol fraction	Partially methylated monosaccharides									
	2,3,4-Me_3-Fuc	2,3,4,6-Me_4-Gal	2,4,6-Me_3-Gal	3,4,6-Me_3-Gal	3,4,6-Me_3-GlcNAc-NMe	3,6-Me_2-GlcNAc-NMe	4.6-Me_2-GlcNAc-NMe	6-Me-GlcNAcNMe	1,4,5,6-Me_4-GalNAc-NMe	1,4,5-Me_3-GalNAc-NMe
	mol/mol oligosaccharide									
3a	0.95	1.2	1.1	0	0.2	0	0	0.5	**1.0**	0.1
3b	0	1.1	0	0	0	0.85	0	0	**1.0**	0
4a	0	1.3	0.3	0	0.8	0.75	0	0	0.15	**1.0**
5a	0	1.1	0	0	0.8	0	0	0	0	**1.0**
5b	0	1.1	0	0	0	0	0.8	0	**1.0**	0
6a	0.3	2.1	0	0	0	1.1	0	traces	0	**1.0**
6b	1.0	0	0	1.2	0	0	0.8	0	**1.0**	0
7	0	2.1	0	0	0	0.6	0	0	0	**1.0**
9	0.8	0	0	1.1	0	0	0	0	**1.0**	0

yields, the chromatographic characteristics (migration relative to GalNAc-ol, $R_{GalNAc-ol}$) and the molar carbohydrate compositions of the (sub)fractions treated in this way, i.e. 2–7 and 9, are summarized in Table 1.

Periodate Oxidation

From their carbohydrate compositions it is evident that fractions 2 and 4b contain disaccharide-alditols, consisting of GlcNAc or Gal, respectively, in addition to GalNAc-ol (see Table 1). For both fractions 2 and 4b, periodate oxidation led to the formation of *N*-acetylthreosaminitol, which was characterized by gas-liquid chromatography [6]. This indicates a (1→3)-type of glycosidic linkage in both disaccharides.

Permethylation Studies

The molar ratios of the various partially *O*-methylated monosaccharide derivatives obtained on hydrolysis, reduction, and acetylation of the permethylated, remaining oligosaccharide-alditol fractions are given in Table 2. From the substitution pattern of the GalNAc-ol derivatives it can be concluded that this residue in the oligosaccharide-alditols 3b, 5b, 6b, and 9 is monosubstituted, namely, at C-3, whereas in the compounds 5a, 6a, and 7 both C-3 and C-6 of GalNAc-ol are involved in glycosidic linkages. For 3a and 4a, monosubstituted as well as disubstituted GalNAc-ol residues are found; therefore, the latter fractions are mixtures of, at least, two oligosaccharide-alditols.

Fractions 3b, 5a, and 5b are trisaccharide-alditols containing equal amounts of Gal, GlcNAc, and GalNAc-ol (see Table 1). Methylation analysis (Table 2) points to linear structures for 3b and 5b, which both contain Gal in terminal position. They differ only in the type of linkage between Gal and GlcNAc. This affords the sequences of 3b and 5b to be as follows: Gal(1→4)GlcNAc(1→3)GalNAc-ol and Gal-(1→3)GlcNAc(1→3)GalNAc-ol, respectively. Compound 5a is a branched trisaccharide-alditol containing Gal and GlcNAc both in terminal position. Two alternative structures have to

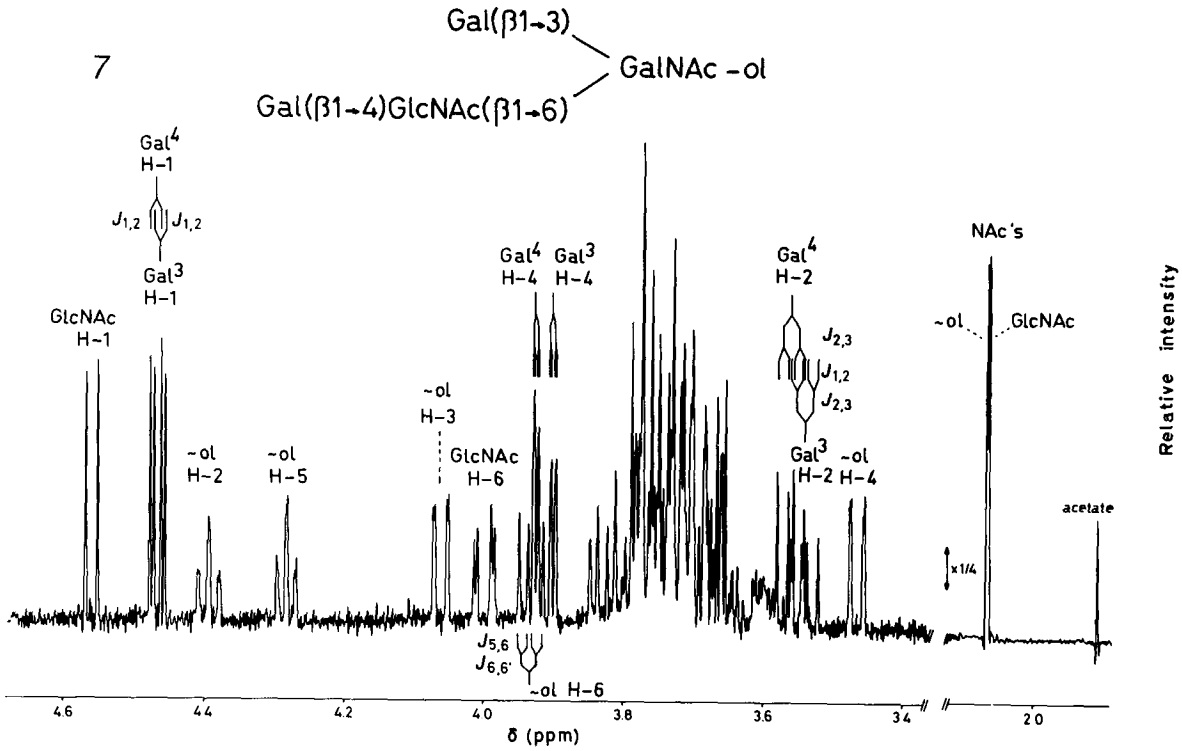


Fig. 3. *Resolution-enhanced 500-MHz ^{1}H-NMR spectrum of bronchial oligosaccharide-alditol fraction 7, possessing the Gal(β1→3)GalNAc-ol core unit, in ^2H_2O at 300 K.* Splitting patterns of signals are indicated in cases of possible obscurity. The relative-intensity scale of the *N*-acetyl proton region of the spectrum differs from that of the other part, as indicated

be considered, namely: Gal(1→3)[GlcNAc(1→6)]GalNAc-ol and GlcNAc(1→3)[Gal(1→6)]GalNAc-ol.

Fraction 9 consists of a Fuc-containing trisaccharide-alditol (see Table 1). Methylation analysis points to the occurrence of Fuc in terminal position and to substitution of Gal at C-2 (see Table 2). Therefore, compound 9 possesses the following structure: Fuc(1→2)Gal(1→3)GalNAc-ol.

For the apparently pure tetrasaccharide-alditols 6b and 7, unambiguous sequences cannot be proposed merely on the basis of the data of sugar and methylation analyses (Tables 1 and 2). Compound 6b is a linear tetrasaccharide-alditol having Fuc in terminal position. Two structural alternatives need to be considered, namely: Fuc(1→2)Gal(1→3)GlcNAc(1→3)-GalNAc-ol, and Fuc(1→3)GlcNAc(1→2)Gal(1→3)GalNAc-ol. The branched compound 7 contains two Gal residues, both in terminal position; one has to be linked directly to GalNAc-ol, the other is attached to C-4 of GlcNAc. Analogous to compound 5a, two structures can be suggested for compound 7, namely, Gal(1→3)[Gal(1→4)GlcNAc(1→6)]GalNAc-ol, and Gal(1→4)GlcNAc(1→3)[Gal(1→6)]GalNAc-ol.

For the oligosaccharide-alditol fractions 3a, 4a, and 6a, the situation is more complicated. The results of sugar analysis (Table 1) seem to indicate that fraction 4a consists, at least mainly, of a tetrasaccharide-alditol, whereas fractions 3a and 6a contain pentasaccharides or even larger oligosaccharide-alditols. For fraction 6a, the relatively low amount of Fuc (Table 1) suggests the presence of at least two components in this fraction, a major one without Fuc, and a minor, fucosylated one. As methylation analysis afforded two differently substituted GalNAc-ol derivatives for both fractions 3a and 4a (see Table 2), they are in fact mixtures of linear and branched oligosaccharide-alditols.

500-MHz ^{1}H-NMR SPECTROSCOPIC STUDIES

500-MHz ^{1}H-NMR spectra were recorded of the 11 bronchial-mucin oligosaccharide-alditol fractions obtained after purification by high-performance liquid chromatography. The eight compounds that appeared to be relatively pure on the basis of carbohydrate composition and methylation analyses (*vide supra*) gave rise to considerably less-complex spectra than did the three fractions known to be mixtures of oligosaccharide-alditols. The difference in complexity of the spectra, directly related to the degree of purity of the fractions, suggests that the discussion of the spectra be divided into two sections.

Fractions Consisting of One Identifiable Oligosaccharide-Alditol

The 500-MHz ^{1}H-NMR spectra of fractions 4b, 9, 5a, and 7 each show the characteristically-shaped H-3 signal of GalNAc-ol at $\delta \approx 4.39$ ppm. (The resonance owes its typical symmetric pattern to the combination of coupling constants, namely, $J_{1,2} \approx 7.0$ Hz, $J_{1',2} \approx 7.8$ Hz, and $J_{2,3} \approx 1.4$ Hz.) As a typical example, the spectrum of 7 is depicted in Fig. 3. However, compounds 2, 3b, 5b, and 6b give rise to a GalNAc-ol H-2 signal at $\delta \approx 4.29$ ppm (the spectrum of the latter fraction is given in Fig. 4). Although detailed ^{1}H-NMR features of various mucin-type oligosaccharide-alditols have been described [8, 15, 16, 28–31], the latter value has not been reported for the chemical shift of GalNAc-ol H-2, so far. For these literature compounds, possessing Gal(β1→3)GalNAc-ol as a common element, H-2 of GalNAc-ol has been found to resonate at $\delta \approx 4.39$ ppm, regardless of the type of extension of the core, either at Gal or at GalNAc-ol or at both

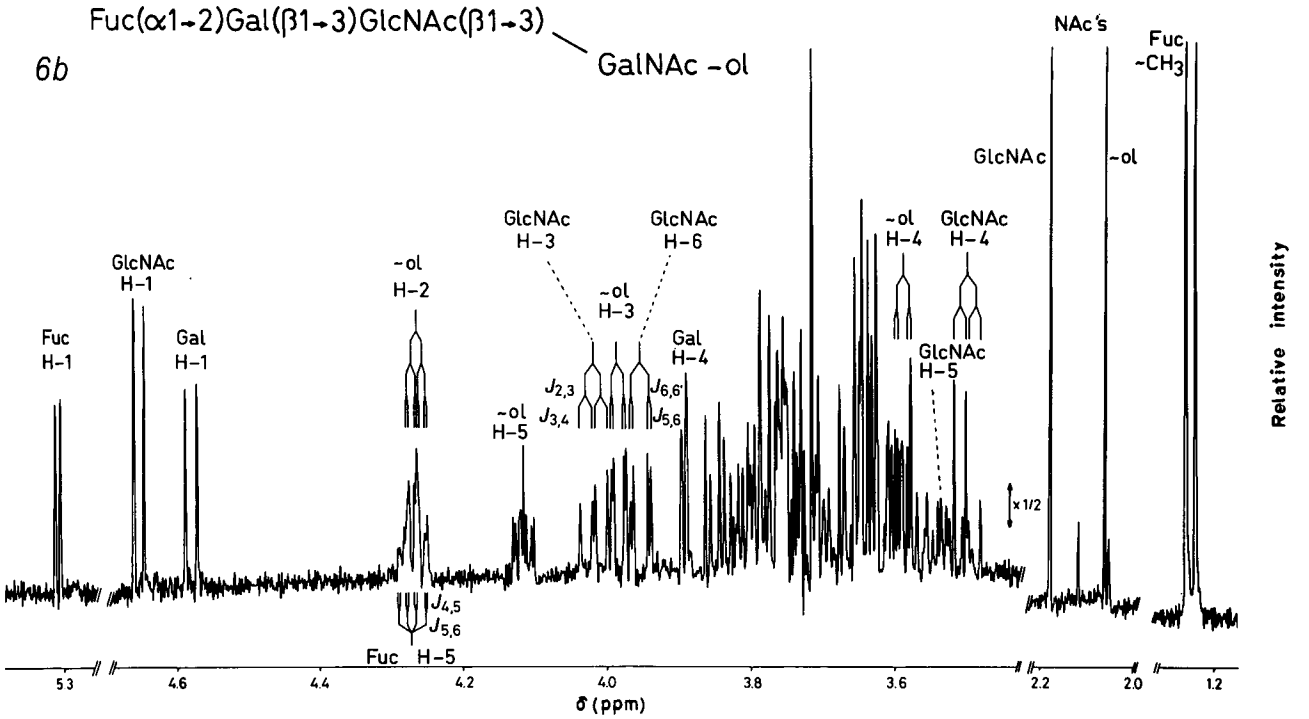


Fig. 4. *Resolution-enhanced 500-MHz ^{1}H-NMR spectrum of bronchial oligosaccharide-alditol fraction 6b, possessing the GlcNAc(β1→3)GalNAc-ol core unit, in ^2H_2O at 300 K.* Splitting patterns of signals are indicated in cases of possible obscurity. The relative-intensity scale of the *N*-acetyl proton region of the spectrum differs from that of the other parts, as indicated

residues. On the basis of the chemical shift for their GalNAc-ol H-2 signal, the eight oligosaccharide-alditols just mentioned are classified into two groups of four compounds. According to this division, the ^{1}H-NMR parameters of the pertinent structural-reporter groups [14–16] of the compounds are compiled in Tables 3 and 4.

Compound 4b can be readily identified as Gal(β1→3)-GalNAc-ol, since the chemical shifts of its structural-reporter groups (see Table 3) are identical to those, acquired at 360 MHz, for the neutral disaccharide-alditol derived from cow milk κ-casein [28] or from hog submaxillary mucin [8]. The value of the coupling constant $J_{1,2}$ of Gal (7.8 Hz) is indicative of the β-glycosidic linkage. Analogously, the 500-MHz ^{1}H-NMR spectral features of compound 9 (see Table 3), including the coupling constants (e.g. $J_{1,2}$ = 7.8 Hz for Gal; $J_{1,2}$ = 3.8 Hz for Fuc), match those acquired at 360 MHz for Fuc(α1→2)Gal(β1→3)GalNAc-ol, obtained from hog submaxillary mucin [8, 15].

The β-configuration of the (1→3)-glycosidic linkage between GlcNAc and GalNAc-ol in the disaccharide-alditol 2 is evident from the chemical shift of GlcNAc H-1 (δ = 4.604 ppm) (see Table 4) in conjunction with its $J_{1,2}$ value (8.4 Hz). Similar values, obtained at 100 MHz (85 °C), have been published for H-1 of GlcNAc in the disaccharide GlcNAc-(β1→3)GalNAc-ol isolated from rat-intestinal mucus glycoproteins [32]. However, as mentioned briefly above, the most striking feature of the spectrum, comparing it with those discussed so far, is the position of the GalNAc-ol H-2 signal: δ = 4.287 ppm. Apparently, this value reflects the presence of β-linked GlcNAc, instead of Gal, as substituent at C-3 of GalNAc-ol. The sensitivity of its chemical shift to this structural feature makes H-2 of GalNAc-ol the ideal reporter for discrimination between the two alternative types of core unit known to occur in mucin-type oligosaccharide-alditols [6–12]. This will be shown for the remaining series of bronchial-mucin fractions.

Also the pattern of the GalNAc-ol H-2 signal for compound 2 differs from that for compounds 4b and 9; in the spectrum of compound 2 it is a well-resolved doublet of triplets ($J_{1,2}$ = $J_{1',2}$ = 7.3 Hz; $J_{2,3}$ = 1.5 Hz). It should be mentioned that this difference in pattern shapes is observable only at highest magnetic-field strength. Analogous to H-2, most of the structural-reporter groups of GalNAc-ol, including its *N*-acetyl methyl protons, show an upfield shift in the step from compound 4b to 2 (see Tables 3 and 4). The well-resolved signals from the skeleton protons of the terminal β-linked GlcNAc residue could be assigned (δ H-3 = 3.584 ppm; δ H-4 = 3.445 ppm; δ H-5 = 3.475 ppm; δ H-6 = 3.950 ppm, outside of the bulk; δ H-6′ = 3.752 ppm), by comparison with those of GlcNAc in GlcNAc(β1→N)Asn [14, 33]. The *N*-acetyl signal at δ = 2.085 ppm is ascribed to GlcNAc, because the chemical shift of this singlet turns out to be sensitive to structural extensions of compound 2 at GlcNAc, in contrast to the signal at 2.037 ppm (*vide infra*).

The spectral data for compound 3b (see Table 4) point to the occurrence of the GlcNAc(β1→3)GalNAc-ol core (for GalNAc-ol, δ H-2 = 4.290 ppm). The chemical shift values for the Gal structural-reporter groups (δ H-1 = 4.455 ppm; δ H-4 = 3.926 ppm) and the value of $J_{1,2}$ (7.85 Hz) demonstrate its terminal position (cf. Table 2) in the trisaccharide, in β-glycosidic linkage. The shift effects introduced by the attachment of Gal to GlcNAc as compared to compound 2, namely, Δδ = 0.027 ppm for H-1 of GlcNAc ($J_{1,2}$ = 7.95 Hz), and Δδ = − 0.002 ppm for its *N*-acetyl signal (see Table 4), are in complete accordance with those described to be specific for the attachment of Gal in (β1→4) linkage to GlcNAc, completing an *N*-acetyllactosamine unit, in *N*-glycosidic glycoprotein carbohydrate chains [14, 34, 35].

Table 3. *^1H chemical shifts of structural-reporter-group protons of constituent monosaccharides for the bronchial oligosaccharide-alditols possessing the Gal(β1→3)GalNAc-ol core unit*

A superscript at the name of a sugar residue indicates to which position of the adjacent monosaccharide it is glycosidically linked; for example, Gal^3 means Gal(β1→3) linked (in this case, to GalNAc-ol). Chemical shifts are relative to internal DSS (using internal acetone at 2.225 ppm) in ^2H_2O at 300 K, acquired at 500 MHz. For the complete structures of the compounds, see Table 5. In the tableheading, the structures are represented by shorthand symbolic notation (cf. [14]); ◇ = GalNAc-ol; ■— = Gal; ●— = GlcNAc and □— = Fuc

Residue	Reporter group	Chemical shift in compound						
		4b	9	5a	7	$6a_2$	$4a_2$	$3a_1$
		ppm						
GalNAc-ol	H-2	4.395	4.399	4.395	4.394	4.393	4.400	4.397
	H-3	4.065	4.091	4.061	4.060	4.060	4.051	4.050
	H-4	3.507	3.522	3.468	3.465	3.454	3.497	3.495
	H-5	4.196	4.163	4.281	4.282	4.270	4.185	4.183
	H-6	3.69	3.68	3.931	3.931	3.927	3.7	3.7
	NAc	2.050	2.046	2.066	2.067	2.067	2.047	2.046
Gal^3	H-1	4.478	4.584	4.468	4.465	4.463	4.464	4.464
	H-4	3.901	3.926	3.901	3.900	3.899	4.126	4.127
$GlcNAc^6$	H-1	–	–	4.538	4.560	4.560	–	–
	H-6	–	–	3.932	3.998	4.010	–	–
	NAc	–	–	2.066	2.064	2.056	–	–
Gal^4	H-1	–	–	–	4.470	4.448	4.481	4.461
	H-4	–	–	–	3.925	3.927	3.927	3.900
$GlcNAc^3$	H-1	–	–	–	–	–	4.688	4.692
	H-6	–	–	–	–	–	3.954	3.966
	NAc	–	–	–	–	–	2.042	2.032
Fuc^2	H-1	–	5.256	–	–	–	–	–
	H-5	–	4.279	–	–	–	–	–
	CH_3	–	1.243	–	–	–	–	–
Fuc^3	H-1	–	–	–	–	5.109	–	5.140
	H-5	–	–	–	–	4.83[a]	–	4.83[a]
	CH_3	–	–	–	–	1.174	–	1.177

[a] Signal is hidden under the relatively broad HO^2H line.

Moreover, a downfield shift is observed for H-6($J_{5,6}$ = 2.3 Hz; $J_{6,6'}$ = − 12.0 Hz) of GlcNAc ($\Delta\delta$ = 0.071 ppm). In conclusion, the structure of 3b is as follows: Gal(β1→4)GlcNAc-(β1→3)GalNAc-ol.

The resonance position of GalNAc-ol H-2 (δ = 4.289 ppm) in the spectrum of fraction 5b defines this oligosaccharide-alditol to be of the GlcNAc(β1→3)GalNAc-ol core type. The chemical shifts of H-1 (δ = 4.461 ppm) and H-4 (δ = 3.919 ppm) of Gal in fraction 5b differ slightly from those of Gal in fraction 3b (see Table 4). In conjunction with the $J_{1,2}$ value (7.7 Hz), these values point to the β-glycosidic linkage of the terminal Gal residue (compare also with fraction 4b, Table 3). The predominant difference between the spectra of fractions 3b and 5b is the appearance of the GlcNAc H-3 signal in the latter, at δ = 3.911 ppm, outside the bulk resonance. The (1→3)-type of linkage between Gal and GlcNAc (see Table 2) is apparently expressed in the relatively large downfield shift ($\Delta\delta \approx$ 0.33 ppm) of GlcNAc H-3, making this proton a structural-reporter group. In addition to the H-3 signal of GlcNAc (characterized by the coupling constants $J_{2,3}$ = 10.4 Hz and $J_{3,4}$ = 8.5 Hz), other GlcNAc reporter-group signals are significantly shifted due to the attachment of Gal in (β1→3) linkage, namely H-1 ($\Delta\delta \approx$ 0.05 ppm; $J_{1,2}$ = 8.4 Hz) and the *N*-acetyl protons ($\Delta\delta \approx$ − 0.01 ppm) (see Table 4). The structure of compound 5b is established to be: Gal(β1→3)GlcNAc(β1→3)GalNAc-ol.

In the spectrum of fraction 6b (see Fig. 4), the GalNAc-ol H-2 resonance is found at δ = 4.267 ppm. Its typical doublet-of-triplets shape cannot readily be recognized, because the signal partly coincides with that of H-5 of Fuc, present as the additional constituent monosaccharide in comparison to fractions 3b and 5b (see Table 1). Nevertheless, the chemical shift makes clear that GalNAc-ol is substituted by GlcNAc in (β1→3) linkage. The H-3 signal of the core GlcNAc residue ($J_{2,3}$ = 10.4 Hz; $J_{3,4}$ = 8.5 Hz) is found at δ = 4.017 ppm. Such a position for H-3, which is even more downfield than in the case of fraction 5b, is indicative of β-substitution of GlcNAc at C-3. The former part of this conclusion is supported by the parameters of the Gal structural-reporter groups, because its H-1 doublet (at δ = 4.580 ppm) shows $J_{1,2}$ = 8.4 Hz. The Fuc residue is present in (α1→2) linkage to Gal. This can be inferred from the spectral features of the Fuc structural-reporter groups: δ H-1 = 5.210 ppm; $J_{1,2}$ = 4.15 Hz; δ H-5 = 4.270 ppm; and δ CH_3 = 1.231 ppm (compare with compound 9, Table 3). Based on these data, the structure of compound 6b is as follows: Fuc(α1→2)Gal(β1→3)GlcNAc(β1→3)GalNAc-ol.

Table 4. *^1H chemical shifts of structural-reporter-group protons of constituent monosaccharides for the bronchial oligosaccharide-alditols possessing the GlcNAc(β1→3)GalNAc-ol core unit*

A superscript at the name of a sugar residue indicates to which position of the adjacent monosaccharide it is glycosidically linked; for example, GlcNAc^3 means GlcNAc(β1→3) linked (in this case, to GalNAc-ol). Gal^{4,6} means Gal(β1→4) linked to GlcNAc which itself is (β1→6) linked to GalNAc-ol; Gal^{x,3} means Gal(β1→x) linked to GlcNAc which itself is (β1→3) linked to GalNAc-ol, wherein $x = 3$ or 4 (see heading). Chemical shifts are relative to internal DSS (using internal acetone at 2.225 ppm) in ^2H_2O at 300 K, acquired at 500 MHz. For the complete structures of the compounds, see Table 5. In the tableheading, the structures are represented by short-hand symbolic notation (cf. [14]); ◇ = GalNAc-ol; ■— = Gal; ●— = GlcNAc and □— = Fuc. The type of linkage in ■——● is (1→4) unless otherwise indicated

Residue	Reporter group	Chemical shift in compound						
		2	3b	5b	6b	4a$_1$	6a$_1$	3a$_2$
		ppm						
GalNAc-ol	H-2	4.287	4.290	4.289	4.267	4.282	4.283	4.280
	H-3	3.996	4.002	4.012	3.982	3.986	3.991	3.988
	H-4	3.546	3.552	3.560	3.587	3.515	3.521	3.5–3.6[a]
	H-5	4.141	4.143	4.138	4.115	4.239	4.240	4.230
	H-6	3.65	3.6–3.7	3.6–3.7	3.6–3.7	3.906	3.91	3.9[a]
	NAc	2.037	2.038	2.034	2.036	2.045	2.045	2.046
GlcNAc^3	H-1	4.604	4.631	4.654	4.656	4.599	4.624	4.597
	H-3	3.584	3.6–3.7	3.911	4.017	3.576	3.6–3.7	3.576
	H-6	3.950	4.021	3.954	3.952	3.951	4.021	3.966
	NAc	2.085	2.083	2.073	2.113	2.081	2.079	2.081
Gal^{x,3}	H-1	–	4.455	4.461	4.580	–	4.456	–
	H-4	–	3.926	3.919	3.891	–	3.927	–
GlcNAc^6	H-1	–	–	–	–	4.564	4.565	4.570
	H-6	–	–	–	–	3.998	3.998	4.0[a]
	NAc	–	–	–	–	2.061	2.062	2.052
Gal^{4,6}	H-1	–	–	–	–	4.473	4.474	4.453
	H-4	–	–	–	–	3.927	3.927	3.921
Fuc^2	H-1	–	–	–	5.210	–	–	–
	H-5	–	–	–	4.270	–	–	–
	CH_3	–	–	–	1.231	–	–	–
Fuc^3	H-1	–	–	–	–	–	–	5.118
	H-5	–	–	–	–	–	–	4.83[b]
	CH_3	–	–	–	–	–	–	1.177

[a] Values could not be determined more accurately, due to the low amount of this compound present in the mixture together with 3a$_1$ (see Fig. 7).
[b] Signal is hidden under the relatively broad HO^{2}H line.

The shift effects of Fuc attachment upon the Gal reporter groups (e.g. $\Delta\delta$ for H-1 ≈ 0.12 ppm, in comparison to fraction 5b) corroborate the (α1→2)-type of linkage. The relatively large shift increment for GlcNAc H-3, concomitant with that for its *N*-acetyl methyl protons, in the step from compound 5b to 6b, is in line with the effects described for elongation of blood-group-active glycosphingolipid oligosaccharide chains by Fuc in (α1→2) linkage to Gal; the latter were studied by ^{1}H-NMR at 360 MHz in dimethylsulfoxide. In fact, these effects were found to be typical of the Fuc-(α1→2)Gal(β1→3)GlcNAc(β1→ ·) structural element [36]. Finally, it should be mentioned that the chemical shift of Fuc H-1 and, to a less extent, that of H-5 and CH_3, for fraction 6b significantly differ from those for fraction 9 (compare Tables 3 and 4). The chemical shifts of the Fuc structural-reporter groups appear to be sensitive to subtle changes of the micro-environment of the residue in the chain, like the presence of GlcNAc in compound 6b instead of GalNAc-ol in compound 9.

The spectral data for compound 5a (see Table 3) reveal that the branched trisaccharide contains Gal as the substituent at C-3 of GalNAc-ol (for GalNAc-ol H-2, δ = 4.395 ppm; compare with 4b and 9). Therefore, GlcNAc must be present at C-6 of GalNAc-ol (Table 2). Both substituents are attached to GalNAc-ol in β-linkage (for Gal, $J_{1,2}$ = 7.85 Hz; for GlcNAc, $J_{1,2}$ = 8.45 Hz). The structure of compound 5a appears to be: Gal(β1→3)[GlcNAc(β1→6)]GalNAc-ol.

As can be deduced from the step from compound 4b to 5a (see Table 3), extension of GalNAc-ol at C-6 by GlcNAc is expressed in a shift increment for H-5 of GalNAc-ol ($\Delta\delta$ = 0.085 ppm). Moreover, the GalNAc-ol H-6 signal ($J_{5,6}$ = 6.0 Hz; $J_{6,6'}$ = − 10.5 Hz) emerges out of the bulk of skeleton proton resonances, towards δ = 3.931 ppm (cf. [29]). The decrease in absolute value of the geminal coupling constant $J_{6,6'}$ in the step from compound 4b (− 11.5 Hz) to 5a proves independently the linkage from GlcNAc to GalNAc-ol to be (1→6) (cf. [14,28,33]). The identity of the GalNAc-ol H-6 signal has been confirmed by selective irradiation of the GalNAc-ol H-5 signal. As usual for a substituent at C-6 of GalNAc-ol [28,29], GlcNAc does not influence the chemical shift of H-2 of GalNAc-ol. The *N*-acetyl signals of GalNAc-ol and GlcNAc in the spectrum of com-

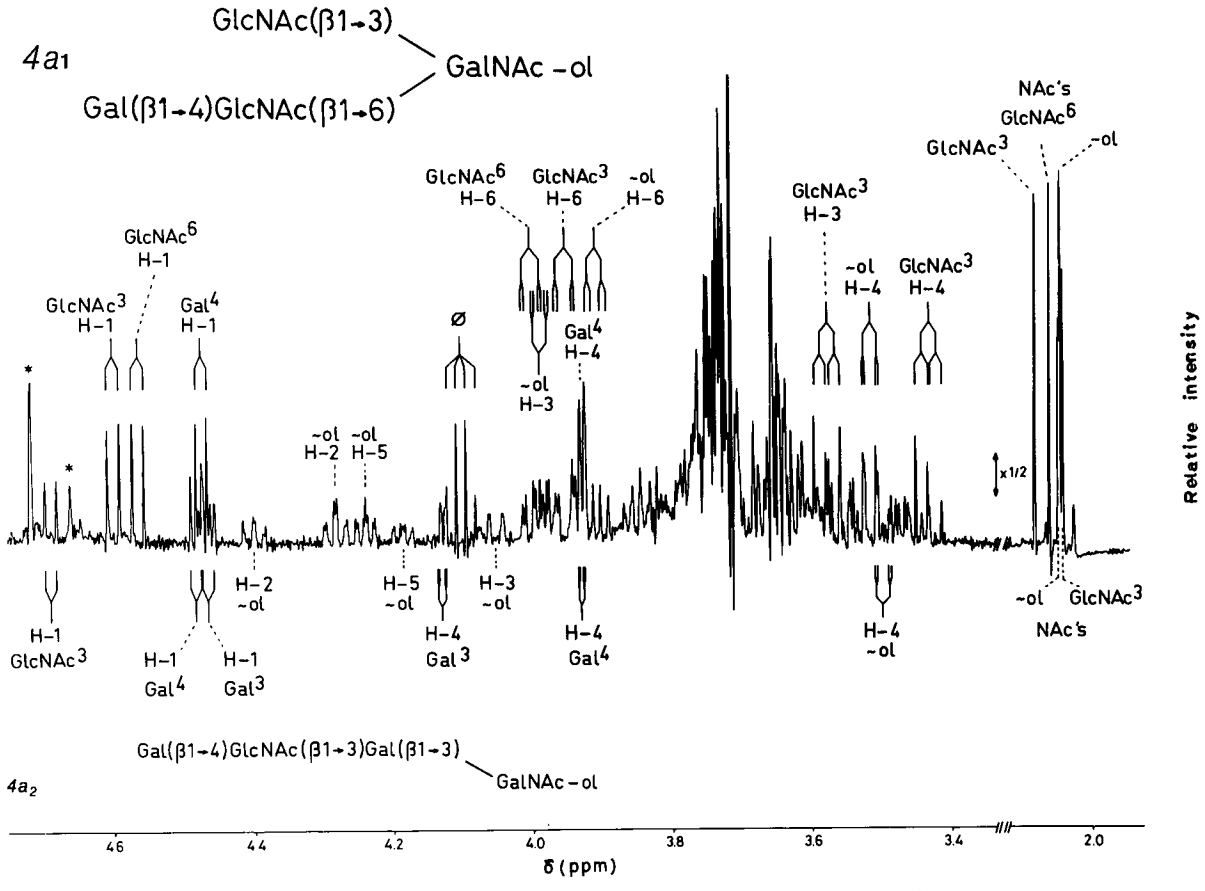


Fig. 5. *Resolution-enhanced 500-MHz ^1H-NMR spectrum of mixture 4a of bronchial oligosaccharide-alditols, in ^2H_2O at 300 K.* Signals belonging to the main component $4a_1$ of the mixture are assigned on top of the spectrum, signals from the minor component $4a_2$ at the bottom. The quartet at $\delta = 4.106$ ppm, marked by ϕ, stems from a frequently occurring, unknown non-carbohydrate, non-protein contaminant (the quartet is coupled with a doublet of thrice its intensity at $\delta \approx 1.32$ ppm). The relative-intensity scale of the *N*-acetyl proton region of the spectrum differs from that of the other part, as indicated. (The asterisk indicates a spinning side-band of the HO^2H line)

pound 5a coincide at $\delta = 2.066$ ppm. Apparently, the GalNAc-ol signal has undergone a downfield shift ($\Delta\delta = 0.016$ ppm) due to introduction of GlcNAc as compared to compound 4b.

From the positions of the GalNAc-ol structural-reporter-group signals in the spectrum of fraction 7 (see Fig. 3), in particular from δ H-2, δ H-5 and δ H-6 (see Table 3) and $J_{6,6'} = -10.5$ Hz, it can be concluded that in this tetrasaccharide-alditol (Table 1) GalNAc-ol is substituted in the same way as in compound 5a. In consequence, Gal is present at C-3 and GlcNAc at C-6 of GalNAc-ol, both in β-linkage. The second Gal residue (δ H-1 $\approx$ 4.47 ppm; $J_{1,2} = 7.8$ Hz) is present in ($\beta1\rightarrow4$) linkage to GlcNAc (δ H-1 = 4.560 ppm; $J_{1,2} = 7.8$ Hz). This can be inferred from the shift effects shown by H-1 ($\Delta\delta = 0.022$ ppm) and by the *N*-acetyl methyl protons ($\Delta\delta = -0.002$ ppm) of GlcNAc, as compared to fraction 5a (cf. the step from compound 2 to 3b, Table 4). Also a downfield shift is observed for H-6 of GlcNAc. Therefore, the structure of compound 7 has been established as Gal($\beta1\rightarrow3$)[Gal($\beta1\rightarrow4$)GlcNAc($\beta1\rightarrow6$)]GalNAc-ol. The assignment of the H-1 doublets at $\delta = 4.465$ and 4.470 ppm to Gal^3 and Gal^4, respectively, is based upon comparison with extensions of compound 7 (*vide infra*, compound $6a_2$).

Evaluating the contribution of 500-MHz ^{1}H-NMR spectroscopy in the structure elucidation of the eight oligosaccharides described above, it can be mentioned that for the disaccharides (2, 4b) and linear trisaccharides (3b, 5b, 9), the sequences proposed on the basis of periodate oxidation or methylation analysis could be confirmed, and completed with the configurations of the glycosidic linkages. For the branched trisaccharide 5a and for the tetrasaccharides 6b and 7, ^{1}H-NMR spectroscopy made it possible to discriminate between the alternative structures which could be drawn up on the basis of methylation analysis. Moreover, the sequences could be supplemented by the anomeric configurations of the glycosidic linkages.

Fractions Consisting of a Mixture of Two Identifiable Oligosaccharide-Alditols

The 500-MHz ^{1}H-NMR spectra of the remaining three bronchial-mucin oligosaccharide-alditol fractions, i.e. 4a, 6a, and 3a, are shown in Fig. 5, 6, and 7, respectively. The common feature of these spectra is the simultaneous occurrence of two GalNAc-ol H-2 signals, one at $\delta \approx 4.39$ ppm, and the other at $\delta \approx 4.29$ ppm, having unequal intensities. This reflects that each of these fractions is a mixture of, at least, two oligosaccharide-alditols, differing in the type of core substituent at C-3 of GalNAc-ol. The remaining structural-reporter-group signals in the spectra can be attributed to either the major or the minor component of the mixture, on the basis of correspondence of their intensity with that of one of the GalNAc-ol H-2 signals. According to the type of core being Gal($\beta1\rightarrow3$)GalNAc-ol or GlcNAc($\beta1\rightarrow3$)Gal-

NAc-ol, the NMR data of the two components of each of the three mixtures are included in Table 3 or 4, respectively. The major and minor components of a mixture are denoted by adding indices 1 and 2, respectively, to the number of the subfraction concerned.

In the spectrum of fraction 4a (Fig. 5), the GalNAc-ol H-2 signals occur at $\delta = 4.400$ and 4.282 ppm, with intensities in the ratio of 1:2. A similar ratio is observed for the intensities of other structural-reporter-group signals. To the collection of higher-intensity signals, corresponding to fraction $4a_1$, belong *inter alia* two anomeric doublets around $\delta \approx 4.6$ ppm ($J_{1,2} > 8$ Hz), and another anomeric doublet at $\delta \approx 4.47$ ppm ($7 < J_{1,2} < 8$ Hz). These stem from two β-linked GlcNAc residues and a β-linked Gal residue, respectively, suggesting that compound $4a_1$ is a tetrasaccharide-alditol consisting of GalNAc-ol, GlcNAc, and Gal in the ratio 1:2:1.

The resonance position of GalNAc-ol H-2 of compound $4a_1$ indicates the GlcNAc(β1→3)GalNAc-ol core type (Table 4). To define the complete substitution pattern of GalNAc-ol, the characteristically-shaped GalNAc-ol H-5 signal is considered. This resonance is found at $\delta = 4.239$ ppm. Comparison of its chemical shift value with that for GalNAc-ol H-5 in compounds 5a and 7 (see Table 3) leads to the conclusion that GalNAc-ol in compound $4a_1$ bears a β-linked GlcNAc at C-6. The H-1 doublet at $\delta = 4.599$ ppm ($J_{1,2} = 8.5$ Hz) is attributed to GlcNAc in (β1→3) linkage to GalNAc-ol. This implies that $GlcNAc^3$ is present in terminal position (compare with compound 2, Table 4). The chemical shifts of the anomeric protons of $GlcNAc^6$ ($\delta = 4.564$ ppm; $J_{1,2} = 8.2$ Hz) and of the terminal Gal residue ($\delta = 4.473$ ppm; $J_{1,2} = 7.8$ Hz) point to the occurrence of an *N*-acetyllactosamine moiety attached to C-6 of GalNAc-ol (compare with compound 7, Table 3). The structure of compound $4a_1$ appears to be: GlcNAc(β1→3)[Gal(β1→4)GlcNAc(β1→6)]-GalNAc-ol.

This conclusion is supported by the chemical shifts of various other structural-reporter groups, e.g. H-6 of GalNAc-ol, H-6 and the *N*-acetyl protons of $GlcNAc^6$, and H-4 of Gal (see Table 4). It should be noted that the chemical shift values of H-1 and H-4 of Gal^4 in compound $4a_1$ support the assignment of the Gal H-1 and H-4 signals for compound 7 (Table 3, Fig. 3). The structural alternative for compound $4a_1$, namely, the tetrasaccharide having Gal in (β1→4) linkage to $GlcNAc^3$ instead of $GlcNAc^6$, can be disproved upon comparison with compound 5a (terminal $GlcNAc^6$). Besides, the type of linkage between Gal and $GlcNAc^6$ is definitely (β1→4) and not (β1→3), since there is no GlcNAc H-3 signal observable at $3.9 < \delta < 4.1$ ppm (compare with compounds 5b and 6b, see Fig. 4).

In the anomeric region of the spectrum of fraction 4a (see Fig. 5), one GlcNAc (δ H-1 = 4.688 ppm; $J_{1,2} = 8.2$ Hz) and two Gal (δ H-1 = 4.481 and 4.464 ppm; $J_{1,2}$=8.0 and 8.1 Hz, respectively) anomeric doublets can be discerned, the intensity of which corresponds with that of the GalNAc-ol H-2 signal at $\delta = 4.400$ ppm. Therefore, compound $4a_2$ is suggested to be a tetrasaccharide-alditol, containing GalNAc-ol, GlcNAc, and Gal, in the ratio 1:1:2.

The GalNAc-ol H-5 signal of lower intensity is found at $\delta = 4.185$ ppm. This means that GalNAc-ol in compound $4a_2$ is not substituted at C-6 (compare with compound 4b). The core Gal^3 residue is substituted at C-3 in β-linkage by a GlcNAc residue. This conclusion is based on the chemical shift of the Gal^3 H-4 atom ($\delta = 4.126$ ppm; $J_{3,4} = 3.5$ Hz; $J_{4,5} \approx 1.0$ Hz), which is known to be characteristic of the →·)GlcNAc(β1→3)Gal(β1→·) sequence [14, 15, 37, 38]. The $GlcNAc^3$ residue is substituted by the second Gal present, in (β1→4) linkage (see Table 3, compare with compound 7). This affords the structure of compound $4a_2$ to be as follows: Gal(β1→4)GlcNAc(β1→3) Gal(β1→3)GalNAc-ol.

The chemical shifts of the $GlcNAc^3$ H-1 and *N*-acetyl protons and those of the Gal^4 and Gal^3 H-1 and H-4 atoms of compound $4a_2$ show close resemblance to those of the corresponding protons in (β1→3)-repeating *N*-acetyllactosamine sequences which occur in *N*-glycosidic carbohydrate chains [14]. The assignments of the Gal H-1 signals for compound $4a_2$ (see Table 3 and Fig. 5) are primarily based on this resemblance; they will be supported by the shift effects upon extending compound $4a_2$ to $3a_1$ (*vide infra*).

In fraction 4a, a contaminant oligosaccharide is present in minute proportion (less than 10% of the mixture), as can be deduced form the additional Gal H-1 signal at $\delta = 4.469$ ppm ($J_{1,2} = 8.0$ Hz). The remainder of its signals is so low that the structure of this third component of 4a could not be identified.

From the spectrum of fraction 6a (see Fig. 6) it is readily discernible that the major component of this mixture, $6a_1$, is of the GlcNAc(β1→3)GalNAc-ol core type (for GalNAc-ol, δ H-2 = 4.283 ppm). The minor constituent, $6a_2$, accounts for the GalNAc-ol H-2 signal at $\delta = 4.393$ ppm. Based on the relative intensities of the GalNAc-ol H-2 signals, compounds $6a_1$ and $6a_2$ occur in the ratio 3:1, which is also reflected in other structural-reporter-group signals. The occurrence of two β-GlcNAc H-1 signals, at $\delta = 4.624$ ppm ($J_{1,2} = 8.1$ Hz) and at $\delta = 4.565$ ppm ($J_{1,2} = 8.2$ Hz), and of two Gal H-1 doublets, at $\delta = 4.474$ ppm ($J_{1,2} = 7.8$ Hz) and $\delta = 4.456$ ppm ($J_{1,2} = 7.8$ Hz), with intensities equal to that of the high-intensity GalNAc-ol H-2 signal, suggests that component $6a_1$ is a pentasaccharide-alditol that contains GalNAc-ol, GlcNAc, and Gal in the ratio 1:2:2.

The GalNAc-ol H-5 signal of compound $6a_1$ is found at $\delta = 4.240$ ppm. As pointed out for compound $4a_1$, this value indicates that GalNAc-ol in compound $6a_1$ bears a GlcNAc residue at C-6, in β-linkage. Further comparison of the spectral data of compound $6a_1$ with those of compound $4a_1$ (Table 4) and 7 (Table 3) reveals that the aforementioned signals at $\delta = 4.565$ and 4.474 ppm can be ascribed to the anomeric protons of the Gal(β1→4)GlcNAc moiety, (β1→6)-linked to GalNAc-ol. This conclusion is substantiated by the similarity of the chemical shifts of other structural-reporter groups of this branch with those of compounds $4a_1$ and 7. In consequence, the remaining anomeric doublets mentioned above belong to GlcNAc and Gal in the (1→3)-linked branch. As compared to compound $4a_1$, H-1 of $GlcNAc^3$ has undergone a downfield shift ($\Delta\delta = 0.025$ ppm), while its *N*-acetyl signal is shifted upfield ($\Delta\delta = -0.002$ ppm). These values are typical of substitution of GlcNAc by a β-linked Gal at C-4. The chemical shifts of the structural-reporter groups of this *N*-acetyllactosamine unit, (β1→3)-linked to GalNAc-ol in compound $6a_1$, are in agreement with those of fraction 3b (see Table 4). These data establish the structure of compound $6a_1$ to be: Gal(β1→4)GlcNAc(β1→3)[Gal(β1→4)-GlcNAc(β1→6)]GalNAc-ol.

The minor component of this mixture, $6a_2$, contains, in addition to the core residues Gal and GalNAc-ol, another Gal residue, because of the presence of two low-intensity anomeric doublets in the spectrum (Fig. 6) at $\delta \approx 4.5$ ppm, having $7 < J_{1,2} < 8$ Hz. Furthermore, a GlcNAc residue forms part of component $6a_2$ (δ H-1 = 4.560 ppm; $J_{1,2} = 8.3$ Hz),

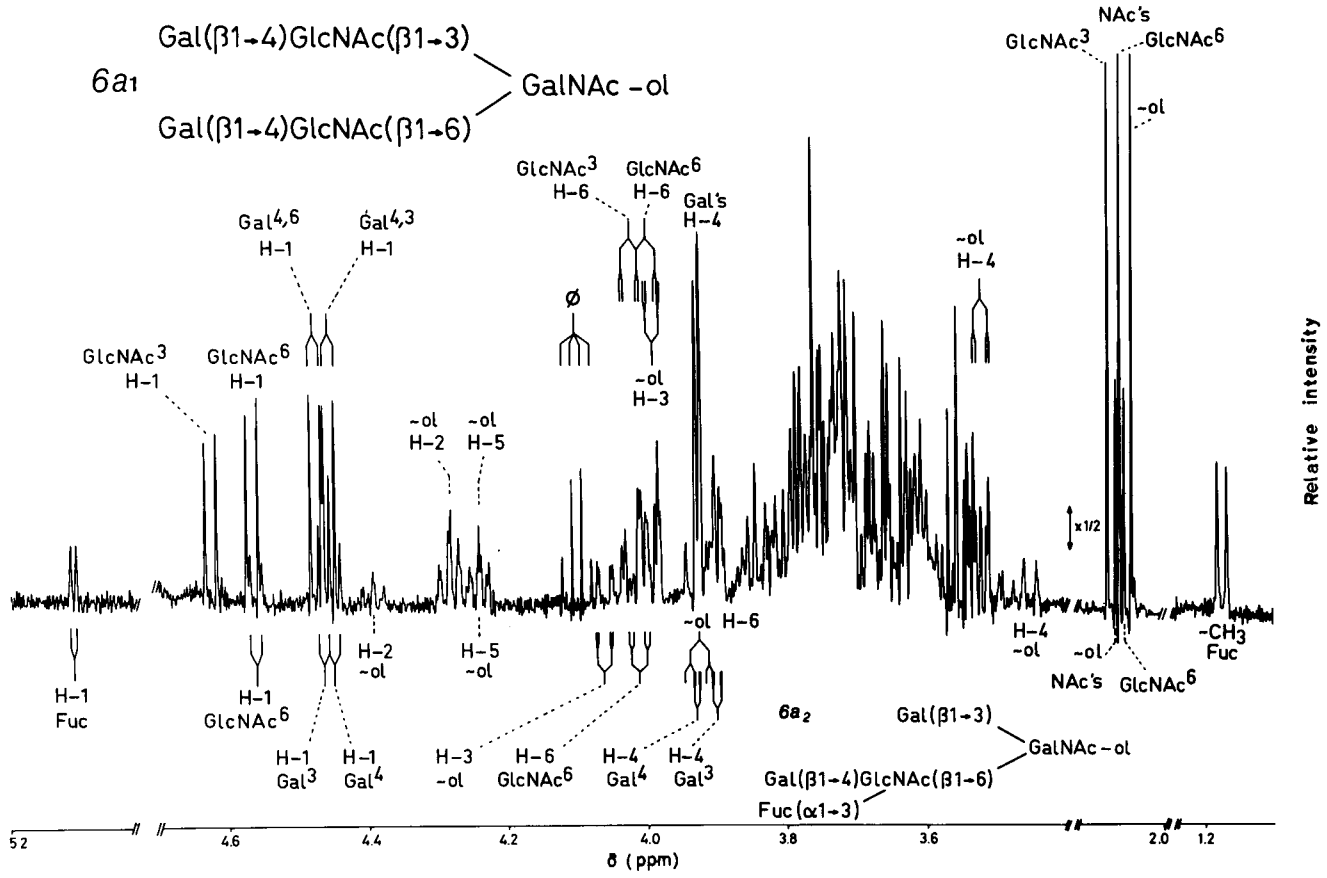


Fig. 6. *Resolution-enhanced 500-MHz ^{1}H-NMR spectrum of mixture 6a of bronchial oligosaccharide-alditols, in ^2H_2O at 300 K.* Signals belonging to the main component $6a_1$ of the mixture are assigned on top of the spectrum, signals from the minor component $6a_2$ at the bottom. The quartet at $\delta = 4.106$ ppm, marked by ϕ, stems from a frequently occurring, unknown non-carbohydrate, non-protein contaminant (the quartet is coupled with a doublet of thrice its intensity at $\delta \approx 1.32$ ppm). The relative-intensity scale of the *N*-acetyl proton region of the spectrum differs from that of the other parts, as indicated

and, in addition, a Fuc residue is present (δ H-1 = 5.109 ppm; $J_{1,2} = 3.9$ Hz; δ CH_3 = 1.174 ppm).

The chemical shift of GalNAc-ol H-5 ($\delta = 4.270$ ppm) is indicative of the C-6 substitution of GalNAc-ol in the pentasaccharide $6a_2$ by GlcNAc in β-linkage (compare with compounds 5a and 7, Table 3).

The core Gal residue of compound $6a_2$ is present in terminal position (see Table 2). The H-1 signal at $\delta = 4.463$ ppm ($J_{1,2} = 7.5$ Hz) and the H-4 signal at $\delta = 3.899$ ppm ($J_{3,4} = 3.5$ Hz; $J_{4,5} < 1.0$ Hz) are attributed to this Gal^3 residue (compare compounds 5a and 7, see Table 3). The chemical shift values of the Fuc H-1 and CH_3 protons point unambiguously to ($\alpha 1 \rightarrow 3$) linkage of this residue to a β-linked GlcNAc which simultaneously bears a β-linked Gal at C-4 [14, 16, 17, 39]. Thus, compound $6a_2$ is an extension of 7, possessing Fuc in ($\alpha 1 \rightarrow 3$)-linkage to $GlcNAc^6$: Gal($\beta 1 \rightarrow 3$)-{Gal($\beta 1 \rightarrow 4$)[Fuc($\alpha 1 \rightarrow 3$)]GlcNAc($\beta 1 \rightarrow 6$)}GalNAc-ol.

This conclusion is supported by the shift effects brought about by introduction of Fuc. The *N*-acetyl signal of $GlcNAc^6$ shows a highly characteristic upfield shift by -0.008 ppm, as compared to compound 7 (Table 3). For H-1 of $Gal^{4,6}$ an upfield shift by -0.022 ppm is observed. The chemical shift of H-1 of $GlcNAc^6$ is not affected, but the signal is considerably broadened. All these spectral features are in accord with those described for the Gal($\beta 1 \rightarrow 4$)[Fuc($\alpha 1 \rightarrow 3$)]GlcNAc($\beta 1 \rightarrow \cdot$) structural element [14–17, 39–41].

The shift alteration for H-1 of $Gal^{4,6}$ in the step from compound 7 to $6a_2$, matching exactly that described earlier [14, 17] for the introduction of this type of Fuc, provides additional proof for the assignment of the Gal H-1 resonances for compound 7 (*vide supra*). Concerning the *N*-acetyl signals, the singlet at $\delta = 2.067$ ppm in the spectrum of compound 7 is unaffected by the attachment of Fuc. Therefore, it is ascribed to GalNAc-ol for compound 7, as well as for compound $6a_2$.

The 500-MHz ^{1}H-NMR spectrum of fraction 3a is presented in Fig. 7. Two GalNAc-ol H-2 signals occur, at $\delta = 4.397$ ppm and at $\delta = 4.280$ ppm, with a relative-intensity ratio of 9:1. The main component of this mixture, $3a_1$, contains Gal in ($\beta 1 \rightarrow 3$) linkage to GalNAc-ol; the minor constituent, $3a_2$, is of the GlcNAc($\beta 1 \rightarrow 3$)GalNAc-ol core type.

To the series of high-intensity reporter-group signals belong two Gal H-1 doublets ($4.4 < \delta < 4.5$ ppm; $7 < J_{1,2} < 8$ Hz), a GlcNAc H-1 doublet (at $\delta \approx 4.7$ ppm; $J_{1,2} > 8$ Hz), and a set of Fuc reporter groups (δ H-1 = 5.140 ppm; $J_{1,2} = 3.9$ Hz; δ CH_3 = 1.177 ppm). From these data it is concluded that component $3a_1$ is a pentasaccharide-alditol containing GalNAc-ol, Gal, GlcNAc, and Fuc in the ratio 1:2:1:1.

Comparison of the data of compound $3a_1$ with those of $4a_2$ (Table 3, Fig. 5) reveals that the former is an extension of

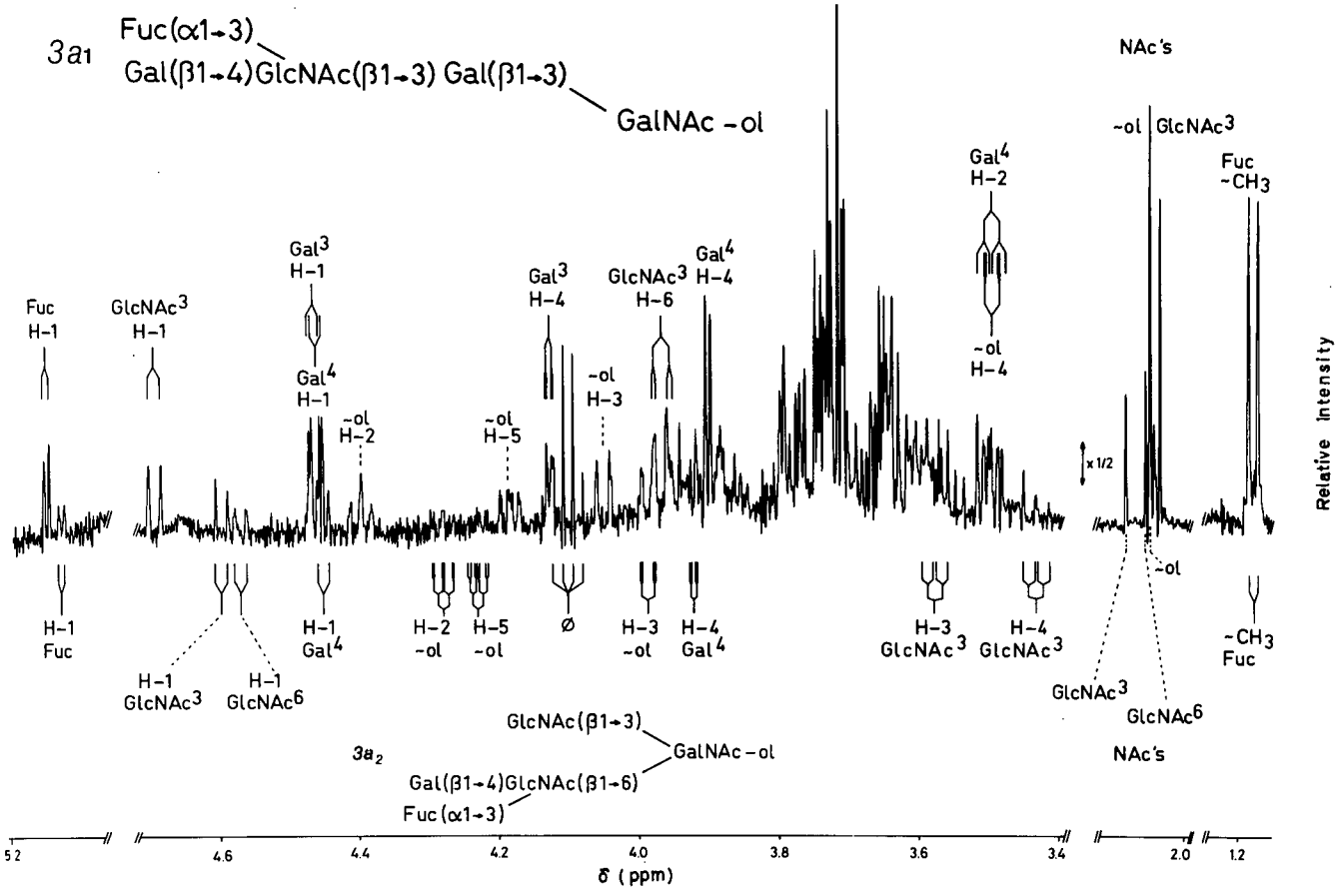


Fig. 7. *Resolution-enhanced 500-MHz ^1H-NMR spectrum of mixture 3a of bronchial oligosaccharide-alditols, in ^2H_2O at 300 K.* Signals belonging to the main component $3a_1$ of the mixture are assigned on top of the spectrum, signals from the minor component $3a_2$ at the bottom. The quartet at $\delta = 4.106$ ppm, marked by ϕ, stems from a frequently occurring, unknown non-carbohydrate, non-protein contaminant (the quartet is coupled with a doublet of thrice its intensity at $\delta \approx 1.32$ ppm). The relative-intensity scale of the *N*-acetyl proton region of the spectrum differs from that of the other parts, as indicated

the latter having an additional Fuc residue in (α1→3) linkage to GlcNAc^3: Gal(β1→4)[Fuc(α1→3)]GlcNAc-(β1→3)Gal(β1→3)GalNAc-ol.

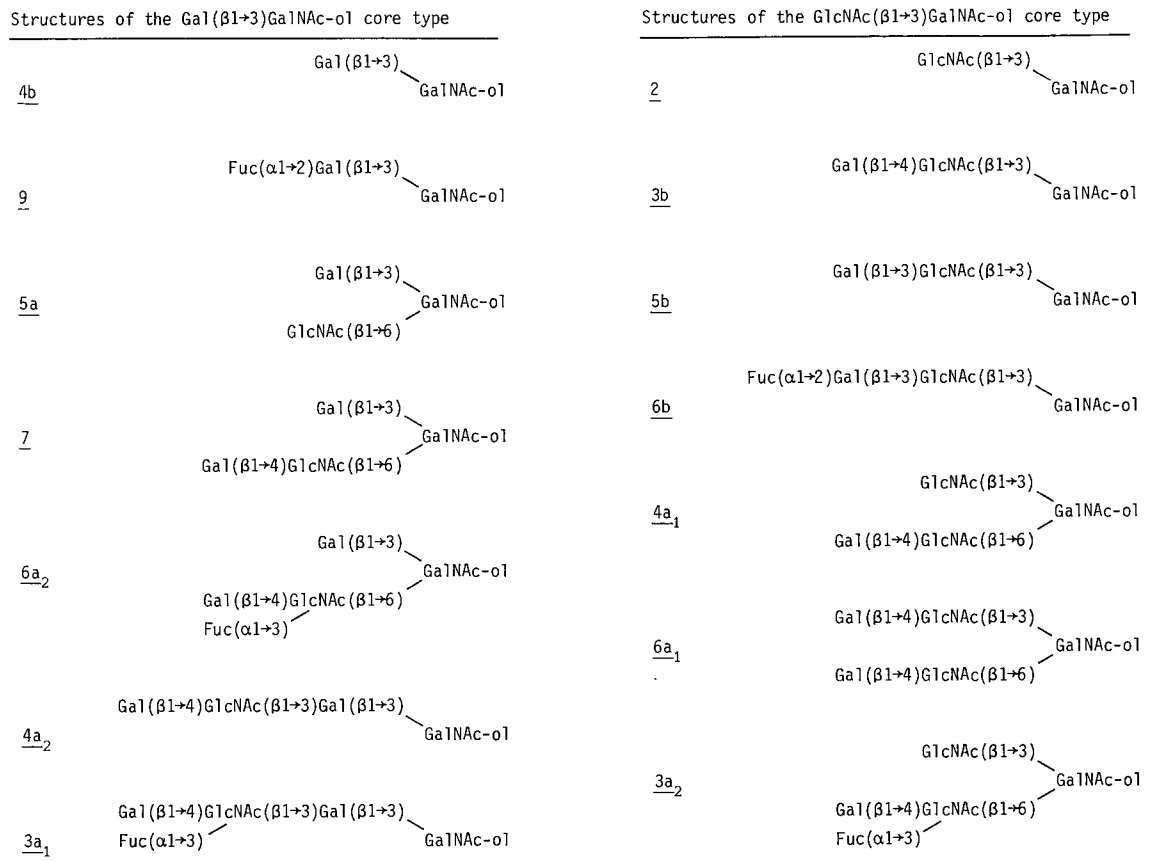
The monosubstitution of GalNAc-ol, at C-3, is evident from the GalNAc-ol reporter-group chemical shifts (i.a. δ H-5 = 4.183 ppm). The substitution of Gal^3 at C-3 is obvious from the chemical shift of its H-4 (δ = 4.127 ppm). In comparison to compound $4a_2$, the *N*-acetyl signal of GlcNAc^3 has undergone an upfield shift ($\Delta\delta$ = − 0.010 ppm), being highly characteristic of the attachment of Fuc in (α1→3) linkage. The chemical shift of GlcNAc^3 H-1 is hardly affected but the signal is relatively broad-lined. The H-1 signal of Gal^4 is shifted by −0.020 ppm. As described above for compound $6a_2$, these effects are indicative of substitution of an *N*-acetyllactosamine unit at C-3 of the GlcNAc residue by Fuc in α-linkage. The chemical shift of the Fuc CH_3 protons (δ = 1.177 ppm) is in line with this conclusion. However, the resonance position of Fuc H-1 (δ = 5.140 ppm) has not been observed before for this type of Fuc residue. This demonstrates that the chemical shifts of Fuc structural-reporter groups, especially that of H-1, are very sensitive to the micro-environment of the structural element containing Fuc.

The minor component, $3a_2$, contains two β-GlcNAc residues (δ H-1 ≈ 4.6 ppm; $J_{1,2} > 8$ Hz), a β-Gal residue (δ H-1 = 4.453 ppm; $J_{1,2}$ = 7.9 Hz) and an α-Fuc residue (δ H-1 = 5.118 ppm; $J_{1,2}$ = 4.3 Hz; δ CH_3 = 1.177 ppm). Thus, compound $3a_2$ is a pentasaccharide-alditol consisting of GalNAc-ol, Gal, GlcNAc, and Fuc in the ratio 1:1:2:1. Comparison of the NMR data of compound $3a_2$ with those of $4a_1$ (see Table 4, Fig. 5) shows that the former is an extension of the latter with Fuc attached at C-3 of GlcNAc^6: GlcNAc-(β1→3){Gal(β1→4)[Fuc(α1→3)]GlcNAc(β1→6)}GalNAc-ol.

The set of the GalNAc-ol H-2 and H-5 chemical shifts (δ = 4.280 and 4.230 ppm, respectively) points to the presence of β-linked GlcNAc residues at C-3 and C-6 of GalNAc-ol (compare with compounds $4a_1$ and $6a_1$, Table 4). The core GlcNAc^3 residue is present in terminal position (δ H-1 = 4.597 ppm; δ NAc = 2.081 ppm; δ H-4 = 3.430 ppm), as can be inferred from comparison with compounds 2 and $4a_1$ (Table 4). The NMR features of the Gal(β1→4)[Fuc-(α1→3)]GlcNAc moiety linked to C-6 of GalNAc-ol in compound $3a_2$ are similar to those for compound $6a_2$ (Table 3). In comparison to compound $4a_1$, the *N*-acetyl signal of GlcNAc^6 is shifted by −0.009 ppm; its H-1 signal is broadened, and the H-1 signal of Gal^{4,6} shows an upfield shift ($\Delta\delta$ = −0.020 ppm). In conjunction with the chemical shift values for H-1 and CH_3 protons of Fuc, the occurrence of the Gal(β1→4)[Fuc(α1→3)]GlcNAc(β1→6) structural element in compound $3a_2$ is proved.

The presence of an additional singlet in the *N*-acetyl region of the spectrum (see Fig. 7), at δ = 2.040 ppm, might indicate

Table 5. *Structures of 14 neutral oligosaccharide-alditols obtained from bronchial mucin of patients suffering from cystic fibrosis*



that the afuco analogue of the main component $3a_1$ is also present in this mixture.

By employing 500-MHz ^{1}H-NMR spectroscopic analysis for the mixtures of bronchial-mucin oligosaccharide-alditols, the structure of six constituents could be established. The resulting structures and their abundances in the various mixtures are not in contradiction with the data from sugar and methylation analyses of the fractions (Tables 1 and 2). However, it cannot be completely excluded that other oligosaccharide-alditols occur in the mixtures, in much lower concentrations than those of the identified components.

The structures established for the 14 oligosaccharide-alditols obtained from cystic fibrosis bronchial-mucus glycoproteins, are compiled in Table 5.

DISCUSSION

From bronchial-mucus glycoproteins stemming from cystic fibrosis patients with blood-group-0 activity, the carbohydrate chains were released in the form of reduced oligosaccharides by alkaline borohydride treatment. Eleven oligosaccharide-alditol fractions were separated and purified in amounts amenable for structural analysis. Eight of these fractions contained rather pure oligosaccharide-alditols, whereas the other three each consisted of two main alditol components. For structural determination, i.e. sugar analysis, periodate oxidation, methylation analysis, and ^{1}H-NMR spectroscopy, the oligosaccharide-alditols were found to be suitable starting compounds.

Sugar analysis in combination with methylation analysis permitted the determination of the sequence of two disaccharides and three linear trisaccharides. However, for branched oligosaccharides and for mixtures, these approaches did not lead to unambiguous structures, although the information obtained on the substitution pattern of the constituent monosaccharides was helpful for interpretation of information obtained with 500-MHz ^{1}H-NMR spectroscopy. The latter technique was found to be a highly effective method in the unraveling of complete primary structures of *O*-glycosidic carbohydrate chains.

The 500-MHz ^{1}H-NMR spectrum of an oligosaccharide-alditol contains a number of structural-reporter-group signals, that is, signals occurring outside of the bulk resonance of the majority of sugar skeleton protons, at clearly distinguishable positions and with highly characteristic patterns. These signals bear the essential information to assign the primary structure. For this series of compounds, the molar sugar composition could be derived. The presence of GalNAc-ol and Fuc is obvious from their typical sets of structural-reporter groups (H-2, H-3, H-4, H-5, NAc, and H-1, H-5, CH_3, respectively). The occurrence of GlcNAc (in β-linkage) is recognizable from its *N*-acetyl signal additional to that of GalNAc-ol, in conjunction with an H-1 doublet in the region $4.50 < \delta < 4.75$ ppm, having $J_{1,2} > 8$ Hz. A (β-linked) Gal residue affords an H-1 doublet at $4.4 < \delta < 4.6$ ppm and is characterized by $J_{1,2}$ between 7 and 8 Hz. The molar ratio of the constituents can be inferred from the intensities of appropriate reporter-group signals. In addition, the type of substitution of each constituent monosaccharide can be deduced.

First of all, the chemical shift for H-2 of GalNAc-ol provides conclusive evidence of the type of core unit, i.e. the nature of the substituent at C-3 of GalNAc-ol. The positions of the H-5 and H-6 resonances of GalNAc-ol report on the presence of a substituent at C-6 of GalNAc-ol. The position of the H-4 signal of Gal indicates about ($\beta 1 \rightarrow 3$) substitution of this Gal residue. The H-3 atom of GlcNAc is a reporter for ($\beta 1 \rightarrow 3$) rather than ($\beta 1 \rightarrow 4$) substitution of this residue. Moreover, the chemical shifts of anomeric protons and *N*-acetyl methyl protons are sensitive to many structural features. The set of chemical shifts for Fuc structural-reporter groups defines the type of linkage, as well as the structural environment of this residue. The anomeric configurations of all glycosidic linkages are reflected in the chemical shifts of the anomeric signals, together with the values of the coupling constants $J_{1,2}$ of the sugars involved. The 500-MHz ^{1}H-NMR spectrum, because of the high resolving power and large sensitivity of the technique, provides a good criterion of the purity of the sample. The method is capable of detecting and identifying oligosaccharide-alditols in complex mixtures of related components, even if present in nanomolar amounts. It was found to be a very powerful tool in mixture analysis of these bronchial-mucus glycoprotein oligosaccharides.

The resulting 14 structures could be grouped into two classes according to the type of core, namely, Gal or GlcNAc in ($\beta 1 \rightarrow 3$) linkage to GalNAc-ol. The GalNAc-ol H-2 signal in the ^{1}H-NMR spectrum was found to be the ideal marker for this classification. The compounds show a gradually increasing complexity going from disaccharides to branched pentasaccharides (see Table 5). These two types of core and some of their extensions described here are known to occur in mucus glycoproteins [6–12,28,29,42–45]. Novel structural elements found in this series of alditols are: the ($\alpha 1 \rightarrow 2$)-fucosylated and non-fucosylated Gal($\beta 1 \rightarrow 3$)GlcNAc moiety directly linked to GalNAc-ol, and the occurrence of the Gal-($\beta 1 \rightarrow 4$)[Fuc($\alpha 1 \rightarrow 3$)]GlcNAc($\beta 1 \rightarrow \cdot$) sequence (SSEA-1 determinant). Considering the structures as entities, those of compounds $3a_1$, $3a_2$, 5b, $6a_2$, and 6b (see Table 5) have not been described, so far.

The data presented show already that a high degree of structural heterogeneity exists in human bronchial glycoproteins. This may be due to the fact that a pool of glycoproteins from six CF patients was used, and/or that the oligosaccharide chains reflect different stages of biosynthesis or degradation. The presence of different oligosaccharide cores in bronchial-mucus glycoproteins suggests that the biosynthetic processing of the various chains may be different. Some glycosyltransferases, e.g. *N*-acetylgalactosaminide ($\beta 1 \rightarrow 3$)galactosyltransferase, *N*-acetylgalactosaminide ($\beta 1 \rightarrow 6$)*N*-acetylglucosaminyltransferase, *N*-acetylglucosaminide ($\beta 1 \rightarrow 3$)galactosyltransferase, galactoside ($\alpha 1 \rightarrow 2$)fucosyltransferase and *N*-acetylglucosaminide ($\alpha 1 \rightarrow 3$)fucosyltransferase have already been characterized in various mucous glands [46–50]. However, the transferase that is able to attach GlcNAc at C-3 of GalNAc is still unknown.

This study forms part of a larger research program aimed at the characterization of the glycoproteins of bronchial mucin in terms of structures of their carbohydrate chains. The set of data presented for fundamental structures can be used for studying structural extensions, e.g. larger neutral and acidic (sialylated and/or sulfated) oligosaccharides. It remains a challenge to find out whether the chains described here are characteristic of cystic fibrosis. This can be verified by studying bronchial mucus from normal individuals and from patients with other respiratory diseases.

This investigation was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO), by the Netherlands Foundation for Cancer Research (KWF, grant UUKC-OC 79-13), by the *Association française de Lutte contre la Mucoviscidose*, and by contract 803025 from *Institut National de la Santé et de la Recherche Médicale*.

REFERENCES

1. Havez, R., Roussel, P., Degand, P. & Biserte, G. (1967) *Clin. Chim. Acta, 17*, 281–295.
2. Roberts, G. P. (1974) *Eur. J. Biochem. 50*, 265–280.
3. Roussel, P., Lamblin, G., Degand, P., Walker-Nasir, E. & Jeanloz, R. W. (1975) *J. Biol. Chem. 250*, 2114–2122.
4. Lafitte, J. J., Lamblin, G., Lhermitte, M., Humbert, P., Degand, P. & Roussel, P. (1977) *Carbohyd. Res. 56*, 333–343.
5. Lamblin, G., Humbert, P., Degand, P. & Roussel, P. (1977) *Clin. Chim. Acta, 79*, 425–436.
6. Lamblin, G., Lhermitte, M., Boersma, A., Roussel, P. & Reinhold, V. N. (1980) *J. Biol. Chem. 255*, 4595–4598.
7. Carlson, D. M. (1968) *J. Biol. Chem. 243*, 616–626.
8. Van Halbeek, H., Dorland, L., Haverkamp, J., Veldink, G. A., Vliegenthart, J. F. G., Fournet, B., Ricart, G., Montreuil, J., Gathmann, W. D. & Aminoff, D. (1981) *Eur. J. Biochem. 118*, 487–495.
9. Slomiany, A. & Slomiany, B. L. (1978) *J. Biol. Chem. 253*, 7301–7306.
10. Derevitskaya, V. A., Arbatsky, N. P. & Kochetkov, N. K. (1978) *Eur. J. Biochem. 86*, 423–437.
11. Slomiany, B. L., Murty, V. L. N. & Slomiany, A. (1980) *J. Biol. Chem. 255*, 9719–9723.
12. Hounsell, E. F., Wood, E., Feizi, T., Fukuda, M., Powell, M. E. & Hakomori, S. I. (1981) *Carbohyd. Res. 90*, 283–307.
13. Yurewicz, E. C. & Moghissi, K. S. (1981) *J. Biol. Chem. 256*, 11895–11904.
14. Vliegenthart, J. F. G., Dorland, L. & Van Halbeek, H. (1982) *Adv. Carb. Chem. Biochem. 41*, in the press.
15. Vliegenthart, J. F. G., Van Halbeek, H. & Dorland, L. (1980) in IUPAC *27th Int. Congr. Pure Appl. Chem.* (Varmavuori, A., ed.) pp. 253–262, Pergamon, Oxford.
16. Vliegenthart, J. F. G., Van Halbeek, H. & Dorland, L. (1981) *Pure Appl. Chem. 53*, 45–77.
17. Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Montreuil, J., Fournet, B. & Schmid, K. (1981) *J. Biol. Chem. 256*, 5588–5590.
18. Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Jouanneau, J. & Bourrillon, R. (1981) *Biochem. Biophys. Res. Commun. 99*, 886–892.
19. Lamblin, G., Lhermitte, M., Boersma, A., Roussel, P., Van Halbeek, H., Dorland, L. & Vliegenthart, J. F. G. (1981) *Monogr. Paediatr. 14*, 46–49.
20. Lamblin, G., Lhermitte, M., Degand, P., Roussel, P. & Slayter, H. S. (1979) *Biochimie (Paris) 61*, 23–43.
21. Boersma, A., Lamblin, G., Degand, P. & Roussel, P. (1981) *Carbohyd. Res. 94*, C7–C9.
22. Reinhold, V. N. (1972) *Methods Enzymol. 25*, 244–249.
23. Lee, Y. C. & Socca, J. R. (1972) *J. Biol. Chem. 247*, 5753–5758.
24. Hakomori, S. I. (1964) *J. Biochem. (Tokyo) 55*, 205–207.
25. Björndal, H., Lindberg, B. & Svensson, S. (1967) *Carbohyd. Res. 5*, 433–440.
26. Boersma, A., Lhermitte, M., Lamblin, G. & Degand, P. (1981) *Carbohyd. Res. 95*, 227–236.
27. Ernst, R. R. (1966) *Adv. Magn. Resonance, 2*, 1–135.
28. Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Fiat, A.-M. & Jollès, P. (1980) *Biochim. Biophys. Acta, 623*, 295–300.
29. Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Fiat, A.-M. & Jollès, P. (1981) *FEBS Lett. 133*, 45–50.
30. Iwasaki, M. & Inoue, S. (1981) *J. Biochem. (Tokyo) 89*, 1067–1074.
31. Inoue, S., Iwasaki, M. & Matsumura, G. (1981) *Biochem. Biophys. Res. Commun. 102*, 1295–1301.
32. Carlsson, H. E., Sundblad, G., Hammarström, S. & Lönngren, J. (1978) *Carbohyd. Res. 64*, 181–188.

33. Dorland, L., Schut, B. L., Vliegenthart, J. F. G., Strecker, G., Fournet, B., Spik, G. & Montreuil, J. (1977) *Eur. J. Biochem. 73*, 93–97.
34. Carver, J. P. & Grey, A. A. (1981) *Biochemistry, 20*, 6607–6616.
35. Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., Spik, G., Fournet, B. & Montreuil, J. (1979) *Eur. J. Biochem. 100*, 569–574.
36. Dabrowski, J., Hanfland, P., Egge, H. & Dabrowski, U. (1981) *Arch. Biochem. Biophys. 210*, 405–411.
37. 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.
38. Dabrowski, J., Egge, H. & Hanfland, P. (1980) *Chem. Phys. Lipids, 26*, 187–196.
39. Vliegenthart, J. F. G. (1980) *Adv. Exp. Med. Biol. 125*, 77–91.
40. Spik, G., Strecker, G., Fournet, B., Bouquelet, S., Montreuil, J., Dorland, L., Van Halbeek, H. & Vliegenthart, J. F. G. (1982) *Eur. J. Biochem. 121*, 413–419.
41. Endo, M., Suzuki, K., Schmid, K., Fournet, B., Montreuil, J., Dorland, L., Van Halbeek, H. & Vliegenthart, J. F. G. (1982) *J. Biol. Chem. 257*, in the press.
42. Pierce-Crétel, A., Pamblanco, M., Strecker, G., Montreuil, J. & Spik, G. (1981) *Eur. J. Biochem. 114*, 169–178.
43. Soulier, S., Sarfati, R. S. & Szabó, L. (1980) *Eur. J. Biochem. 108*, 465–472.
44. Lohmander, L. S., De Luca, S., Nilsson, B., Hascall, V. C., Caputo, C. B., Kimura, J. H. & Heinegård, D. (1980) *J. Biol. Chem. 255*, 6084–6091.
45. Saito, T., Itoh, T., Adachi, S., Suzuki, T. & Usui, T. (1981) *Biochim. Biophys. Acta, 678*, 257–267.
46. Schachter, H., McGuire, E. J. & Roseman, S. (1971) *J. Biol. Chem. 246*, 5321–5328.
47. Baker, A. P. & Munro, J. R. (1971) *J. Biol. Chem. 246*, 4358–4362.
48. Williams, D., Longmore, G., Matta, K. L. & Schachter, H. (1980) *J. Biol. Chem. 255*, 11253–11261.
49. Scudder, P. R. & Chantler, E. N. (1981) *Biochim. Biophys. Acta, 660*, 128–135.
50. Sheares, B. T., Lau, J. T. Y. & Carlson, D. M. (1982) *J. Biol. Chem. 257*, 599–602.

H. van Halbeek, L. Dorland, and J. F. G. Vliegenthart,
Organisch-Chemisch Laboratorium, Afdeling Bio-Organische Chemie, Rijksuniversiteit (te) Utrecht,
Croesestraat 79, NL-3522 AD Utrecht, The Netherlands

W. E. Hull, Bruker Analytische Messtechnik, GmbH, Silberstreifen,
D-7512 Rheinstetten 4-Forchheim, Federal Republic of Germany

G. Lamblin, M. Lhermitte, A. Boersma, and P. Roussel,
Unité de Recherches sur la Biochimie des Protéines (Unité 16 de l'Institut National de la Santé et de la Recherche Médicale),
Place de Verdun, F-59045 Lille-Cedex, France