Eur. J. Biochem. *157*, 139–146 (1986)


# Typing of core and backbone domains of mucin-type oligosaccharides from human ovarian-cyst glycoproteins by 500-MHz $^1$H-NMR spectroscopy

Johanna H. G. M. MUTSAERS[1], Herman VAN HALBEEK[1], Johannes F. G. VLIEGENTHART[1], Albert M. WU[2] and Elvin A. KABAT[3]

[1] Department of Bio-Organic Chemistry, University of Utrecht
[2] Department of Veterinary Pathology, Texas A & M University, Galveston, Texas
[3] Departments of Microbiology, Human Genetics and Development and Neurology and the Cancer Center, College of Physicians and Surgeons, Columbia University, Columbia, Missouri

(Received December 18, 1985) – EJB 85 1376

Human blood-group A active glycoproteins from ovarian-cyst fluid were subjected to Smith degradation and subsequent β-elimination. The resulting oligosaccharide-alditols represent the core and backbone domains of the *O*-linked carbohydrate chains. Nine of these, ranging in size from disaccharides to hexasaccharides, were investigated by $^1$H-NMR spectroscopy. Their primary structures could be adequately characterized. In particular, the core types, i.e. the substitution patterns of *N*-acetylgalactosaminitol (GalNAc-ol) as well as the types of backbone, i.e. the linkage types of alternating Gal-GlcNAc sequences, were unambiguously identified. The core type GlcNAcβ(1–3)GalNAc-ol is described for the first time as occurring in ovarian-cyst glycoprotein.

The carbohydrate chains of the water-soluble blood-group active glycoproteins purified from human ovarian-cyst fluid have been investigated extensively [1–5]. They mainly comprise large-size *O*-glycosidically linked oligosaccharides, which may bear blood-group determinants like A, B, H or $Le^a$, $Le^b$ at their peripheral domains.

In order to study the core and backbone portions of these chains, the terminal sugars constituting these blood-group determinants may be removed by Smith degradation of the glycoprotein. Subsequently, the remaining carbohydrates are released from the peptides by alkaline borohydride treatment. Thus, the resulting oligosaccharide-alditols have been characterized to a certain extent by methylation analysis, periodate oxidation and gas-liquid chromatography/mass spectrometry analysis of the permethylated, *N*-trifluoroacetylated intact compounds [3]. Here, we report on the structural analysis of these oligosaccharide-alditols employing an independent technique, namely, 500-MHz $^1$H-NMR spectroscopy. This method is well suited for the elucidation of the complete primary structure of mucin-type oligosaccharide-alditols, in particular in typing of the core units [6, 7].

## MATERIALS AND METHODS

Oligosaccharide-alditols were obtained from Smith-degraded blood-group A active glycoproteins from ovarian-cyst fluid, as described [3]. Nine of these were analysed by 500-MHz $^1$H-NMR spectroscopy.

Samples were repeatedly exchanged in $^2H_2O$ (99.96 atom% $^2$H), with intermediate lyophilization. 500-MHz $^1$H-NMR spectra were recorded on a Bruker WM 500 instrument (SON hf-NMR facility, Department of Biophysical Chemistry, Nijmegen University, The Netherlands) operating in the pulsed Fourier transform mode. The probe temperature was kept at 22.0 (± 0.1) °C, in order to ensure complete visualization of the H-1 signals in the spectral region $4.4 < \delta < 4.8$ ppm ($\delta$ $HO^2H$ = 4.81 ppm). Resolution enhancement of the spectra was achieved by Lorentzian-to-Gaussian transformation from quadrature phase detection. Chemical shifts ($\delta$) are expressed at 22 °C downfield from internal sodium 4,4-dimethyl-4-silapentane-1-sulfonate, but were actually measured by reference to internal acetone ($\delta$ = 2.225 ppm) with accuracy of 0.002 ppm.

## RESULTS

A large number of oligosaccharide-alditols were obtained by alkaline borohydride treatment of Smith-degraded blood-group A active glycoproteins from ovarian-cyst fluid [3]. Sugar analysis revealed the presence of Gal, GlcNAc and GalNAc-ol as the only sugar constituents. Nine of the oligosaccharides were available in sufficient amount to permit 500-MHz $^1$H-NMR spectroscopic analysis. For reasons of consistency, the compounds are designated according to the denoting system used in [3] (see Scheme 1).

The resonance position of GalNAc-ol H-2 in the $^1$H-NMR spectrum of such oligosaccharide-alditols is known to be sufficiently discriminative to disclose the presence and character of the substituent monosaccharide at C-3 of GalNAc-ol. When observed at $\delta$ = 4.29 ppm, it points to the presence of β-linked GlcNAc [7]; a position around $\delta$ = 4.39 ppm is characteristic of the occurrence of β-linked Gal [7] whereas $\delta$ = 4.24 ppm has been found for compounds lacking a substituent at C-3 [8–10]. On the basis of the chemical shift value of their GalNAc-ol H-2 atom, the ovarian-cyst oligosaccharide-alditols were classified into three groups. According to this division, the $^1$H-NMR parameters of the structural-reporter groups of the compounds are compiled in Tables 1–3.

In Table 1, the NMR data of compound "8B" (for spectrum see Fig. 1) are compared with those of some refer-

*Correspondence to* J. F. G. Vliegenthart, Afdeling Bio-Organische Chemie Transitorium III, Postbus 80.075, NL-3508-TB Utrecht, The Netherlands

"8B" Galβ(1-3)GlcNAcβ(1-3)
\
GalNAc-ol

10 Galβ(1-3)
\
GalNAc-ol

5 Galβ(1-3)GlcNAcβ(1-3)Galβ(1-3)
\
GalNAc-ol

2A Galβ(1-3)GlcNAcβ(1-3)Galβ(1-3)GlcNAcβ(1-3)Galβ(1-3)
\
GalNAc-ol

2B Galβ(1-3)GlcNAcβ(1-3)Galβ(1-3)
/ \
Galβ(1-4)GlcNAcβ(1-6) GalNAc-ol

2C Galβ(1-3)GlcNAcβ(1-3)Galβ(1-3)
\
GalNAc-ol
/
Galβ(1-4)GlcNAcβ(1-6)

9B GalNAc-ol
/
GlcNAcβ(1-6)

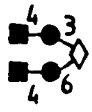
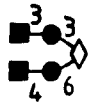
7 GalNAc-ol
/
Galβ(1-4)GlcNAcβ(1-6)

"3C" GalNAc-ol
/
Galβ(1-3)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)

Scheme 1. *Structures of the backbone oligosaccharide-alditols obtained from Smith-degraded blood-group A active human ovarian-cyst glycoproteins as revealed by $^1$H-NMR spectroscopic analysis.* The quotation marks (for 8B and 3C) are used to indicate that the structure based on NMR spectroscopy differs from the one reported previously [3]

Table 1. *$^1$H chemical shifts of structural-reporter-group protons of the constituent monosaccharides for oligosaccharide "8B" derived from ovarian-cyst glycoproteins, together with those of three reference compounds (R1–R3)*

Chemical shifts are expressed downfield from sodium 4,4-dimethyl-4-silapentane-1-sulfonate but were actually measured by reference to internal acetone ($\delta$ = 2.225 ppm) in $^2H_2O$ at 22°C. The complete structures of the compounds are in Scheme 1. In the table heading the structures are represented by short-hand symbolic notation: ◇ = GalNAc-ol, ■ = Gal, ● = GlcNAc. A superscript at the name of a sugar residue indicates to which position of the adjacent monosaccharide it is glycosidically linked. More than one superscript are used to discriminate between identically bound residues; they indicate the types of linkages of the neighbouring residues in the sequence. n.d. not determined

| Residue | Reporter group | Chemical shift in | | | |
|---|---|---|---|---|---|
| | | R1 [7] | "8B" | R2 [11] | R3 [11] |
| | |  |  |  |  |
| | | ppm | | | |
| GalNAc-ol | H-2 | 4.290 | 4.287 | 4.284 | 4.287 |
| | H-3 | 4.002 | 4.011 | n.d. | n.d. |
| | H-4 | 3.552 | 3.563 | n.d. | n.d. |
| | H-5 | 4.143 | 4.135 | 4.242 | 4.233 |
| | NAc | 2.038 | 2.034 | 2.045 | 2.043 |
| GlcNAc$^3$ | H-1 | 4.631 | 4.651 | 4.623 | 4.648 |
| | H-3 | 3.6–3.7 | 3.913 | n.d. | n.d. |
| | H-6 | 4.021 | 3.953 | 4.021 | 3.955 |
| | NAc | 2.083 | 2.072 | 2.079 | 2.069 |
| Gal$^{4,3}$ | H-1 | 4.455 | – | 4.451 | – |
| | H-4 | 3.926 | – | 3.923 | – |
| Gal$^3$ | H-1 | – | 4.460 | – | 4.453 |
| | H-4 | – | 3.918 | – | 3.92 |
| GlcNAc$^6$ | H-1 | – | – | 4.558 | 4.559 |
| | NAc | – | – | 2.062 | 2.062 |
| Gal$^{4,6}$ | H-1 | – | – | 4.473 | 4.472 |
| | H-4 | – | – | 3.923 | 3.92 |

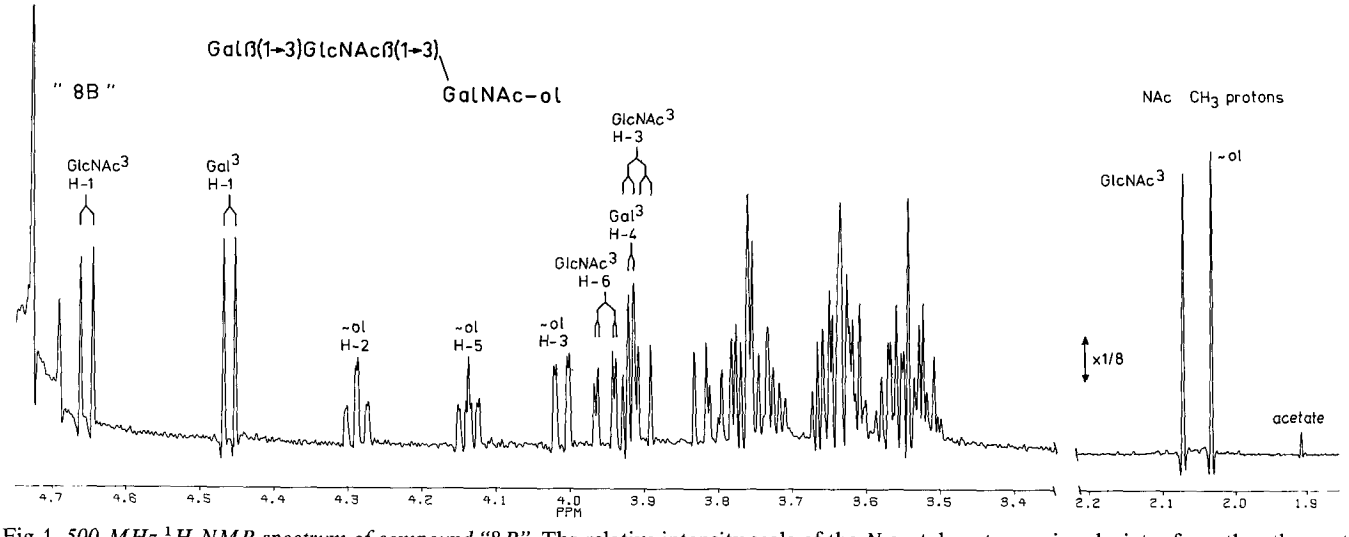
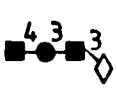
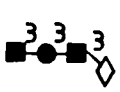
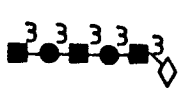
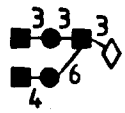
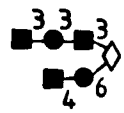



Fig. 1. *500-MHz $^1$H-NMR spectrum of compound "8B".* The relative intensity scale of the *N*-acetyl proton region deviates from the other part of the spectrum as indicated

Table 2. *$^1H$ chemical shifts of structural-reporter-group protons of the constituent monosaccharides for oligosaccharides 10, 5, 2A, 2B and 2C derived from ovarian-cyst glycoproteins, together with those of reference compound R4*

Chemical shifts are expressed as described in Table 1. For complete structures of the compounds see Scheme 1 and for symbolic notation and explanation of superscripts see Table 1. The number of superscripts is limited to three. In case of compound 2A $Gal^3$ is the first Gal residue, $Gal^{3,3}$ the second and $Gal^{3,3,3}$ the third residue encountered starting from the GalNAc-ol side. n.d. not determined

| Residue | Reporter group | Chemical shift in | | | | | |
|---|---|---|---|---|---|---|---|
| | | 10 | R4 [11] | 5 | 2A | 2B | 2C |
| | |  |  |  |  |  |  |
| | | ppm | | | | | |
| GalNAc-ol | H-2 | 4.393 | 4.400 | 4.401 | 4.401 | 4.389 | 4.406 |
| | H-3 | 4.063 | 4.051 | 4.050 | 4.052 | 4.020 | 4.056 |
| | H-4 | 3.506 | 3.497 | 3.493 | 3.497 | 3.503 | 3.443 |
| | H-5 | 4.193 | 4.185 | 4.184 | 4.188 | 4.142 | 4.275 |
| | NAc | 2.050 | 2.047 | 2.046 | 2.048 | 2.046 | 2.067 |
| $Gal^3$ | H-1 | 4.476 | 4.464 | 4.465 | 4.463 | 4.468 | 4.453 |
| | H-4 | 3.901 | 4.126 | 4.128 | 4.128 | 4.118 | 4.133 |
| $GlcNAc^3$ | H-1 | – | 4.688 | 4.701 | 4.696 | 4.701 | 4.700 |
| | H-3 | – | n.d. | 3.912 | 3.91 | 3.949 | 3.91 |
| | H-6 | – | 3.954 | 3.900 | 3.899 | 3.903 | 3.901 |
| | NAc | – | 2.042 | 2.034 | 2.039 | 2.032 | 2.032 |
| $Gal^4$ | H-1 | – | 4.481 | – | – | 4.468 | 4.472 |
| | H-4 | – | 3.927 | – | – | 3.922 | 3.926 |
| $Gal^{3,3}$ | H-1 | – | – | 4.449 | 4.441 | 4.451 | 4.453 |
| | H-4 | – | – | 3.912 | 4.143 | 3.912 | 3.916 |
| $GlcNAc^{3,3}$ | H-1 | – | – | – | 4.742 | – | – |
| | H-3 | – | – | – | 3.91 | – | – |
| | H-6 | – | – | – | 3.899 | – | – |
| | NAc | – | – | – | 2.024 | – | – |
| $Gal^{3,3,3}$ | H-1 | – | – | – | 4.448 | – | – |
| | H-4 | – | – | – | 3.910 | – | – |
| $GlcNAc^6$ | H-1 | – | – | – | – | 4.607 | 4.556 |
| | H-6 | – | – | – | – | 3.995 | 3.998 |
| | NAc | – | – | – | – | 2.053 | 2.059 |

ence compounds of related structure. It appears that the reporter-group chemical shifts of "8B" match those of the type-1 trisaccharide Gal$\beta$(1 – 3)GlcNAc$\beta$(1 – 3)GalNAc-ol (compound 5b [7]), obtained from bronchial mucins of cystic fibrosis patients. Therefore, the structure of compound "8B" is

Gal$\beta$(1 – 3)GlcNAc$\beta$(1 – 3)
\
GalNAc-ol.

This result deviates from the structure of 8B reported in [3], in the type of linkage of GlcNAc to GalNAc-ol being $\beta$(1 – 3) rather than $\beta$(1 – 6). For comparison of the data of "8B" with those of oligosaccharides that do have GlcNAc$\beta$(1 – 6)GalNAc-ol, see Table 3.

To establish some general rules for defining the linkage type in a Gal-GlcNAc unit by $^1$H-NMR spectroscopy as $\beta$(1 – 4) or $\beta$(1 – 3), a comparison was made of the chemical-shift data of compounds "8B" and R1, i.e. Gal$\beta$(1 – 4)-GlcNAc$\beta$(1 – 3)GalNAc-ol [7] (see Table 1). This shows that the chemical shifts of the structural-reporter groups of both Gal and GlcNAc undergo significant shift alterations in the step from Gal$\beta$(1 – 4)GlcNAc (type 2) to Gal$\beta$(1 – 3)GlcNAc (type 1) in the backbone sequence (see Table 4). In addition, it may be mentioned that upon substitution of GlcNAc at C-3, the GlcNAc H-3 emerges out of the bulk of skeleton protons ($\delta \approx 3.7$ ppm) to $\delta = 3.913$ ppm, becoming a reporter group for the $\beta$(1 – 3) type of linkage. The $\Delta\delta$ values (Table 4) are not influenced by the presence of an additional substituent at C-6 of GalNAc-ol, as may become evident from comparison of the chemical-shift data listed for compounds R2 and R3 [11] (see Tables 1 and 4). However, some of them are affected by the type of linkage in which GlcNAc is involved and the nature of the residue to which it is attached (cf. compound 5- and R4, see below and *N*-type asialo glycopeptides from hemopexin [12]; Table 4). It was concluded that compounds 10, 5, 2A, 2B and 2C have in common the Gal$\beta$(1 – 3)Gal-NAc-ol core type (GalNAc-ol H-2 at $\delta = 4.39$ ppm; see Table 2).

In the spectrum of compound 10 only signals belonging to the aforementioned Gal$\beta$(1 – 3)GalNAc-ol sequence are present. The spectrum is identical to that of Gal$\beta$(1 – 3)-

GalNAc-ol obtained from bronchial mucin [7]. This result is in accordance with [3].

The NMR spectrum of compound 5 (see Fig. 2a) shows two *N*-acetyl signals around $\delta$ = 2.0 ppm and three doublets in the $\beta$-anomeric region. This suggests the presence of one GlcNAc, one GalNAc-ol and two Gal residues, making 5 a tetrasaccharide. Furthermore, the H-5 signal of GalNAc-ol is observed at $\delta$ = 4.184 ppm showing that no substituent is present at C-6 of GalNAc-ol [6, 7]. The core Gal[3] residue ($\delta$ H-1 = 4.465 ppm) bears a $\beta$-linked GlcNAc ($\delta$ H-1 = 4.701 ppm) at C-3 as can be deduced from the chemical shift of H-4 of Gal[3], being 4.128 ppm (compare R4 in Table 2). The GlcNAc[3] residue is substituted with a $\beta$-linked Gal residue ($\delta$ H-1 = 4.449 ppm). To identify the position of the linkage between Gal and GlcNAc[3] the chemical shift of the structural reporters of the terminal Gal and the GlcNAc[3] residue of 5 were compared to the corresponding residues in compound R4 Gal$\beta$(1−4)GlcNAc$\beta$(1−3)Gal$\beta$(1−3)GalNAc-ol (see Table 2). The chemical shift differences are included in Table 4. Moreover, for compound 5 the H-3 signal of GlcNAc[3] is found at $\delta$ = 3.912 ppm. These chemical shift deviations with respect to compound R4 reveal compound 5 to possess the type-1 rather than the type-2 Gal-GlcNAc sequence:

Gal$\beta$(1−3)GlcNAc$\beta$(1−3)Gal$\beta$(1−3)
\
GalNAc-ol .

A similar trend of chemical shift alteration, at least for GlcNAc H-1 and Gal H-1, have been observed going from lacto-*N*-neotetraose to lacto-*N*-tetraose [13]. The structure of compound 5 based on NMR is in accordance with [3].

Compounds 2A, 2B and 2C were found to be hexasaccharides, consisting of Gal, GlcNAc and GalNAc-ol residues in the molar ratio 3:2:1, as can be derived from the number of *N*-acetyl signals, the number of anomeric signals and the occurrence of GalNAc-ol H-2 and H-5 signals in the NMR spectra.

Compound 2A (for spectrum see Fig. 2b) was found to be an extension of compound 5 with another Gal-GlcNAc unit. The attachment of this unit in $\beta$-linkage to C-3 of Gal[3,3] was deduced from the chemical shift of the H-4 signal of Gal[3,3] being $\delta$ = 4.143 ppm. The H-1 and H-4 signals of the terminal Gal residue at $\delta$ = 4.448 ppm and 3.910 ppm, respectively, indicate the terminal Gal to be $\beta$(1−3) linked to GlcNAc[3,3]. Therefore, the structure of 2A is

Gal$\beta$(1−3)GlcNAc$\beta$(1−3)Gal$\beta$(1−3)GlcNAc$\beta$(1−3)-
Gal$\beta$(1−3)
\
GalNAc-ol .

This is in accordance with [3].

Table 3. *$^1H$ chemical shifts of structural-reporter-group protons of the constituent monosaccharides for oligosaccharides 9B, 7 and "3C" derived from ovarian-cyst glycoproteins*
Chemical shifts are expressed as described in Table 1. For complete structures of the compounds see Scheme 1. For explanation of the symbolic notation and superscripts see Table 1

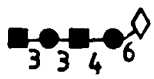
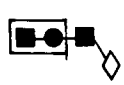
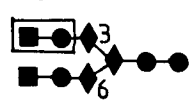
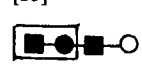
| Residue | Reporter group | Chemical shift in | | |
|---|---|---|---|---|
| | | 9B  | 7  | "3C"  |
| | | ppm | | |
| GalNAc-ol | H-2 | 4.242 | 4.243 | 4.242 |
| | H-3 | 3.841 | 3.843 | 3.841 |
| | H-4 | 3.379 | 3.379 | 3.379 |
| | H-5 | 4.021 | 4.026 | 4.021 |
| | H-6 | 3.933 | 3.932 | 3.932 |
| | NAc | 2.046 | 2.045 | 2.042 |
| GlcNAc[6] | H-1 | 4.553 | 4.577 | 4.573 |
| | H-6 | 3.928 | 3.993 | 3.990 |
| | NAc | 2.059 | 2.061 | 2.060 |
| Gal[4] | H-1 | – | 4.471 | 4.460 |
| | H-4 | – | 3.925 | 4.153 |
| GlcNAc[3] | H-1 | – | – | 4.722 |
| | H-6 | – | – | 3.898 |
| | NAc | – | – | 2.028 |
| Gal[3] | H-1 | – | – | 4.440 |
| | H-4 | – | – | 3.911 |

Table 4. *Chemical shift differences for pertinent reporter groups going from a Gal$\beta$(1−4)GlcNAc$\beta$(1−.) to a Gal$\beta$(1−3)GlcNAc$\beta$(1−.) sequence*
Chemical shift differences are expressed in the step from a compound possessing a Gal$\beta$(1−4)GlcNAc$\beta$(1−.) to a Gal$\beta$(1−3)GlcNAc$\beta$(1−.) sequence. These sequences are outlined in the table-heading. In the table-heading the structures are represented by short-hand symbolic notation: ◇ = GalNAc-ol, ■ = Gal, ● = GlcNAc, ◆ = Man, and ○ = Glc. n.d. not determined

| Residue | Reporter group | Chemical shift differences in | | | | |
|---|---|---|---|---|---|---|
| | |  |  |  | [12]  | [13]  |
| | | ppm | | | | |
| GlcNAc | H-1 | +0.020 | +0.025 | +0.013 | +0.018 | +0.032 |
| | H-6 | −0.068 | −0.066 | −0.054 | n.d. | n.d. |
| | NAc | −0.011 | −0.010 | −0.008 | −0.005 | n.d. |
| Gal | H-1 | +0.005 | +0.002 | −0.032 | −0.022 | −0.043 |
| | H-4 | −0.008 | n.d. | −0.015 | n.d. | n.d. |

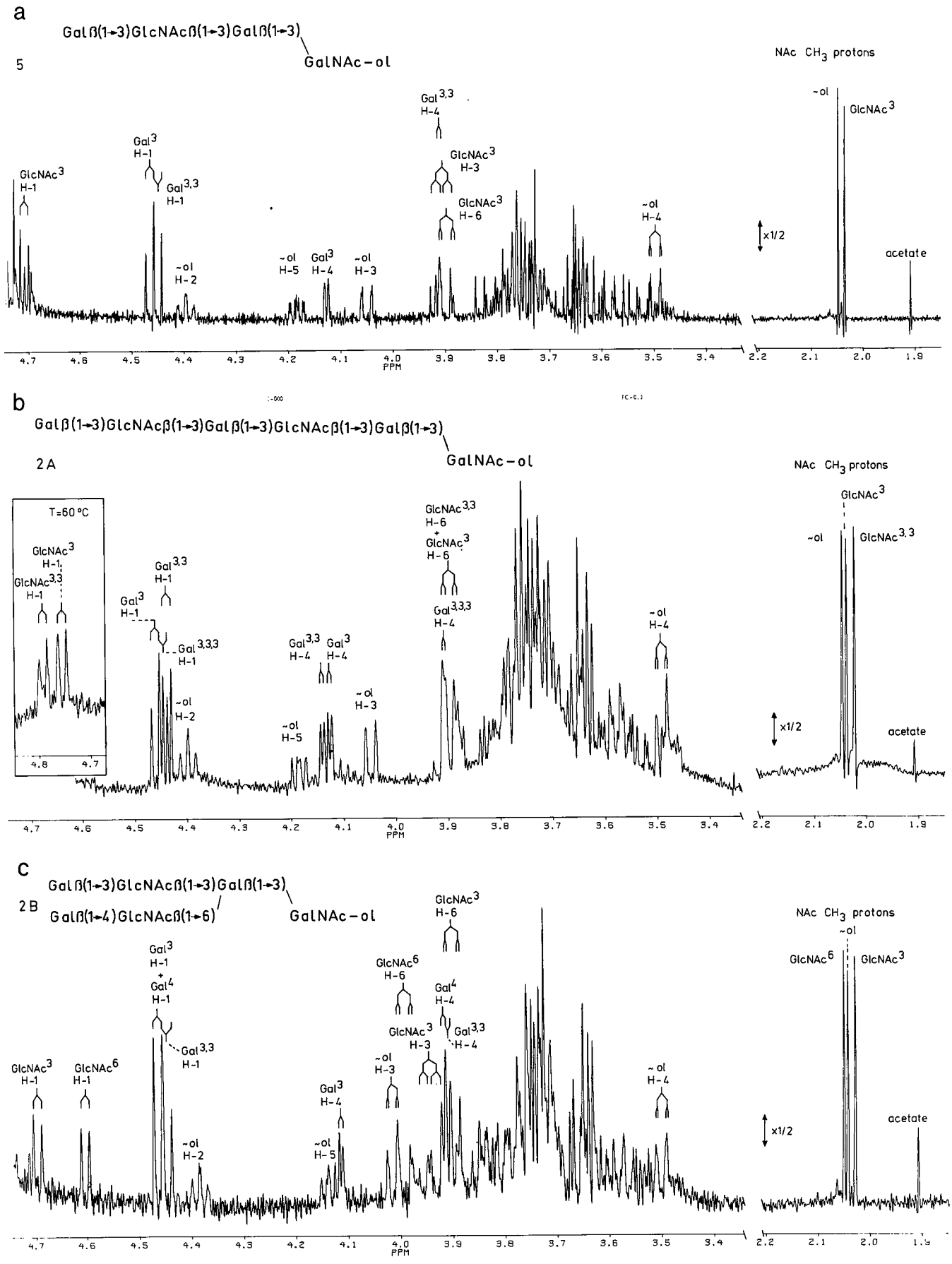



Fig. 2. *500-MHz $^1$H-NMR spectrum of (a) compound 5, (b) compound 2A (the inset shows the GlcNAc H-1 signals at 60 °C) and (c) compound 2B.* The relative intensity scale of the *N*-acetyl proton region deviates from that of the other part of the respective spectra, as indicated

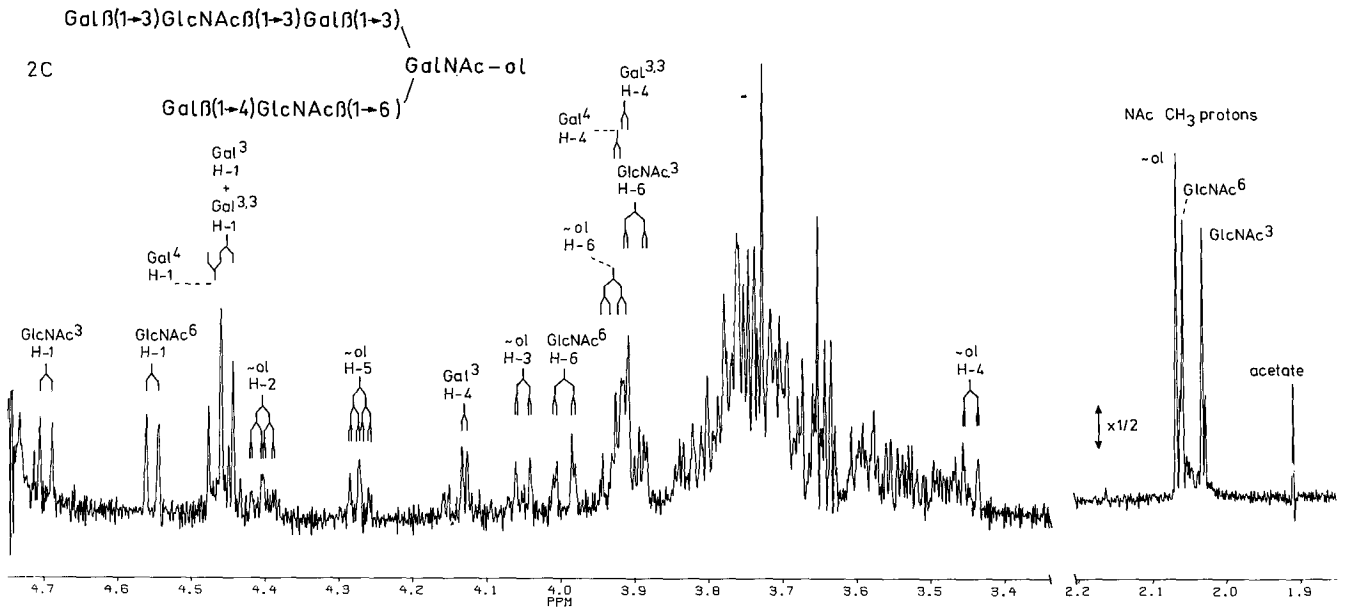



Fig. 3. *500-MHz $^1$H-NMR spectrum of compound 2C*. The relative intensity scale of the *N*-acetyl proton region deviates from the other part of the spectrum as indicated

Compound 2B (for spectrum see Fig. 2c) is also an extension of compound 5 with a Gal-GlcNAc unit, but attached at a different position. The H-1 signal of the additional GlcNAc found at 4.607 ppm, suggests that the additional Gal-GlcNAc unit is linked $\beta(1-6)$ to Gal$^3$. This was previously observed in [14]. The Gal$^3$ residue is thus serving as a branching point in the structure. In the additional Gal-GlcNAc unit the H-1 and H-4 signals of the terminal Gal and the H-6 signal of GlcNAc$^6$ are observed at 4.468 ppm, 3.922 ppm and 3.995 ppm, respectively. They show that this Gal residue is $\beta(1-4)$ linked to GlcNAc$^6$. Further support for the presence of a Gal$\beta(1-4)$GlcNAc$\beta(1-.)$ sequence linked to C-6 of Gal$^3$ stems from the chemical shift effects of the H-3 and H-5 signals of GalNAc-ol and on the H-4 of Gal$^3$. These are in line with those observed earlier [14]. Therefore, the structure of 2B is proposed to be as follows

Gal$\beta(1-3)$GlcNAc$\beta(1-3)$Gal$\beta(1-3)$
/ \
Gal$\beta(1-4)$GlcNAc$\beta(1-6)$ GalNAc-ol .

This structure is in accordance with [3].

Compound 2C (for spectrum see Fig. 3) is another extension of compound 5 with a Gal-GlcNAc unit. The branch attached to C-3 of GalNAc-ol is identical to that of compound 5 (see Table 2). The position of the H-5 signal of GalNAc-ol ($\delta$ = 4.275 ppm) shows that a $\beta$-linked GlcNAc is present at C-6 of GalNAc-ol [7]. The additional Gal residue is $\beta(1-4)$ linked to this GlcNAc$^6$ as may become clear from comparison of the chemical shift data with those of R2 and R3 [11] (see Table 1). Therefore, the structure of compound 2C is

Gal$\beta(1-3)$GlcNAc$\beta(1-3)$Gal$\beta(1-3)$
\
GalNAc-ol.
/
Gal$\beta(1-4)$GlcNAc$\beta(1-6)$

This result is in accordance with [3].

The NMR spectra of the compounds 9B, 7 and "3C" have in common the H-2 signal of GalNAc-ol at 4.24 ppm. This position is indicative of the absence of a substituent sugar at GalNAc-ol C-3 (cf. [8, 9]).

The $\beta$-anomeric region of the spectrum of 9B (see Fig. 4a) shows only one doublet; therefore, compound 9B is a disaccharide. It consists of a GlcNAc residue ($\delta$ H-1 = 4.553 ppm, $J_{1,2}$ = 8.1 Hz, $\delta$ H-6 = 3.928 ppm and $\delta$ NAc = 2.059 ppm) $\beta(1-6)$ linked to GalNAc-ol [9]. Therefore, the structure of 9B is

GalNAc-ol
/
GlcNAc$\beta(1-6)$.

This is in accordance with [3].

Compound 7 (for spectrum see Fig. 4b) was found to be an extension of 9B with a Gal residue ($\delta$ H-1 = 4.471 ppm, $J_{1,2}$ = 8.0 Hz, $\delta$ H-4 = 3.925 ppm). The latter is $\beta(1-4)$ linked to GlcNAc$^6$, as could be readily deduced by comparison of the relevant chemical shift values of compound 7 with those of R2 and R3 [11] (see Table 1) cf. [9]. Therefore, the structure of compound 7 is

GalNAc-ol
/
Gal$\beta(1-4)$GlcNAc$\beta(1-6)$.

This is in accordance with [3].

In turn, compound "3C" (for spectrum see Fig. 4c) is an extension of compound 7 with another Gal-GlcNAc unit. The H-4 signal of Gal$^4$ at 4.155 ppm shows the additional GlcNAc to be $\beta(1-3)$ linked to Gal$^4$. The H-1 signal of this GlcNAc$^3$ at 4.721 ppm ($J_{1,2}$ = 8.4 Hz) corroborates the $\beta$-configuration of the linkage. The set of chemical shifts of H-1, H-6 and NAc of GlcNAc$^3$, and of H-1 and H-4 of the terminal Gal residue (Table 3), point to the $\beta(1-3)$ type of linkage of this Gal to GlcNAc$^3$. Therefore, the structure of compound "3C" is

GalNAc-ol
/
Gal$\beta(1-3)$GlcNAc$\beta(1-3)$Gal$\beta(1-4)$GlcNAc$\beta(1-6)$.

This result is not in accordance with [3].

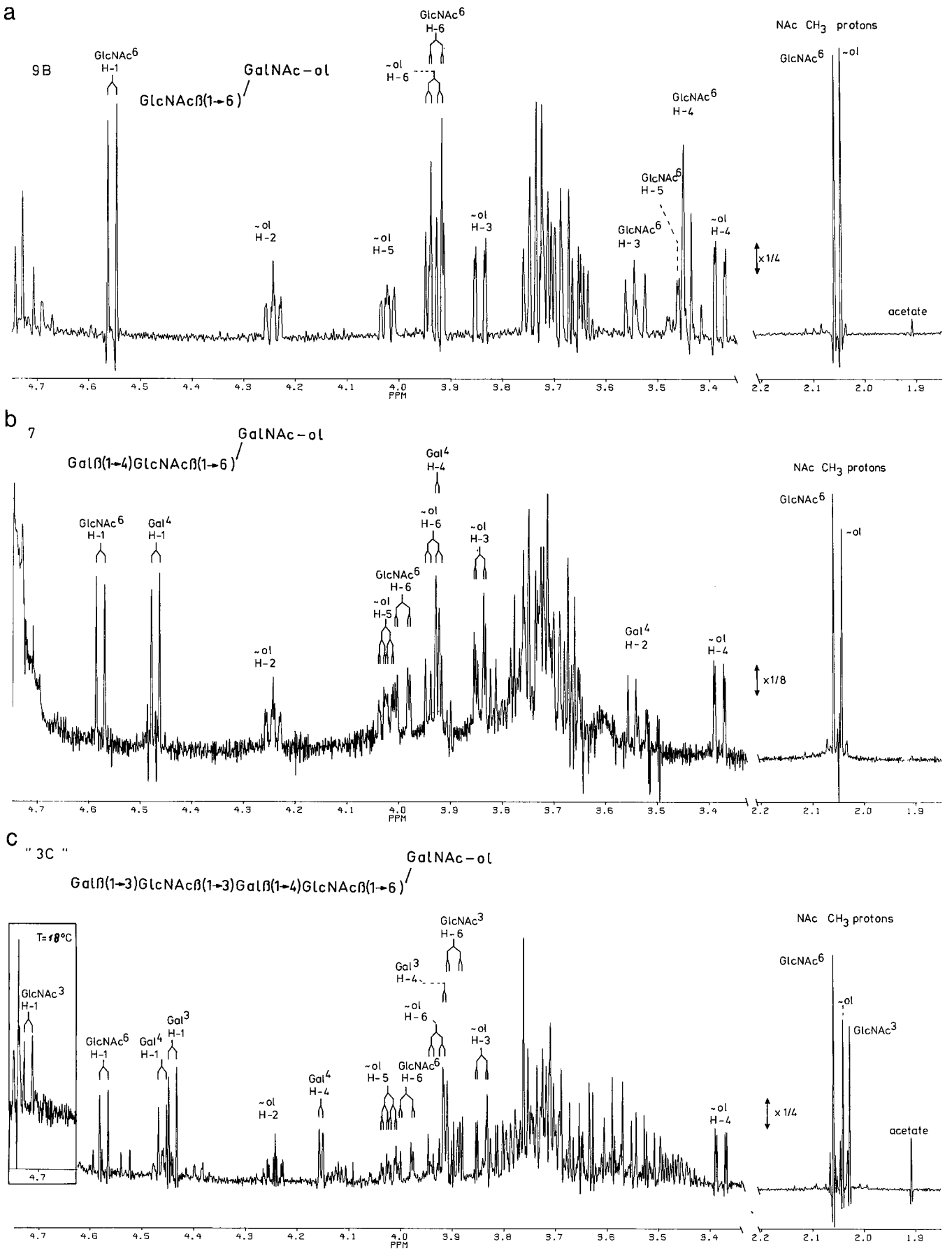
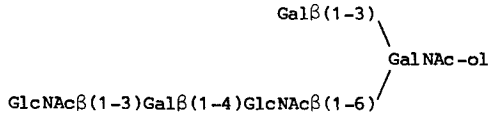



Fig. 4. *500-MHz $^1$H-NMR spectrum of (a) compound 9B, (b) compound 7, (c) compound "3C".* The inset shows the GlcNAc H-1 signals at 18 °C. The relative intensity scale of the *N*-acetyl proton region deviates from that of the other part of the respective spectra, as indicated

## DISCUSSION

500-MHz $^1$H-NMR spectroscopy of nine oligosaccharide-alditols from Smith-degraded blood-group A active glycoproteins from ovarian-cyst fluid allowed a further extension of the scope of this technique to the typing of core and backbone structural elements of mucin-type carbohydrates. From the chemical shifts of the H-2 and H-5 signals of GalNAc-ol, the presence and nature of the substituents at C-3 and C-6 could be readily inferred [6, 7]. In addition, the number of Gal and GlcNAc residues and the types of linkage between residues in linear and branched sequences could be established. In particular, for the sequence Gal$\beta$(1-*x*)GlcNAc the *x* could be established to be 3 or 4 on the basis of subtle, but significant differences in chemical shift of structural reporter groups (Table 4). The elucidation of the structures of the hexasaccharide positional isomers 2A, 2B and 2C is an excellent illustration of the power of high-resolution $^1$H-NMR spectroscopy. It should be noted that comparison of the chemical shift data and the spectrum of 2C with those presented for the core pentasaccharide Cc in [16] suggests that the same compounds are concerned; nonetheless, Cc had been interpreted in terms of the structure




The structures elucidated here for the backbones of the carbohydrate chains from blood-group A active glycoproteins from ovarian-cyst fluid are in line with those reported previously [3] with two exceptions: compounds "8B" and "3C". The reasons for the discrepancies are not clear, as yet. Compound "3C" deviates from the branched structure reported [3] since it appeared to be a linear pentasaccharide in which the GalNAc-ol is substituted at C-6 only. On the other hand, the structure of compound "8B" is novel for ovarian-cyst glycoproteins. Structures possessing a GlcNAc residue $\beta$(1-3) linked to GalNAc-ol were not reported before for human ovarian-cyst glycoproteins [1, 3, 15, 16], but they occur in horse gastric, sheep gastric, rat colonic and human bronchial mucins [7, 17-19].

Structures possessing a $\beta$(1-6) linked GlcNAc as the only substituent at GalNAc-ol seldom occur in nature. They have been reported for human meconium [8] and human $\kappa$-casein [9]. The compounds "3C", 7 and 9B described here resulted from Smith-degraded glycoprotein. Hence in the original material a substituent at C-3 of GalNAc-Ser/Thr must have been present.

This investigation was supported by the Netherlands Foundation for Chemical Research (SON/ZWO), by the Netherlands Foundation for Cancer Research (KWF, grant UUCK 83-13), by the National Science Foundation (grant PCM 81-02321) to E. A. K., by Cancer Center Support (grant CA 13696) to Columbia University, and by the Texas Agricultural Experiment Station (grant TAES H6194) and USDA/SEA Formula Animal Health Funds (project 6648).

## REFERENCES

1. Rovis, L., Anderson, B., Kabat, E. A., Gruezo, F. & Liao, J. (1973) *Biochemistry 12*, 5340-5353.
2. Maisonrouge-McAuliffe, F. & Kabat, E. A. (1976) *Arch. Biochem. Biophys. 175*, 90-113.
3. Wu, A. M., Kabat, E. A., Nilsson, B., Zopf, D. A., Gruezo, F. & Liao, J. (1984) *J. Biol. Chem. 259*, 7178-7186.
4. Watkins, W. M. (1972) in *Glycoproteins* (Gottschalk, A., ed.) 2nd edn, pp. 830-891, Elsevier, Amsterdam.
5. Watkins, W. M., Yates, A. D. & Greenwell, P. (1981) *Biochem. Soc. Trans. 9*, 186-191.
6. Van Halbeek, H. (1984) *Biochem. Soc. Trans. 12*, 601-605.
7. Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Hull, W. E., Lamblin, G., Lhermitte, M., Boersma, A. & Roussel. P. (1982) *Eur. J. Biochem. 127*, 7-20.
8. Hounsell, E. F., Lawson, A. M., Feeney, J., Gooi, H. C., Pickering, N. J., Stoll, M. S., Lui, S. G. & Feizi, T. (1985) *Eur. J. Biochem. 148*, 367-377.
9. Van Halbeek, H., Vliegenthart, J. F. G., Fiat, A.-M. & Jollès, P. (1985) *FEBS Lett. 187*, 81-88.
10. Vliegenthart, J. F. G., Dorland, L., Van Halbeek, H. & Haverkamp, J. (1982) *Cell Biol. Monogr. 10*, 127-172.
11. Lamblin, G., Boersma, A., Lhermitte, M., Roussel, P., Mutsaers, J. H. G. M., Van Halbeek, H. & Vliegenthart, J. F. G. (1984) *Eur. J. Biochem. 143*, 227-236.
12. Bernard, N., Engler, R., Strecker, G., Montreuil, J., Van Halbeek, H. & Vliegenthart, J. F. G. (1984) *Glycoconjugate J. 1*, 123-140.
13. Dua, V. K. & Bush, C. A. (1983) *Anal. Biochem. 133*, 1-8.
14. Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Kochetkov, N. K., Arbatsky, N. P. & Derevitskaya, V. A. (1982) *Eur. J. Biochem. 127*, 21-29.
15. Tanaka, M., Dube, V. E. & Anderson, B. (1984) *Biochim. Biophys. Acta 798*, 283-290.
16. Dua, V. K., Dube, V. E., Li, Y. T. & Bush, C. A. (1985) *Glycoconjugate J. 2*, 17-30.
17. Newman, W. & Kabat, E. A. (1976) *Arch. Biochem. Biophys. 172*, 535-550.
18. Hounsell, E. F., Fukuda, M., Powell, M. E., Feizi, T. & Hakomori, S. (1980) *Biochem. Biophys. Res. Commun. 92*, 1143-1150.
19. Slomiany, B. L., Murty, V. L. N. & Slomiany, A. (1980) *J. Biol. Chem. 255*, 9719-9723.