THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 261, No. 5, Issue of February 15, pp. 1992-1997, 1986
Printed in U.S.A.

Structure of Sialyloligosaccharides Isolated from Bonnet Monkey (*Macaca radiata*) Cervical Mucus Glycoproteins Exhibiting Multiple Blood Group Activities*

(Received for publication, July 17, 1985)

Nasir-Ud-Din and Roger W. Jeanloz

From the Laboratory for Carbohydrate Research, Department of Biological Chemistry and Medicine, Harvard Medical School and Massachusetts General Hospital, Boston, Massachusetts 02114

Geneviève Lamblin and Philippe Roussel‡

From the Institut National de la Santé et de la Recherche Médicale, Unité No. 16, Biochimie des Protéines, Place de Verdun, F-59045 Lille, France

Herman van Halbeek, Johanna H. G. M. Mutsaers, and Johannes F. G. Vliegenthart

From the Department of Bio-organic Chemistry, University of Utrecht, Croesestraat 79, NL-3522 AD Utrecht, The Netherlands

Mucin glycoproteins purified from cervical epithelial secretion of the bonnet monkey (*Macaca radiata*) exhibit multiple blood group activities. Alkaline borohydride reductive cleavage resulted in a mixture of neutral and acidic oligosaccharide-alditols. By high-performance liquid chromatography, seven oligosaccharides (A-4-1 to A-4-7) have been purified from the monosialyloligosaccharide fraction (A-4). Based on the results of 500-MHz ^{1}H NMR spectroscopy, in conjunction with sugar analysis and immunological assays, we propose the following structures for these oligosaccharides.

```
                    GalNAc-ol
                   /                                  (A-4-1)
        NeuAcα(2→6)
          Galβ(1→3)
                    \
                    GalNAc-ol                         (A-4-2)
                   /
        NeuAcα(2→6)
   Fucα(1→2)Galβ(1→3)
                    \
                    GalNAc-ol                         (A-4-3)
                   /
        NeuAcα(2→6)
              Galβ(1→3)
                    \
                    GalNAc-ol                         (A-4-4)
                   /
GalNAcβ(1→4)Galβ(1→4)GlcNAcβ(1→6)
                   /
        NeuAcα(2→3)
```

```
                    Fucα(1→2)Galβ(1→3)
                                      \
                                       GalNAc-ol      (A-4-5)
                                      /
GalNAcβ(1→4)Galβ(1→4)GlcNAcβ(1→6)
                   /
        NeuAcα(2→3)
              GalNAcα(1→3)
                    \
                    Fucα(1→2)Galβ(1→3)
                                      \
                                       GalNAc-ol      (A-4-6)
                                      /
GalNAcβ(1→4)Galβ(1→4)GlcNAcβ(1→6)
                   /
        NeuAcα(2→3)
                    Galα(1→3)
                    \
                    Fucα(1→2)Galβ(1→3)
                                      \
                                       GalNAc-ol      (A-4-7)
                                      /
GalNAcβ(1→4)Galβ(1→4)GlcNAcβ(1→6)
                   /
        NeuAcα(2→3)
```

These structures imply that either the A, B, or H determinant may be found in combination with the Cad/Sd^a determinant; the oligosaccharides identified, together, account for the blood group activities exhibited by the cervical mucus.

Cervical mucus is a gel-like, hydrophilic epithelial secretion, playing a significant role in reproductive physiology (1). The physical and chemical properties of human (2) and bovine (3) cervical secretions change with the state of ovarian function.

* This investigation was supported in part, by Grant AM-03564 from the National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases, by Grant HD-12431 from the National Institute of Child Health and Human Development, National Institutes of Health, by the Netherlands Foundation for Chemical Research with financial aid from the Netherlands Organization for the Advancement of Pure Research, and by Grant UUKC 83-13 from the Netherlands Foundation for Cancer Research. This is publication 989 of the Robert W. Lovett Memorial Group for the Study of Diseases Causing Deformities, Harvard Medical School and Massachusetts General Hospital. A preliminary account of this study was presented at the XIIth International Carbohydrate Symposium, Utrecht, The Netherlands, July 1-7, 1984. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Dedicated to Luis Leloir on the occasion of his 80th birthday.

‡ To whom correspondence should be sent: Institut National de la Santé et de la Recherche Médicale Unité No. 16, Biochimie des Protéines, Place de Verdun, F-59045 Lille, France.

Near the time of ovulation, the production of mucus increases, and the secretion becomes profuse, thin, and receptive to sperm. In the postovulatory phase, the mucus becomes thick and scanty, and impedes sperm penetration. Abnormalities in this secretion may be a factor in infertility. Furthermore, it has been shown that the receptors for various steroid hormones exist in cervical tissues (4). The rheological properties of the cervical mucus are largely determined by high-molecular-weight glycoproteins (5).

For the understanding of the physiological role of the glycoproteins of the cervical secretion during the menstrual cycle, knowledge of the carbohydrate structure is essential in relation to the biophysical properties and physiological function of the mucus at various phases of the ovulatory cycle (6, 7). For human cervical proteins, the structure of a few carbohydrate chains has been determined (8, 9). This kind of investigation is hampered by the fact that mucus of human origin is scarce and difficult to obtain. The bonnet monkey secretes large amounts of mucus, the rheological properties of which resemble those of human mucus (10); moreover, the simian menstrual cycle is similar to the human cycle. This makes the mucus of the bonnet monkey a suitable model system.

Fractionation of midcycle (periovulatory) cervical mucus provides two high-molecular-weight glycoproteins having a chemical composition characteristic of mucin-type structure (10–14). This paper describes further characterization of midcycle mucus glycoproteins, in particular the structure determination of their sialyloligosaccharides.

EXPERIMENTAL PROCEDURES AND RESULTS[1]

DISCUSSION

The isolation of the secretory glycoproteins present in cervical mucus of bonnet monkeys was readily accomplished by gel filtration on Bio-Gel P-200. The fractionation of the glycoprotein material by Sepharose 2B, followed by ion exchange chromatography on Ecteola-cellulose, yielded a main fraction 1-A (80%) (14). The minor fraction (1-B) differs significantly from 1-A in its sulfate and cystine content. In inhibition tests of hemagglutination against human anti-blood group A, anti-blood group B, and anti-blood group H, fraction 1-A showed a positive reaction. The feature of the combined occurrence of A, B, and H blood group activities could possibly have been avoided if typing of blood group activities of the monkeys had been carried out on the salivary mucin, before pooling the cervical mucus (37). Sialic acid in fraction 1-A was present as *N*-acetylneuraminic acid. This is similar to human cervical mucus (8) and different from bovine which contains the *N*-glycolylneuraminic acid (38). Cysteine was present only in a small amount as a component of glycoprotein 1-A. No cross-linking fraction containing cystine, as in the case of bovine cervical mucus (39, 40) was isolated.

Alkaline borohydride treatment of fraction 1-A resulted in a mixture of oligosaccharide-alditols which was subsequently fractionated on Bio-Gel P-4. The main fraction A-4 was further separated on hplc,[2] affording 7 subfractions (A-4-1 to A-4-7). The structures elucidated can be divided into two groups on the basis of the structure of the core, namely Galβ(1→3)[NeuAcα(2→6)]GalNAc-ol (A-4-1 to A-4-3) and Galβ(1→3)[GlcNAcβ(1→6)]GalNAc-ol (A-4-4 to A-4-7). The larger structures (A-4-5 to A-4-7) exhibit multiple blood group determinants.

The occurrence of a terminal nonreducing sequence GalNAcβ(1→4)[NeuAcα(2→3)]Gal is a feature already described in oligosaccharides from glycophorins with blood group Cad specificity (31) and from Tamm and Horsfall urinary glycoproteins with Sd^a activity (32, 41). Sd^a activity has also been detected in urinary mucin (42) and meconium (43). GalNAcβ(1→4)[NeuGlα(2→3)]Gal sequences have been observed in fish egg glycoproteins (44) and GalNAcβ(1→4)Gal in a cloned murine cytotoxic T lymphocyte line (45). In oligosaccharides (A-4-5 to A-4-7) of cervical mucins from the bonnet monkey, the Cad blood group determinant can occur together with an A, B, or H determinant.

[1] Portions of this paper (including "Experimental Procedures," "Results," Figs. 1–7, and Tables I–IV) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 85 M-2359, cite the authors, and include a check or money order for $9.20 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

[2] The abbreviations used are: hplc, high-performance liquid chromatography; glc, gas-liquid chromatography; GalNAc-ol, *N*-acetylgalactosaminitol; Fuc, fucose.

REFERENCES

1. Montgomery, R. (1970) in *The Glycoconjugates* (Pigman, W., and Horton, D., eds) Vol. IIB, pp. 627–709, Academic Press, New York
2. Iacobelli, S., Garcea, N., and Angeloni, C. (1971) *Fertil. Steril.* **22,** 727–734
3. Gibbons, R. A. (1959) *Biochem. J.* **73,** 209–217
4. Holt, J. A., Schumacher, G. F. B., Jacobson, H. I., and Swartz, D. P. (1979) *Fertil. Steril.* **32,** 170–176
5. Moghissi, K. S., and Blandau, R. J. (1972) *J. Reprod. Med.* **8,** 21–25
6. Van Kooij, R. J., Roelofs, H. J. M., Kathmann, G. A. M., and Kramer, M. F. (1980) *Fertil. Steril.* **34,** 226–233
7. Wolf, D. P., Sokoloski, J. E., and Litt, M. (1980) *Biochim. Biophys. Acta* **630,** 545–558
8. Yurewicz, E. C., and Moghissi, K. S. (1981) *J. Biol. Chem.* **256,** 11895–11904
9. Yurewicz, E. C., Matsuura, F., and Moghissi, K. S. (1982) *J. Biol. Chem.* **257,** 2314–2322
10. Ovadia, J., Mc Arthur, J. W., Kopito, L., and Ulfelder, H. (1971) *Biol. Reprod.* **5,** 127–145
11. Hatcher, V. B., Schwarzmann, G. O. H., Jeanloz, R. W., and Mc Arthur, J. W. (1977) *Biochemistry* **16,** 1518–1524
12. Nasir-Ud-Din, Jeanloz, R. W., Vercelotti, J. R., and Mc Arthur, J. W. (1981) *Biochim. Biophys. Acta* **678,** 483–496
13. Nasir-Ud-Din, Walker-Nasir, E., Mc Arthur, J. W., Nash, T. E., and Shaler, M. (1982) *Fertil. Steril.* **37,** 483–486
14. Nasir-Ud-Din, Jeanloz, R. W., Roussel, P., and Mc Arthur, J. W. (1983) *Biol. Reprod.* **28,** 1189–1199
15. Dubois, M., Gilles, K. A., Hamilton, J. K., Reber, P. A., and Smith, F. (1956) *Anal. Chem.* **28,** 350–356
16. Slayter, H. S., Lamblin, G., Le Treut, A., Galabert, A., Houdret, N., Degand, P., and Roussel, P. (1984) *Eur. J. Biochem.* **142,** 209–218
17. Havez, R., Roussel, P., Degand, P., and Lamblin, G. (1973) *Clin. Chim. Acta* **17,** 281–295
18. Lamblin, G., Boersma, A., Klein, A., Roussel, P., van Halbeek, H., and Vliegenthart, J. F. G. (1984) *J. Biol. Chem.* **259,** 9051–9058
19. Lamblin, G., Klein, A., Boersma, A., Nasir-Ud-Din, and Roussel, P. (1983) *Carbohydr. Res.* **118,** C1–C4
20. Vliegenthart, J. F. G., Van Halbeek, H., and Dorland, L. (1981) *Pure Appl. Chem.* **53,** 45–77
21. Gros, C., and Labouesse, B. (1969) *Eur. J. Biochem.* **7,** 463–470
22. Watkins, W. H., and Morgan, W. T. J. (1962) *Vox Sang.* **7,** 129–150
23. Matsumoto, I., and Osawa, T. (1970) *Arch. Biochem. Biophys.* **140,** 484–491
24. Iyer, R. N., and Carlson D. M. (1971) *Arch. Biochem. Biophys.* **142,** 101–105
25. van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Fiat, A.-M., and Jolles, P. (1980) *Biochim. Biophys. Acta* **623,** 295–300
26. Vliegenthart, J. F. G., Dorland, L., Van Halbeek, H., and Haverkamp, J. (1982) *Cell Biol. Monogr.* **10,** 127–172
27. Van Halbeek, H., Dorland, L., Haverkamp, J., Veldink, G. A., Vliegenthart, J. F. G., Fournet, B., Ricart, G., Montreuil, J., Gathmann, W. D., and Aminoff, D. (1981) *Eur. J. Biochem.* **118,** 487–495
28. Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Hull, W. E., Lamblin, G., Lhermitte, M., Boersma, A., and Roussel, P. (1982) *Eur. J. Biochem.* **127,** 7–20
29. Van Halbeek, H. (1984) *Biochem. Soc. Trans.* **12,** 601–605
30. Lamblin, G., Boersma, A., Lhermitte, M., Roussel, P., Mutsaers, J. H. G. M., van Halbeek, H., and Vliegenthart, J. F. G. (1984) *Eur. J. Biochem.* **143,** 227–236
31. Blanchard, D., Cartron, J.-P., Fournet, B., Montreuil, J., van Halbeek, H., and Vliegenthart, J. F. G. (1983) *J. Biol. Chem.* **258,** 7691–7695
32. Williams, J., Marshall, R. D., Van Halbeek, H., and Vliegenthart, J. F. G. (1984) *Carbohydr. Res.* **134,** 141–155
33. Akiyama, K., Simons, E. R., Bernasconi, P., Schmid, K., van Halbeek, H.,

Vliegenthart, J. F. G., Haupt, H., and Schwick, H. G. (1984) *J. Biol. Chem.* **259,** 7151–7154
34. Yates, A. D., Feeney, J., Donald, A. S. R., and Watkins, W. M. (1984) *Carbohydr. Res.* **130,** 251–260
35. Lemieux, R. U., Bock, K., Delbaere, L. T. J., Koto, S., and Rao, V. S. (1980) *Can. J. Chem.* **58,** 631–653
36. Dua, V. K., Dube, V. E., Li, Y. T., and Bush, C. A. (1985) *Glycoconjugate J.* **2,** 17–30
37. Socha, W. W., Mooz-Jankowski, J., Wiener, A. S., Risser, D. R., and Plonski, H. (1976) *Am. J. Phys. Anthropol.* **45,** 485–492
38. Gibbons, R. A. (1963) *Biochem. J.* **89,** 380–391
39. Gibbons, R. A., and Selwood, R. J. (1973) *The Biology of the Cervix* (Blandau, R. J., and Moghissi, K. S., eds) pp. 251–265, University of Chicago Press, Chicago
40. Bhushna Rao, K. S. P., and Masson, P. L. (1977) in *Mucus in Health and Disease* (Elstein, M., and Parke, D. V., eds) pp. 275–282, Plenum Press, New York
41. Donald, A. S. R., Yates, A. D., Soh, C. P. C., Morgan, W. T. J., and Watkins, W. M. (1984) *Biochem. Soc. Trans.* **12,** 596–599
42. Cartron, J.-P., Kornprobst, M., Lemonnier, M., Lambin, P., Piller, F., and Salmon, C. (1982) *Biochem. Biophys. Res. Commun.* **106,** 331–337
43. Morton, J. A., Pickles, M. M., and Terry, A. M. (1970) *Vox. Sang.* **19,** 472–482
44. Iwasaki, M., Inoue, S., Kitajima, K., Nomoto, H., and Inoue, Y. (1984) *Biochemistry* **23,** 305–310
45. Conzelman, A., and Kornfeld, S. (1984) *J. Biol. Chem.* **259,** 12528–12535

SUPPLEMENTARY MATERIAL TO

"STRUCTURE OF SIALYL-OLIGOSACCHARIDES ISOLATED FROM BONNET-MONKEY (MACACA RADIATA) CERVICAL MUCUS GLYCOPROTEINS EXHIBITING MULTIPLE BLOOD-GROUP ACTIVITES"

by Nasir-ud-din, Roger W. Jeanloz, Geneviève Lamblin, Philippe Roussel, Herman van Halbeek, Johanna H.G.M. Mutsaers and Johannes F.G. Vliegenthart

EXPERIMENTAL PROCEDURES

Materials : Bio-Gel P-200, Bio-Gel P-4, AG50W-X8 (100-200 mesh), AG1-X2 ion-exchange resins and Ecteola-cellulose (Cellex E) were purchased from Bio-Rad Laboratories, Sepharose 2B from Pharmacia Fine Chemicals Inc., agarose from MCI Biomedical Division of Marine Colloids Inc. and l'Industrie Biologique Francaise. Ulex europeus, Triticium vulgaris, Helix pomatia and Ricinus communis 120 lectins were purchased from Sigma.

Analytical methods. The hexose content of column eluates was measured with the phenol-sulfuric acid method (15), and the protein content by determining the absorbance at 280 nm. Amino acid analysis was performed with a Beckman Model 116 amino acid analyzer, after hydrolysis of the sample with 6 M HCl at 105°C for 20 h in an atmosphere of N_2, followed by dilution with water and lyophilization.

Polyacrylamide gel electrophoresis was performed according to (16), agarose gel electrophoresis in veronal buffer at pH 8.2 (ionic strength 0.1) according to (17). Gels were stained with periodate-Schiff reagent, amido black, Coomassie blue, toluidine blue or Sudan black (16,17).

Quantitative analysis of carbohydrates by gas-liquid chromatography (glc)[1] was performed according to (18). The samples were per-O-(trimethyl) silylated with Sylon HTP (Supelco) ; myo-inositol was used as internal standard. The glc analyses were performed with a Perkin-Elmer Model 900 gas chromatograph, on a column (180 x 0.3 cm) containing 3 % OV 17 on Chromosorb WHP 80-100 mesh (Supelco, Bellefonte, Pa).

Hplc was performed on a 5 µ Lichrosorb-NH_2 column. The elution was performed with a linear gradient of 4:1 to 1:1 acetonitrile-water containing 2.5 mM ammonium hydrogencarbonate for 70 min at room temperature and at a flow rate of 1 ml/min (19). Prior to ^{1}H-NMR spectroscopic analysis, samples were repeatedly exchanged in D_2O (99.96 mol% D, Aldrich) with intermediate lyophilization. The pD of the solution was adjusted to 7. ^{1}H-NMR spectroscopic analysis was performed on a Bruker WM-500 spectrometer (SON hf-NMR facility, Department of Biophysical Chemistry, Nijmegen University, the Netherlands) operating at 500 MHz in the Fourier transform mode at probe temperatures of 5, 10 or 27°C (20). Chemical shifts are given for neutral solutions at 27°C, relative to internal sodium 4,4-dimethyl-4-silapentane-1-sulfonate, but were actually measured by reference to internal acetone (δ2.225) with an accuracy of 0.002 ppm.

Isolation and purification of mucus glycoproteins. The crude pooled mucus obtained from 4 bonnet monkeys at the mid-cycle (12 to 15 days of the menstrual cycle) was partially solubilized in 50 mM sodium phosphate buffer, pH 7.0, containing 0.02 % NaN_3 by stirring at 4°C for 16 h. The cellular debris and other insoluble materials were removed by centrifugation (2 000 x g) ; the supernatant was dialyzed against distilled water and the retentate was lyophilized affording a mixture of (glyco)proteins. The residue was solubilized in 50 mM sodium phosphate buffer, pH 7.0, containing 0.02 % NaN_3 by stirring at 4°C overnight. The complex mixture of macromolecules was purified by filtration over Bio-Gel P-200 followed by fractionation on Sepharose-2B (11,12). The main fraction (100 mg) was subfractionated on an Ecteola-cellulose column (2.7 x 65 cm). The column was eluted with 0.1 M NaCl (200 ml) followed by a gradient of 0.1 M to 1 M NaCl in 10 mM HCl (Fig. 1).

End group analysis. The amino-terminal end-group was determined by dansylation of the major glycoprotein fraction (1-A, see below) (0.5 mg) and hydrolysis (21). The dansylated amino acids were separated and identified by thin layer chromatography on polyamide plates.

Blood-group activity and inhibition of lectin-induced hemagglutination. Hemagglutination inhibition tests were performed according to Watkins and Morgan (22) using anti-A, anti-B and anti-H immune sera (CRTS, Lille, France). The tests were performed by mixing one volume of mucin solution (25 µl) and one volume of antiserum representing 2 complete agglutinating doses. The mixtures were incubated for 15 min at room temperature and examined for agglutination.

Inhibitory activity towards lectins was assayed for glycoprotein fraction 1-A. Type O human red blood cells treated with neuraminidase (3.0 x 10^8 cells with 150 µl neuraminidase, 75 units from V. cholerae) were used with peanut agglutinin and Ricinus communis, whereas untreated cells were used with Ulex europeus, Ricinus communis, Helix pomatia and wheat germ agglutinins. The titration and inhibition assays were performed according to (23).

Alkaline borohydride treatment. The major fraction eluted from the Ecteola column was treated with 2 M $NaBH_4$ in 50 mM NaOH for 18 h at 45°C according to (24). Then the mixture was adjusted to pH 5.4 with 4 M acetic acid. For preparation of [^{3}H]-labeled oligosaccharides, β-elimination on a portion of glycoprotein (5 mg) was performed using NaB[^{3}H]$_4$ (5 mCi) under the conditions described above. The two reaction mixtures were combined and desalted on a column of AG50W-X8 (H^+, 100-200 mesh). Reduced oligosaccharides were separated into neutral (25 %) and acidic compounds (75 %) on a column (3.4 x 70 cm) of AG1-X2 (OAc^-, 200-400 mesh). The column was eluted with water, 0.5 M pyridine-acetic acid (pH 5.4) and then with 1 M acetic acid. The acidic oligosaccharides were further purified by filtration on Bio-Gel P-4 (200-400 mesh) in pyridine-acetic acid.

RESULTS

Purification and characterization of cervical mucus glycoproteins. The mucus glycoproteins obtained from bonnet monkey at menstrual mid-cycle were purified by gel filtration on Bio-Gel P-200 followed by fractionation on Sepharose 2B (12). The main fraction (85 %) after Sepharose 2B chromatography (denoted fraction 1) was analysed by polyacrylamide (15 %) gel electrophoresis. The glycoprotein did not enter the gel and no contaminating (glyco)proteins or (glyco)lipids were observed by staining with Coomassie blue, periodate-Schiff reagent or Sudan black. Ecteola-cellulose chromatography of fraction 1 afforded two subfractions, denoted 1-A (80 % by weight) and 1-B (Fig. 1). The carbohydrate and amino acid compositions and the sulfate contents of these fractions are reported in Table I.

Agarose electrophoresis of fraction 1-A showed a single band with periodate-Schiff reagent and toluidine blue. The carbohydrate chains consisted of fucose, galactose, N-acetylglucosamine, N-acetylgalactosamine and sialic acid. The glycoprotein contained a relatively high proportion of threonine, serine, alanine and glycine residues, while cysteine was detected in only small amounts (Table 1). Threonine and a trace of glycine were found as amino terminal end groups.

Blood-group activity and inhibition of lectin hemagglutination. Fraction 1-A had blood group A, B and H activities. Under our experimental conditions, inhibition of anti-A serum was obtained with 31 µg of fraction 1-A, inhibition of anti-B and anti-H immunesera was obtained with 250 µg of fraction 1-A.

The hemagglutination inhibition assays using glycoprotein fraction 1-A showed no activity with wheat germ and peanut agglutinins. A weak inhibitory activity using relatively high concentration of glycoprotein (100 µg/ml, 50 µl solution used) was observed with H. pomatia and U. europeus agglutinins. With R. communis low concentrations of glycoproteins demonstrated inhibition.

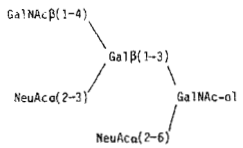
Preparation of sialyl oligosaccharide-alditols. Fraction 1-A (70 mg) was subjected to alkaline borohydride reductive cleavage (24) yielding a mixture of oligosaccharide-alditols. A decrease of serine and threonine and a corresponding increase of alanine and appearance of α-aminobutyric acid were noticed, concomitant with conversion of part of the GalNAc residues into GalNAc-ol (see Table I). The acidic oligosaccharides, eluted from the column of AG1-X2 by 0.5 M pyridine-acetic acid, pH 5.4, were separated on a Bio-Gel P-4 column into five fractions (Fig. 2). Fractions A-1 and A-2 represent mainly glycopeptide material as indicated by very low percentage of N-acetylgalactosaminitol (Table II) and the presence of hydroxylated amino acids (data not shown). The major fraction A-4 (12 mg) was further investigated. The molar carbohydrate composition of fraction A-4 is included in Table II. A-4 was fractionated by hplc into seven subfractions (A-4-1 to A-4-7) (Fig. 3).

Structure determination of sialyl oligosaccharides. The molar carbohydrate composition of A-4-1 to A-4-7 is given in Table III. Starting from these data, the complete primary structures of the oligosaccharides could be elucidated by employment of high-resolution (500-MHz) ^{1}H-NMR spectroscopy. The chemical shifts of the structural-reporter groups of these compounds are compiled in Table IV.

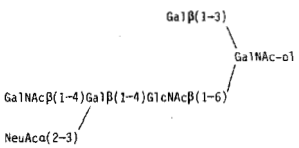
Compound A-4-1 could be identified as the disaccharide-alditol NeuAcα(2-6) GalNAc-ol (cf. Table III). Its NMR characteristics (Table IV) match exactly those described for this compound, obtained from other sources (20,25,26). Similarly, comparison of the NMR data of A-4-2 with those of the branched trisaccharide Gal β(1-3) [NeuAc α(2-6)] GalNAc-ol obtained from cow κ-casein (25,26) revealed that their structures are identical.

Compound A-4-3 was found to be the extension of A-4-2 with a Fuc residue in α(1-2) linkage to Gal[3] (for explanation of superscript notation, see Table IV, footnote a). This structure is essentially identical to the acidic tetrasaccharide Fucα(1-2)Galβ(1-3)[NeuGlα(2-6)]GalNAc-ol (27), except for the type of sialic acid in A-4-3 being NeuAc instead of NeuGl.

For compound A-4-4, the positions of the signals of H-2 (δ4.392) and H-5 (δ4.280) of GalNAc-ol in the ^{1}H-NMR spectrum point to the presence of a Galβ(1-3)[GlcNAcβ(1-6)]GalNAc-ol type of core (18,28,29). The H-1 signal of Gal[3] found at δ4.462 ($J_{1,2}$ = 7.8 Hz), together with the H-4 signal at δ3.907, indicate that A-4-4 contains this core-Gal residue in terminal position (20,28,30). The (1-6)-linked branch in A-4-4 could be identified to contain the Cad blood-group structure, by comparing the NMR data of A-4-4(Table IV) with those of the Cad pentasaccharide structure



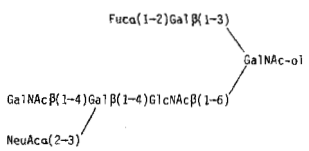
obtained from glycophorin-A of erythrocytes with blood-group Cad activity (31). The set of chemical shifts observed in A-4-4 for NeuAc H-3ax and H-3eq (δ1.93 and δ2.66, respectively) is known to be typical for the occurrence of a so-called internal sialic acid residue (31,32). Such a NeuAc residue is α(2-3)-linked to a Gal residue which also bears a sugar residue at C-4. The occurrence of NeuAc in α(2-3)-linkage to Gal is corroborated by the appearance of the Gal H-3 signal at 4.15. The substituent at C-4 of Gal is β-linked GalNAc (compare Table III), characterized by its H-1 doublet at 4.72 ($J_{1,2}$ = 8.1 Hz) and its N-acetyl signal at δ2.014. The former signal is partly hidden under the HOD-line when recording the NMR spectrum at room temperature, but could be visualized by lowering the sample temperature to 5-10°C (compare inset Fig. 4). The C-3,C-4 disubstituted Gal residue itself is β(1-4)-linked to the core $GlcNAc^6$. This can be derived from the chemical shifts of H-1, H-6 and the N-acetyl signal of $GlcNAc^6$, being 4.550, 4.000 and 2.062 respectively (28-33). Therefore, the structure of A-4-4 is the following :



The small but significant difference in chemical shift of the N-acetyl protons of β-GalNAc in A-4-4 compared to the Cad pentasaccharide (2.014 vs. 2.025 respectively) must be attributed to the difference in branch location of the Cad determinant. Analogously, the set of H-2, H-3 and H-4 resonances of the disubstituted β-Gal in this sequence appear to be slightly different for the two compounds (δ3.35/4.15/4.11 vs.δ3.42/4.16/4.09). Therefore, the combination of the chemical shifts of these structural-reporter groups seems to be suitable for branch localization of the Cad determinant in more complex oligosaccharides.

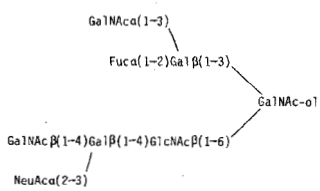
The application of this finding to the spectra of the remaining three compounds (A-4-5 to A-4-7) allows the recognition of the Cad sequence in the (1-6)-linked branch. By consequence, the structural differences between compounds A-4-4 to A-4-7 must be limited to the (1-3) branch. The carbohydrate composition of the oligosaccharides (Table III) suggests that A-4-5 may be an extension of A-4-4 with a Fuc residue whereas A-4-6 and A-4-7 have, in addition to Fuc, an extra GalNAc or Gal residue, respectively.

In the 500-MHz ^{1}H-NMR spectrum of A-4-5 (Fig. 4) the structural-reporter-group signals characteristic of Fucα(1-2) linked to Gal^3 of the core (δH-1 5.23, δH-5 4.27 and δCH_3 1.24) are readily recognized (28-30). The α(1-2)-linkage to Gal^3 is confirmed by the H-1 signal of Gal^3 at δ4.57 (compare A-4-3). Therefore, the structure of A-4-5 is :



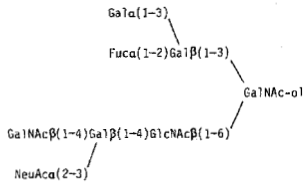
This implies that this oligosaccharide contains both the H and the Cad blood-group determinant. Comparison of the NMR spectra of A-4-4 and A-4-5 shows that the apparent extension of Gal^3 with Fuc in α(1-2) linkage causes some remarkable effects on the chemical shifts of other structural-reporter groups. The H-5 signal of GalNAc-ol was found to shift from δ4.280 to δ4.254 (Table IV). The N-acetyl signals of GalNAc-ol and $GlcNAc^6$ have shifted from δ2.067 and δ2.062, respectively, both to 2.053 in A-4-5. The effects are in line with those observed in the step from A-4-2 to A-4-3 as far as the GalNAc-ol signals are concerned (compare also ref. 30, in particular the step from compound 11 to 12 therein).

In A-4-6 (for spectrum see Fig. 5), identification of the (1-3)-linked branch is readily achieved by comparison of its 500-MHz ^{1}H-NMR parameters with those of the tetrasaccharide GalNAcβ(1-3)[Fucα(1-2)]Galβ(1-4)Glc (34) and with those of the acidic pentasaccharide-alditol GalNAcβ(1-3)[Fucα(1-2)]Galβ(1-3) [NeuGlα(2-6)]GalNAc-ol given in (27) (it should be noted that the spectrum of the latter compound (Fig. 6) contains a couple of corrected assignments in comparison with earlier 360-MHz data (cf. refs 34-36)). The simultaneous occurrence in the spectrum of A-4-6 of two H-1 signals of α-linked residues (δ5.378, Fuc^2 ; δ5.185, $GalNAc^3$) in conjunction with the presence of Fuc H-5 at δ4.322 and of $GalNAc^3$ H-2, H-4 and H-5 at δ4.251, δ4.020 and δ4.160 respectively, point to the occurrence of the A-determinant sequence in A-4-6. Therefore, the structure of A-4-6 is :



Conceiving A-4-6 as an extension of A-4-5 with an α-GalNAc, profound shift effects are observed on the signals of H-2 and H-5 of GalNAc-ol, H-1 and H-4 of Gal^3 and H-1, H-5 and CH_3 of Fuc (Table IV). These effects are essentially identical to those observed in the step from the acidic H^+ tetrasaccharide Fucα(1-2)Galβ(1-3)[NeuGlα(2-6)]GalNAc-ol to the acidic A^+ pentasaccharide (27,35).

The ^{1}H-NMR spectrum of A-4-7 (Fig. 7) shows, like that of A-4-6, two α-anomeric signals, but now at δ5.349 and δ5.259. In combination with the knowledge of the presence of an additional Gal residue as compared to A-4-5 (Table III), these are assigned to α-linked Fuc and Gal, respectively. Comparison with literature ^{1}H-NMR data on the blood-group B determinant (29,35) learns that A-4-7 possesses this sequence in its (1-3)-branch. So it can be concluded that A-4-7 possesses a blood-group Cad and blood-group B determinant jointly in one oligosaccharide as follows :



The apparent extension of A-4-5 (blood-group H determinant) with an α-linked Gal residue to A-4-7 (blood-group B determinant) causes profound shift effects on H-2 and H-5 of GalNAc-ol, H-1 and H-4 of Gal^3 and on H-1, H-5 and CH_3 of Fuc (Table IV) (cf. 35).

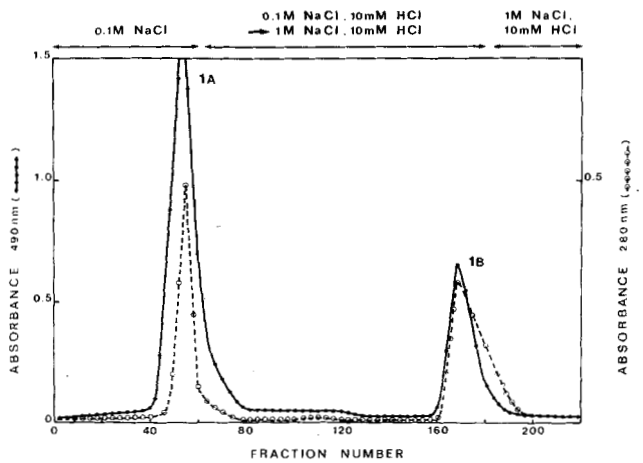


Fig.1 - Fractionation of Sepharose 2B purified glycoprotein 1 on an Ecteola-cellulose column (65 x 2.7 cm). The column was eluted with 0.1 M NaCl (200 ml), with a gradient of 0.1 M NaCl in 10 mM HCl/1 M NaCl in 10 mM HCl (450 ml) followed in 1 M NaCl in 10 mM HCl (200 ml). Fractions of 5 ml were collected and every third fraction was examined for the presence of hexoses and for protein. 100 mg of glycoprotein was applied to the column.

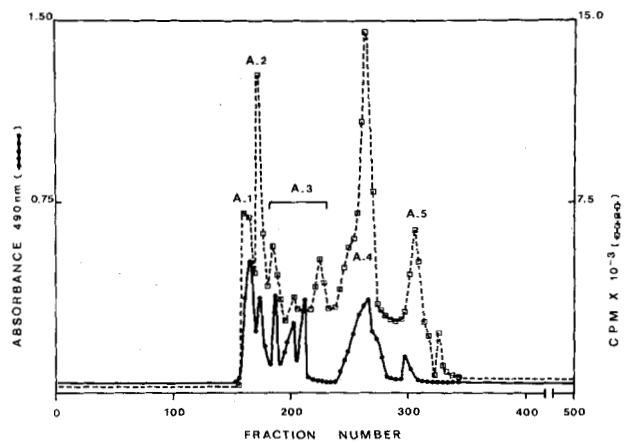


Fig. 2 - Fractionation of acidic oligosaccharides on a column (175 x 3.5 cm) of Bio-Gel P-4. The column was eluted with 50 mM pyridine acetic acid (pH 5.0, 800 ml), followed by a gradient of 0.1 to 1 M acetic acid. Fractions of 2.5 ml were collected and every third fraction was examined for hexoses and tritium. Five main fractions A1-A5 were pooled according to the bars (A-1, 161-168, 1.4 mg ; A-2, 170-184, 3 mg ; A-3, 191-236, 7 mg ; A-4, 240-294, 12 mg and A-5, 296-315, 2 mg).

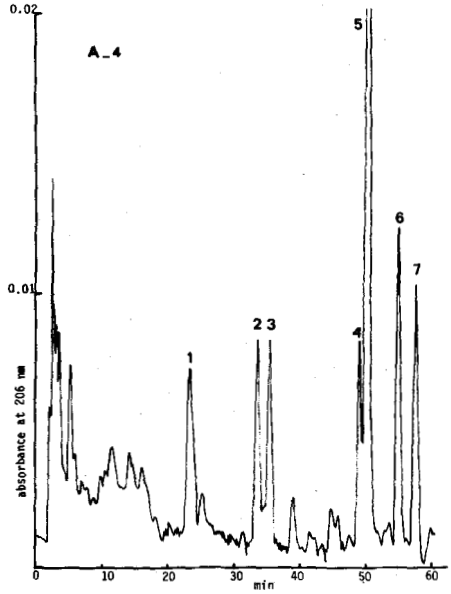


Fig. 3 - Hplc separation of sialyl-oligosaccharides A-4-1 to A-4-7 (peaks 1 to 7) of fraction A-4 on a Lichrosorb-NH_2 column eluted with a linear gradient of 4:1 to 1:1 (v/v) acetonitrile-water containing 2.5 mM ammonium hydrogen carbonate for 70 min at room temperature and at a flow rate of 1 ml/min (19). The relative yields of the recovered fractions were : A-4-1, 9.3 % ; A-4-2, 9.6 % ; A-4-3, 10.5 % ; A-4-4, 6.8 % ; A-4-5, 33.6 % ; A-4-6, 15.7 % ; A-4-7, 14.5 %.

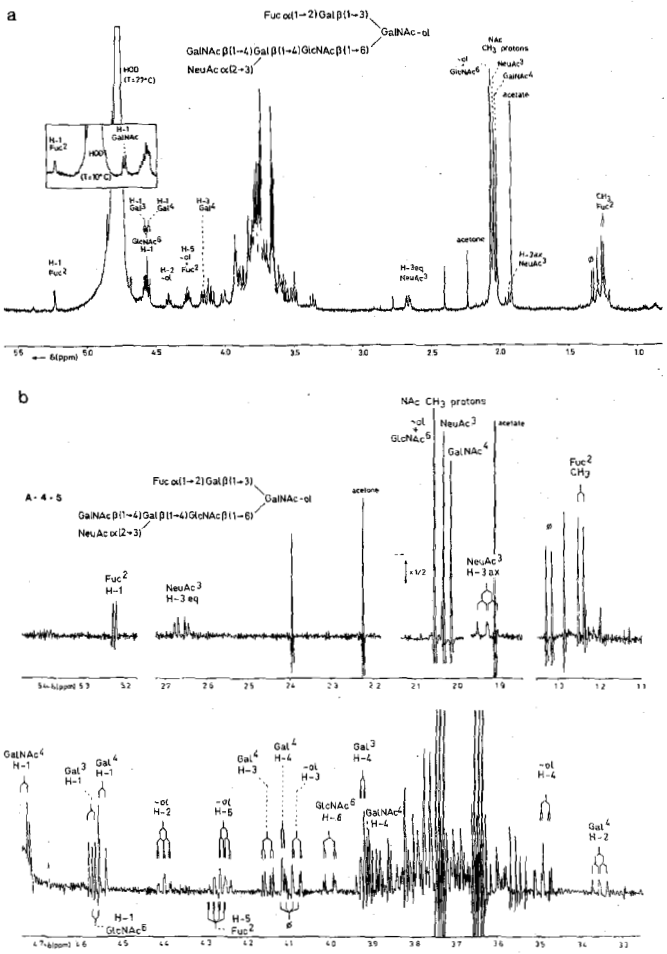


Fig. 4 - (a) 500-MHz ^{1}H-NMR spectrum of compound A-4-5 in D_2O at 27°C. The inset shows the H-1 signal of GalNAc which is visualized by lowering the temperature to 10°C. The signal marked by Ø stems from a frequently occurring, non-carbohydrate contaminant. (b) Resolution-enhanced 500-MHz ^{1}H-NMR spectrum of the oligosaccharide-alditol A-4-5 in D_2O at 27°C. The relative-intensity scale of the N-acetyl proton region of the spectrum differs from that of the other parts, as indicated.

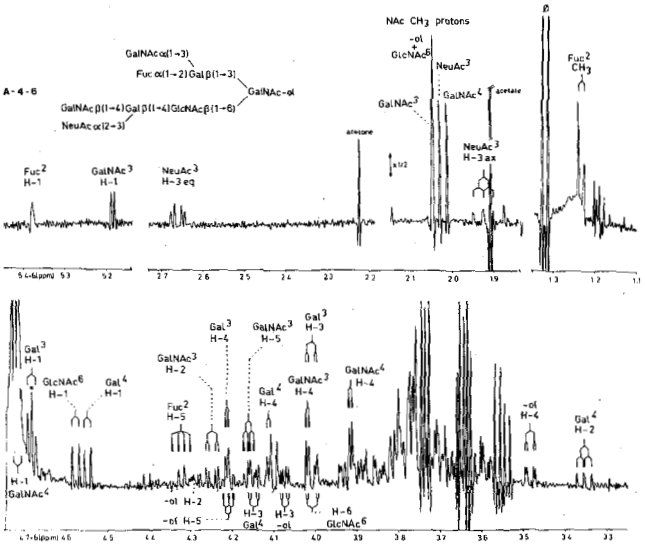


Fig. 5 - Resolution-enhanced 500-MHz ^{1}H-NMR spectrum of the oligosaccharide alditol A-4-6 in D_2O at 27°C. The relative-intensity scale of the N-acetyl proton region of the spectrum differs from that of the other parts, as indicated (* spinning side band).

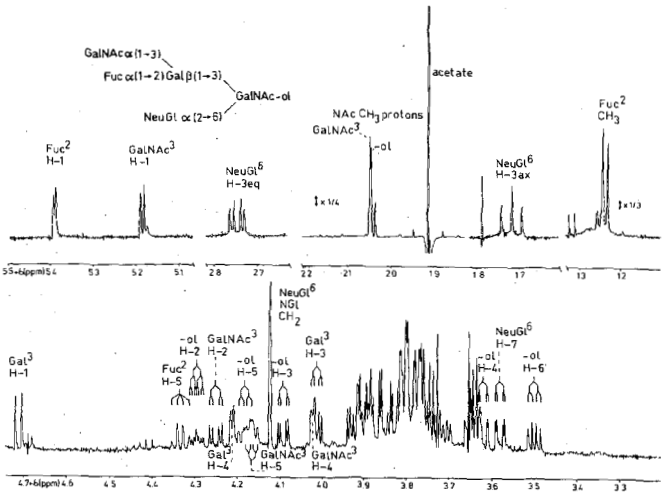


Fig. 6 - Resolution-enhanced 500-MHz ^{1}H-NMR spectrum of the acidic pentasaccharide-alditol with blood-group A-determinant supplied with the corrected (compare ref. 27) assignments. The spectrum is recorded in D_2O at 20°C.

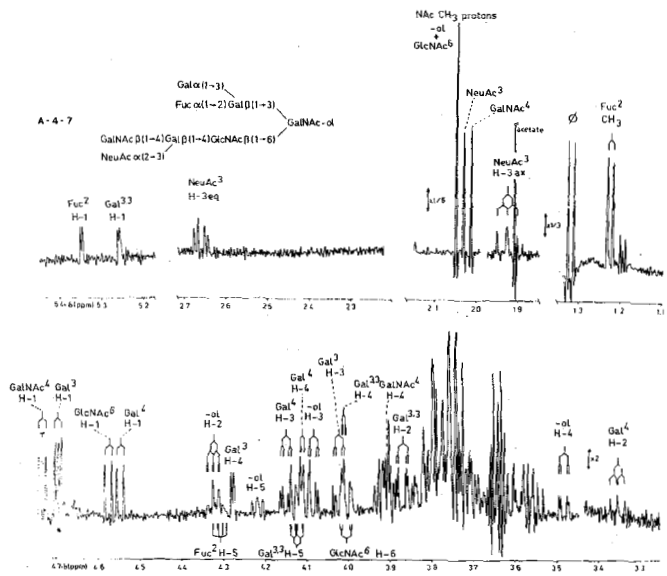


Fig. 7 - Resolution-enhanced 500-MHz ^{1}H-NMR spectrum of the oligosaccharide-alditol A-4-7 in D_2O at 27°C. The relative-intensity scale of the N-acetyl proton region of the spectrum differs from that of the other parts, as indicated (* spinning side band).

TABLE I : Carbohydrate and amino acid composition of cervical-mucus glycoprotein fractions 1-A and 1-B from bonnet monkey separated on Ecteola-cellulose, and amino acid composition of alkali-treated fraction 1-A glycoprotein.

	1-A		1-B	
	%	Molar ratio[a]	%	Molar ratio[a]
Fuc	7	0.9	6	1.1
Gal	24	3.0	15	2.6
GlcNAc	10	1.0	7	1.0
GalNAc	18	1.8	13	1.8
NeuAc	15	1.1	10	1.0
Total carbohydrate	74		51	
Sulfate	1		5	
Amino acid	14			n.d.[c]

	1-A	1-A after treatment with alkali	1-B
Asp[b]	44		48
Thr	228	-123	250
Ser	117	- 51	118
Glu	64		60
Pro	58		87
Gly	84		79
Ala	101	+ 41	99
Aba		+111	
Cys	14		3
Val	61		53
Ileu	59		44
Leu	62		62
Tyr	14		7
Phe	19		17
Lys	31		32
His	16		13
Arg	28		28

[a]Molar ratio relative to GlcNAc [c]not determined

[b]Residues per 1 000 residues

TABLE II : Carbohydrate composition of acidic oligosaccharide fractions separated by Bio-Gel P-4 filtration of the β-eliminative cleavage product of fraction 1-A glycoprotein from bonnet monkey cervical mucus.

Monosaccharide	Content in fraction							
	A-1	A-2	A-3		A-4		A-5	
	%	%	%	Molar[a] ratio	%	Molar[a] ratio	%	Molar[a] ratio
Fuc	6.8	6.9	6.4	0.11	6.3	0.59	4.5	0.25
Gal	18.9	14.0	17.4	4.50	23.4	1.99	20.4	1.00
GlcNAc	16.8	15.7	13.9	2.76	13.9	0.99	4.9	0.20
GalNAc	18.6	18.0	17.3	3.62	16.1	1.12	5.0	0.2
NeuAc	19.8	18.9	21.1	3.13	19.1	0.95	27.9	0.82
GalNAc-ol	1.8	1.0	4.8	1.00	14.5	1.00	24.6	1.00

[a]Molar ratio relative to GalNAc-ol.

TABLE III : Molar composition of sialyl oligosaccharides obtained by fractionation of cervical mucus glycan fraction of A-4 by hplc.

A-4 subfraction	Molar ratio[a] of monosaccharides					
	Fuc	Gal	GlcNAc	GalNAc	NeuAc	GalNAc-ol
A-4-1	-	-	-	-	0.8	1
A-4-2	-	1.0	-	-	0.8	1
A-4-3	0.7	0.9	-	-	0.7	1
A-4-4	-	2.0	0.8	1.2	0.8	1
A-4-5	0.7	2.1	0.9	1.1	1.0	1
A-4-6	0.6	1.6	0.8	1.4	0.9	1
A-4-7	0.5	3.4	1.0	1.2	1.0	1

[a]relative to GalNAc-ol

Table IV

^{1}H chemical shifts of structural reporter groups of constituent monosaccharides for the monosialyl oligosaccharide-alditols A-4-1 to A-4-7 derived from cervical mucus glycoprotein A of the bonnet monkey

Residue[a]	Reporter group	Chemical shift[b] in[c]						
		A-4-1	A-4-2	A-4-3	A-4-4	A-4-5	A-4-6	A-4-7
								
GalNAc-ol	H-2	4.246	4.378	4.381	4.392	4.399	4.294	4.324
	H-3	3.846	4.055	4.083	4.058	4.085	4.076	4.085
	H-4	3.411	3.534	3.540	3.485	3.482	3.482	3.482
	H-5	4.023	4.244	4.221	4.280	4.254	4.210	4.219
	H-6	3.846	n.d.	n.d.	3.937	3.93	3.925	3.923
	H-6	3.529	3.486	3.482	n.d.	n.d.	n.d.	n.d.
	NAc	2.055	2.047	2.038	2.067	2.053	2.049[d]	2.050
Gal[3]	H-1	-	4.474	4.583	4.462	4.571	4.676	4.697
	H-3	-	n.d.	n.d.	n.d.	n.d.	4.012	4.026
	H-4	-	3.894	3.916	3.907[e]	3.923	4.214	4.278
GlcNAc[6]	H-1	-	-	-	4.550	4.560	4.576	4.576
	H-6	-	-	-	4.000	4.003	4.010	4.002
	NAc	-	-	-	2.062	2.053	2.049[d]	2.050
Gal[4]	H-1	-	-	-	4.546	4.544	4.547	4.547
	H-2	-	-	-	3.352	3.354	3.354	3.355
	H-3	-	-	-	4.150	4.152	4.152	4.152
	H-4	-	-	-	4.112	4.112	4.113	4.113
NeuAc[6]	H-3ax	1.701	1.692	1.700	-	-	-	-
	H-3eq	2.727	2.726	2.730	-	-	-	-
	NAc	2.032	2.033	2.034	-	-	-	-
NeuAc[3]	H-3ax	-	-	-	1.925	1.927	1.925	1.925
	H-3eq	-	-	-	2.659	2.660	2.659	2.660
	NAc	-	-	-	2.032	2.031	2.032	2.031
GalNAc[4]	H-1	-	-	-	4.711[f]	4.722[g]	4.713[f]	4.738 (4.714[f])
	H-4	-	-	-	3.917[e]	3.914	3.916	3.916
	NAc	-	-	-	2.014	2.014	2.015	2.014
Fuc[2]	H-1	-	-	5.268	-	5.226	5.378	5.349
	H-5	-	-	4.269	-	4.272	4.322	4.317
	CH_3	-	-	1.235	-	1.244	1.232	1.227
GalNAc[3]	H-1	-	-	-	-	-	5.185	-
	H-2	-	-	-	-	-	4.251	-
	H-4	-	-	-	-	-	4.030	-
	H-5	-	-	-	-	-	4.160	-
	NAc	-	-	-	-	-	2.052[d]	-
Gal[3,3]	H-1	-	-	-	-	-	-	5.259
	H-2	-	-	-	-	-	-	3.872
	H-4	-	-	-	-	-	-	4.015
	H-5	-	-	-	-	-	-	4.119

[a] A superscript at the name of a sugar residue indicates to which position of the adjacent monosaccharide it is glycosidically linked. A second superscript is used to discriminate between (1-3)-linked Gal residues; it indicates the type of linkage of the neighbouring monosaccharide.

[b] Chemical shifts are in ppm relative to internal DSS (sodium 4,4-dimethyl-4-silapentane-1-sulfonate) but were actually measured to internal acetone (δ 2.225), in D_2O at 27° C.

[c] For complete structures of the compounds, see text. In the table-heading the structures are represented by short-hand symbolic notation; ◊ =GalNAc-ol, ■=β-Gal, ● =GlcNAc, △ =NeuAc[3], ○ =NeuAc[6], ◆ =β-GalNAc, □ =Fuc, ⊗ =α-Gal, ◆ =α-GalNAc

[d,e]Assignments may have to be interchanged.

[f] Values determined at 5°C. [g] Value determined at 10°C.