STRUCTURAL CHARACTERIZATION OF A NOVEL ACIDIC OLIGOSACCHARIDE UNIT DERIVED FROM COW COLOSTRUM κ-CASEIN

A 500 MHz ^{1}H-NMR study

H. van HALBEEK, L. DORLAND, J. F. G. VLIEGENTHART, A.-M. FIAT+ and P. JOLLÈS+

Department of Bio-Organic Chemistry, University of Utrecht, Croesestraat 79, 3522 AD Utrecht, The Netherlands and *+Laboratoire des Protéines, Université de Paris V, 45 rue des Saint-Pères, 75270 Paris Cedex 06, France*

Received 17 August 1981

1. Introduction

The glycan part of the cow milk glycoprotein κ-casein is constituted of only three different monosaccharides: galactose (Gal), *N*-acetylgalactosamine (GalNAc), and *N*-acetylneuraminic acid (NeuAc). The carbohydrate chains exhibit heterogeneity in their primary structures. So far, as products after alkaline borohydride reductive cleavage, the bisialo tetrasaccharide-alditol NeuAcα(2 → 3)Galβ(1 → 3)[NeuAcα-(2 → 6)]GalNAc-ol as well as two monosialo trisaccharide-alditols, NeuAcα(2 → 3)Galβ(1 → 3)GalNAc-ol and Galβ(1 → 3)[NeuAcα(2 → 6)]GalNAc-ol which can be conceived as partial structures of the tetrasaccharide-alditol, have been identified [1–4].

The carbohydrate moiety of κ-casein obtained from cow colostrum is much more complex than that from cow milk [1,5,6]. *N*-acetylglucosamine (GlcNAc) was found to be a typical, additional constituent of colostrum κ-casein. This observation indicates that κ-casein undergoes a maturation process leading to simpler carbohydrate chains. It is reasonable to propose that the maturation is related to the development of κ-casein function in the milk clotting process.

Relatively little is known on the location and type of linkage of the additional GlcNAc in the carbohydrate part of colostrum κ-casein. Recently, the structure of a monosialo tetrasaccharide-alditol, namely GlcNAcβ(1 → 3)Galβ(1 → 3)[NeuAcα(2 → 6)]GalNAc-ol, has been described [7]. Here, we report on the identification of an acidic pentasaccharide-alditol as a novel carbohydrate chain of colostrum κ-casein. The characterization of this compound has been done by 500 MHz ^{1}H-NMR spectroscopy, a method which allows the deduction of primary structures of carbohydrates derived from glycoproteins, even if low quantities are available in mixtures of closely related compounds [8–11].

2. Materials and methods

Cow colostrum κ-caseinoglycopeptide was prepared according to [6] from colostrum obtained 15 min after calving. The caseinoglycopeptide (30 mg) was treated with alkaline borohydride (0.05 M NaOH and 1.0 M $NaBH_4$) for 18 h at 50°C under nitrogen in the dark [12]. After desalting on Dowex 50W-X2 (H^+) with 2 mM formic acid as eluent and washing with methanol, the sugar moieties were isolated by filtration on Biogel P4 (250 × 0.9 cm) with water as eluent. Sugar-positive fractions, stained with the orcinol/sulphuric acid reagent, were further purified by descending paper chromatography (Whatman no. 1) for 18 h in the solvent system ethyl acetate/pyridine/water/acetic acid (5:5:3:1, by vol.). Sugars were detected with the periodate/benzidine reagent.

One of the fractions isolated from the Biogel P4 eluate seemed to contain a hitherto unknown (R_{Gal} = 0.52) oligosaccharide-alditol in a rather pure state. Sugar analysis of this fraction was performed by gas–liquid chromatography after methanolysis and trimethylsilylation as in [13].

A solution of the intact oligosaccharide-alditol was neutralized and exchanged several times in D_2O (99.96 atom% D, Aldrich) with intermediate lyophilization. NMR spectral analysis was carried out on a Bruker WM-500 spectrometer, operating in the Fourier trans-

Published by Elsevier/North-Holland Biomedical Press
00145793/81/0000–0000/$02.75

form mode with quadrature phase detection, at a probe temperature of 300 K. For solvent peak suppression a WEFT pulse sequence was used. Resolution enhancement of the spectrum was achieved by Lorentzian–Gaussian transformation [14]. Chemical shifts are given relative to sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS) (indirectly to acetone in D_2O: δ = 2.225 ppm).

3. Results and discussion

After alkaline borohydride treatment, filtration on Biogel P4 and paper chromatography, 0.9 mg purified oligosaccharide-alditol fraction (R_{Gal} = 0.52) was obtained from cow colostrum κ-caseinoglycopeptide. Sugar analysis of the sample afforded GalNAc-ol, Gal, GlcNAc and NeuAc in the ratio 1.0:2.0:0.9:1.2.

In order to elucidate the primary carbohydrate structure of the oligosaccharide-alditol(s) present, the sample was analyzed by 500 MHz ^{1}H-NMR spectroscopy. The resolution-enhanced 500 MHz ^{1}H-NMR spectrum is presented in fig.1. Relevant NMR data of the main component (**X**) of the oligosaccharide-alditol mixture are listed in table 1. The signals in the spectrum were assigned by using the ^{1}H-NMR data, acquired at 500 MHz, of the reference oligosaccharide-alditols **A**, **B**, **C** and **D**. The structures of the reference substances are presented in fig.2. Compounds **A** and **D** were isolated from bronchial mucin glycoproteins

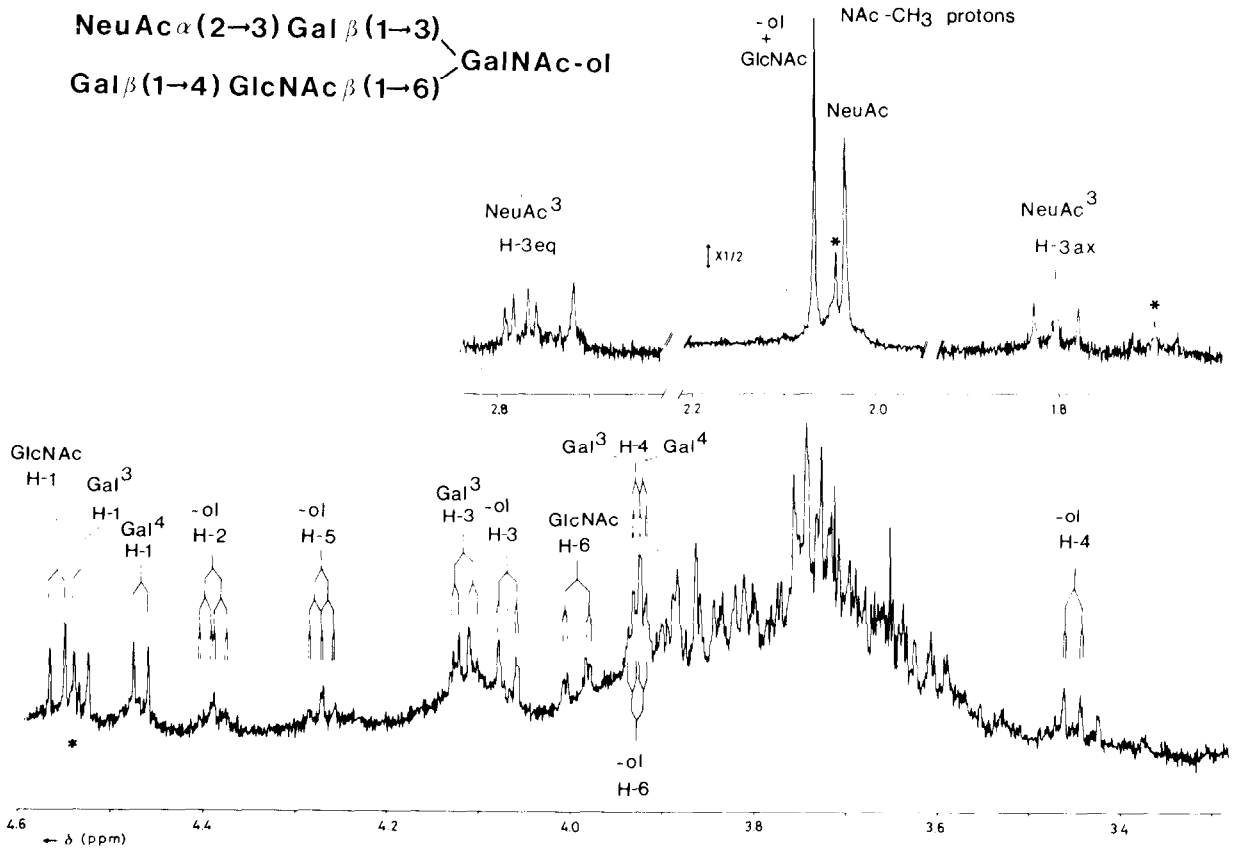


Fig.1. Resolution-enhanced 500 MHz ^{1}H-NMR spectrum (D_2O, 300 K, p$D \simeq 7$) of an acidic oligosaccharide-alditol fraction obtained from cow colostrum κ-caseinoglycopeptide. The sample mainly consists of the pentasaccharide-alditol, the structure of which is on top of the spectrum. It contains a small amount (~12%) of NeuAcα(2 → 3)Galβ(1 → 3)[NeuAcα(2 → 6)]GalNAc-ol as carbohydrate contaminant; this is particularly evident from the signals marked by asterisks (see text). The relative intensity scale of the *N*-acetyl proton region differs from that of the other parts of the spectrum as indicated. A superscript at the name of a sugar residue indicates the position of the C-atom in the adjacent monosaccharide, involved in the glycosidic linkage.

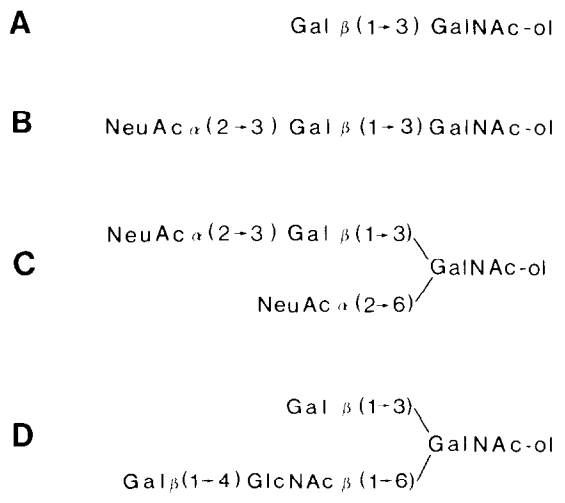


Fig.2. Structures of the mucin-type oligosaccharide-alditols, used as reference compounds.

stemming from patients suffering from cystic fibrosis [15]. Compounds **B** and **C** were obtained from cow milk κ-casein in [3].

The 360 MHz ^{1}H-NMR characteristics of **A** and **B** have been described in detail [3,16,17]. The NMR data of these compounds, refined by 500 MHz ^{1}H-NMR spectroscopy, are compiled in table 1. To summarize, the set of chemical shifts of the GalNAc-ol structural reporter groups, especially of H-2 (δ = 4.39 ppm) and H-5 (δ = 4.19 ppm), are indicative of a mono-substitution of GalNAc-ol at C-3 by Gal in β-linkage. In the case when Gal is substituted at C-3 by NeuAc in α-linkage (**B**), this can be inferred from the set of chemical shifts of the NeuAc H-3 atoms (δ H-3ax = 1.800 ppm; δ H-3eq = 2.774 ppm) in combination with those of Gal H-1 (δ = 4.547 ppm) and H-3 (δ = 4.122 ppm).

The 500 MHz ^{1}H-NMR spectrum of the acidic tetrasaccharide-alditol **C** has appeared in [8]. The NeuAc residue present in α(2 → 6) linkage to GalNAc-ol is characterized by the unique set of its H-3 chemical shifts (δ H-3ax = 1.692 ppm; δ H-3eq = 2.723 ppm). The chemical shifts of the GalNAc-ol H-5 (δ = 4.240 ppm) and H-6′ (δ = 3.475 ppm) give independent evidence of the location of this NeuAc at C-6 of the alditol [3,8].

The set of chemical shifts of H-2 (δ = 4.394 ppm) and H-5 (δ = 4.282 ppm) of GalNAc-ol in the neutral tetrasaccharide-alditol **D** indicate that this residue bears two substituents, namely at C-3 and C-6. The H-2 chemical shift points to the occurrence of the Galβ(1 → 3)GalNAc-ol moiety in **D**. This can be deduced from comparison with the data of other compounds possessing this core unit [3,8,15–17]. (Oligosaccharide-alditols containing the GlcNAcβ(1 → 3)-GalNAc-ol core moiety give rise to a significantly different chemical shift of GalNAc-ol H-2 [15].) The attachment of an *N*-acetyllactosamine unit in β-linkage to C-6 of GalNAc-ol comes to expression in shift increments for H-5 ($\Delta\delta \simeq 0.09$ ppm) and for H-6 ($\Delta\delta \simeq 0.24$ ppm) of GalNAc-ol, as compared to **A**. Owing to this substitution the doublet of doublets of the latter proton appears apart from the bulk of skeleton protons. In the case of mono-substitution of GalNAc-ol at C-3 (for example in **A** and **B**) the GalNAc-ol H-6 signal is located in the bulk ($\delta \simeq 3.7$ ppm). Attachment of NeuAc in α(2 → 6) linkage to GalNAc-ol affects only slightly the chemical shift of H-6 ($\delta \simeq 3.84$ ppm) [8], but leads to a considerable upfield shift of H-6′ (compound **C**, table 1). Therefore, the resonance positions of GalNAc-ol H-6 and H-6′ enable the rapid discrimination between substitution of the alditol at C-6 by β-linked GlcNAc or by α-linked NeuAc, respectively. The spectral features of the *N*-acetyllactosamine unit in **D** resemble very closely those of such units linked to the trimannosyl-*N*,*N*′-diacetylchitobiose core moiety of *N*-glycosidic carbohydrate chains derived from glycoproteins [8,10,18]. Assignments of signals in the spectrum of **D** having similar chemical shifts, in particular those of the anomeric protons of the terminal Gal residues (designated Gal^3 and Gal^4, respectively, after the type of glycosidic linkage wherein they are involved), and also the *N*-acetyl singlets of GalNAc-ol and of GlcNAc, are based upon data of reference compounds. Details are published in [15]. It should be emphasized that the chemical shift of H-1 of Gal^3 is slightly affected by the introduction of GlcNAc in β(1 → 6) linkage to GalNAc-ol ($\Delta\delta = -0.013$ ppm, when comparing **D** with **A**, table 1).

The 500 MHz ^{1}H-NMR spectrum of the purified colostrum oligosaccharide-alditol fraction (fig.1) shows that the sample consists of two components. This is particularly evident from the *N*-acetyl proton region ($2.0 < \delta < 2.1$ ppm) showing three signals with intensity ratios 1.6:0.1:1.0. The main component (**X**) contains GalNAc-ol substituted at C-3 as well as at C-6. This conclusion is based on the chemical shifts of H-2 (δ = 4.390 ppm) and of H-5 (δ = 4.272 ppm) of GalNAc-ol (see table 1). The former reflects that

Table 1
^{1}H chemical shifts of structural reporter groups of constituent monosaccharides for the pentasaccharide-alditol **X** obtained from cow colostrum κ-caseinoglycopeptide, together with those for the reference oligosaccharide-alditols **A** to **D**[a]

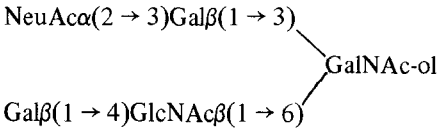
Residue	Reporter group	Chemical shift (ppm) in[b]				
		A	**B**	**C**	**D**	**X**
GalNAc-ol	H-2	4.395	4.390	4.378	4.394	4.390
	H-3	4.065	4.074	4.067	4.060	4.072
	H-4	3.507	3.498	3.524	3.465	3.456
	H-5	4.196	4.187	4.240	4.282	4.272
	H-6	3.69	3.68	3.84	3.931	3.927
	H-6′	3.628	3.65	3.475	3.7	3.7
	NAc	2.050	2.046	2.042	2.067	2.066
Gal^{3c}	H-1	4.478	4.547	4.541	4.465	4.534
	H-3	3.67	4.122	4.117	3.66	4.116
	H-4	3.901	3.931	3.927	3.900	3.922
GlcNAc	H-1	–	–	–	4.560	4.559
	H-6	–	–	–	3.998	3.993
	NAc	–	–	–	2.064	2.066
Gal^{4c}	H-1	–	–	–	4.470	4.470
	H-3	–	–	–	3.68	3.7
	H-4	–	–	–	3.925	3.931
NeuAc^{3c}	H-3ax	–	1.800	1.800	–	1.801
	H-3eq	–	2.774	2.774	–	2.774
	NAc	–	2.034	2.032	–	2.033
NeuAc^{6c}	H-3ax	–	–	1.692	–	–
	H-3eq	–	–	2.723	–	–
	NAc	–	–	2.032	–	–

[a] For the complete structure of compound **X** see fig.1; for the structures of **A**–**D** see fig.2
[b] All data were acquired at 500 MHz, for neutral solutions of the compounds in D_2O at 300 K
[c] Sugar residues possessing the same name are discriminated by a superscript indicating to which position of the adjacent monosaccharide they are linked

Gal is present in β(1 → 3) linkage to the alditol (see above). At C-6, GalNAc-ol bears a GlcNAc residue in β-linkage, which can primarily be deduced from the H-5 chemical shift. The latter conclusion is supported by the presence of the H-6 signal of GalNAc-ol at δ = 3.927 ppm, resonating outside of the bulk as could be proved by selective irradiation of GalNAc-ol H-5. The second Gal residue which is present in **X** according to the molar sugar composition of the fraction, is β(1 → 4) linked to GlcNAc as is evident from comparison of the chemical shifts of GlcNAc H-1 and H-6 of **X** with those of **D** (table 1).

One of the Gal residues in **X** is present in a terminal position (δ H-1 = 4.470 ppm); the other one is substituted by NeuAc in α(2 → 3) linkage because the spectrum shows NeuAc H-3 signals at δ = 1.801 ppm (H-3ax) and δ = 2.774 ppm (H-3eq) together with a Gal H-1 doublet at δ = 4.534 ppm and a Gal H-3 signal at δ = 4.116 ppm. The resonance positions of the latter signals match exactly the corresponding ones of the linear trisaccharide-alditol **B**, with the exception of H-1 (see table 1). The small upfield shift of this Gal H-1 in the step from **B** to **X** (Δδ = –0.013 ppm) relects the influence of the introduction of an

N-acetyllactosamine unit $\beta(1 \rightarrow 6)$ linked to GalNAc-ol upon the chemical environment of the core Gal H-1, as outlined above for the extension of **A** to **D**. Therefore, strong evidence exists for the location of the NeuAc residue in **X** in $\alpha(2 \rightarrow 3)$ linkage to Gal^3. Moreover, the chemical shifts of the H-1 and H-4 signals of the terminal Gal in **X** are identical with those of Gal^4 in **D** (table 1). This affords the following structure for the main compound **X**:



In comparison to **D**, the structural reporter skeleton protons of GalNAc-ol in **X** show small shift effects due to the attachment of NeuAc in $\alpha(2 \rightarrow 3)$ linkage to Gal^3 ($\Delta\delta$ H-2 = −0.004 ppm; $\Delta\delta$ H-3 = 0.012 ppm; $\Delta\delta$ H-4 = −0.009 ppm; $\Delta\delta$ H-5 = −0.010 ppm; $\Delta\delta$ H-6 = −0.004 ppm, see table 1). The complete accordance of these effects, as to sign and absolute value, with those observed going from **A** to **B**, lends additional support to the location of NeuAc in the upper branch of **X**. Thus, the possibility of the positionally isomeric, alternative structure for **X** possessing NeuAc in $\alpha(2 \rightarrow 3)$ linkage to Gal^4 instead of Gal^3, can be ruled out primarily on the basis of the chemical shifts of the Gal^3 and Gal^4 anomeric protons, and secondly on the basis of the set of chemical shifts of the GalNAc-ol skeleton protons.

In the *N*-acetyl proton region the singlet at δ = 2.033 ppm is assigned to NeuAc (compare with **B** and **C**), that at δ = 2.066 ppm to the coinciding resonances of GlcNAc and GalNAc-ol (compare with **D**). The occurrence of the low intensity singlet at δ = 2.047 ppm in conjunction with that of the NeuAc H-3ax signal at δ = 1.693 ppm and the Gal H-1 doublet at δ = 4.541 ppm, point to the presence of the bisialo tetrasaccharide-alditol **C** as minor constituent (~12%) of this sample, which is in agreement with the data of the sugar analysis.

4. Concluding remarks

A novel type of carbohydrate structure has been deduced for an acidic oligosaccharide chain of cow colostrum κ-casein, namely NeuAcα(2 → 3)Galβ(1 → 3)-[Galβ(1 → 4)GlcNAcβ(1 → 6)]GalNAc-ol. This report is the first account on the location of GlcNAc, being the typical constituent monosaccharide of cow colostrum glycoprotein, as forming part of an *N*-acetyllactosamine unit which is linked to C-6 of the authentically protein-bound GalNAc. This structural element has been demonstrated to occur also in ovine colostrum κ-casein, employing chemical and enzymic methods of structural analysis [19]. Furthermore, it is for the second time that an *O*-glycosidic, mucin-type structure containing both a sialic acid residue as well as an *N*-acetyllactosamine unit linked to the Galβ(1 → 3)-GalNAc-ol core, has been found [20].

The NMR data compiled in table 1 reveal that the chemical shifts of the structural reporter groups of the novel pentasaccharide-alditol (**X**) can be predicted from the reference data of compounds **A**, **B** and **D**. Taking the data of Galβ(1 → 3)GalNAc-ol (**A**) as starting point, the shift increments for attachment of NeuAc in $\alpha(2 \rightarrow 3)$ linkage to Gal (step from **A** to **B**) and the influences of extension of GalNAc-ol with a $\beta(1 \rightarrow 6)$ linked *N*-acetyllactosamine unit (step from **A** to **D**) lead to predicted values for the chemical shifts of the structural reporter groups of **X** which are in complete agreement with the experimental values. It has been well-established from a large set of structurally related carbohydrate chains derived from *N*-glycosidic glycoproteins that shift increments, which arise as a result of structural variations, can be handled independently of each other [21]. Now, it turns out that such shift increments can also be used to deduce structures of *O*-glycosidic chains, on the basis of the data of predecessors in complexity.

Acknowledgements

The authors thank Professor Dr C. Alais for providing the cow colostrum caseinoglycopeptide sample, Mrs J. Chevan for skilful technical assistance and Dr C. A. G. Haasnoot and Mr P. A. W. van Dael for their help with the recording of the 500 MHz ^{1}H-NMR spectrum. 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), the Institut National de la Santé et de la Recherche Médicale (Unité U-116) and the Centre National de la Recherche Scientifique (Equipe de Recherches no. 102).

References

[1] Fournet, B., Fiat, A.-M., Montreuil, J. and Jollès, P. (1975) Biochimie 57, 161–165.
[2] Fournet, B., Fiat, A.-M., Alais, C. and Jollès, P. (1979) Biochim. Biophys. Acta 576, 339–346.
[3] Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Fiat, A.-M. and Jollès, P. (1980) Biochim. Biophys. Acta 623, 295–300.
[4] Saito, T., Itoh, T. and Adachi, S. (1980) Agric. Biol. Chem. 44, 1023–1030.
[5] Alais, C. and Jollès, P. (1961) Biochim. Biophys. Acta 51, 315–322.
[6] Guérin, J., Alais, C., Jollès, J. and Jollès, P. (1974) Biochim. Biophys. Acta 351, 325–332.
[7] Saito, T., Itoh, T. and Adachi, S. (1981) Biochim. Biophys. Acta 673, 487–494.
[8] Vliegenthart, J. F. G., Van Halbeek, H. and Dorland, L. (1981) Pure Appl. Chem. 53, 45–77.
[9] Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Jouanneau, J. and Bourrillon, R. (1981) Biochem. Biophys. Res. Commun. 99, 886–892.
[10] Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Montreuil, J., Fournet, B. and Schmid, K. (1981) J. Biol. Chem. 256, 5588–5590.
[11] Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Spik, G., Chéron, A. and Montreuil, J. (1981) Biochim. Biophys. Acta 675, 293–296.
[12] Carlson, D. M. (1961) J. Biol. Chem. 241, 2984–2986.
[13] Clamp, J. R. (1974) Biochem. Soc. Symp. 40, 3–16.
[14] Ernst, R. R. (1966) Adv. Mag. Res. 2, 1–135.
[15] Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Hull, W. E., Lamblin, G., Lhermitte, M., Boersma, A. and Roussel, P. (1981) submitted.
[16] Vliegenthart, J. F. G., Van Halbeek, H. and Dorland, L. (1980) in: 27th IUPAC Int. Congress Pure Appl. Chem., Helsinki 1979 (Varmavuori, A. ed) pp. 253–262, Pergamon, Oxford.
[17] 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. 119, in press.
[18] Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Schmid, K., Montreuil, J., Fournet, B. and Hull, W. E. (1980) FEBS Lett. 114, 11–16.
[19] Soulier, S., Sarfati, R. S. and Szabó, L. (1980) Eur. J. Biochem. 108, 465–472.
[20] Lohmander, L. S., De Luca, S., Nilsson, B., Hascall, V. C., Caputo, C. B., Kimura, J. H. and Heinegård, D. (1980) J. Biol. Chem. 255, 6084–6091.
[21] Vliegenthart, J. F. G., Dorland, L. and Van Halbeek, H. (1982) Adv. Carbohydr. Chem. Biochem. 41, in press.