THE APPLICATION OF 500-MHz ^{1}H-NMR SPECTROSCOPY FOR THE STRUCTURE ELUCIDATION OF *N*-ACETYLLACTOSAMINE TYPE ASPARAGINE-BOUND CARBOHYDRATE CHAINS OF GLYCOPROTEINS

H. van HALBEEK, L. DORLAND, J. F. G. VLIEGENTHART, K. SCHMID*, J. MONTREUIL**, B. FOURNET** and W. E. HULL***

*Department of Bio-Organic Chemistry, University of Utrecht, Croesestraat 79, 3522 AD Utrecht, The Netherlands; *Department of Biochemistry, Boston University School of Medicine, Boston MA 02118, USA; **Laboratoire de Chimie Biologique, Université des Sciences et Techniques de Lille I, Villeneuve d'Ascq, France,* and ****Bruker Analytische Messtechnik, GmbH, D-7512 Rheinstetten-Fo., GFR*

Received 14 March 1980

1. Introduction

As we have shown previously, the type of branching in *N*-acetyllactosamine type carbohydrate chains of glycoproteins can be distinguished by means of 360-MHz ^{1}H-NMR spectroscopy [1–3]. The spectra obtained at 360 MHz are sufficiently different to characterize the various classes of antennary structures. However, upon increasing structural complexity a number of signals are not well resolved in these spectra. In particular the structural reporter groups [4] of the various *N*-acetyllactosamine units have very similar chemical shifts.

Recently a 500-MHz ^{1}H-NMR spectrometer was developed, which, together with an improved resolution enhancement routine based on [5], offered the possibility of establishing to which extent this new approach could lead to the unraveling of more spectral details. As will be shown now an identification of all individual lactosamine units can be achieved and inter alia an improvement is obtained in the assignment of the H-2's of the mannose residues.

2. Materials and methods

Asialo-glycopeptides from human plasma α_1-acid glycoprotein were prepared and designated according to Schmid et al. [6].

The glycopeptides were repeatedly exchanged in D_2O (99.96 atom %D, Aldrich) with intermediate lyophilization.

The 500-MHz ^{1}H-NMR spectra were recorded on a Bruker WM-500 spectrometer, operating in the Fourier Transform mode at a probe temperature of 300 K. Chemical shifts are given relative to sodium-2,2-dimethyl-2-silapentane-5-sulphonate (indirectly to acetone in D_2O:δ = 2.225 ppm). Resolution enhancement of the 500-MHz spectra was achieved by Lorentzian to Gaussian transformation according to Ernst [5].

3. Results and discussion

500-MHz ^{1}H-NMR spectra were recorded for bi-, tri- and tetra-antennary asparagine-bound carbohydrate chains of the *N*-acetyllactosamine type. The chemical shifts of the structural reporter groups, i.e. the anomeric protons, mannose H-2's and *N*-acetyl methyl protons are summarized in table 1. For the representatives of the classes A(bi), B(tri) and C(tetra) spectral regions for these groups of protons are given in figs.1 and 2.

As pointed out before [1] the substitution pattern of the mannose residues in each class is reflected in the chemical shifts of the H-1's and H-2's of these residues. The 500-MHz ^{1}H-NMR spectra show clearly that the signals of the H-2's of the mannose residues 3 and 4 in class B and C compounds do not coincide which is a large improvement in comparison to the 360-MHz spectra [1,2] (e.g. for class B δ H-2 of Man 3: 4.209 ppm; δ H-2 of Man 4: 4.218 ppm). It has to be noted that the difference between the

Table 1
^{1}H Chemical Shifts of Structural Reporter Groups of Constituent Monosaccharides for Bi-, Tri- and Tetra-antennary *N*-acetyllactosamine Type Glycopeptides

	Compound / Residue[a]	Bi-antenna (asn, lys) (class A)	Tri-antenna (asn, lys) (class B)	Tetra-antenna (asn, gly, thr) (class C)
H-1 of	1	5.094	5.092	5.053
	2	4.616	4.614	4.614
	3	4.765	4.755	4.757
	4	5.121	5.120	5.129
	4′	4.928	4.924	4.868
	5	4.582	4.570	4.573
	5′	4.582	4.580	4.596
	6	4.467	4.468	4.470
	6′	4.473	4.473	4.472
	7	–	4.545	4.547
	7′	–	–	4.553
	8	–	4.462	4.465
	8′	–	–	4.481
H-2 of	3	4.246	4.209	4.210
	4	4.190	4.218	4.224
	4′	4.109	4.108	4.092
NAc of	1	2.004	2.003	2.008
	2	2.079	2.078	2.078
	5	2.050	2.048	2.054[b]
	5′	2.046	2.045	2.041[b]
	7	–	2.075	2.079
	7′	–	–	2.042[b]

[a] For coding of monosaccharide residues and complete structures see fig.1A,B,C
[b] Assignments may be interchanged

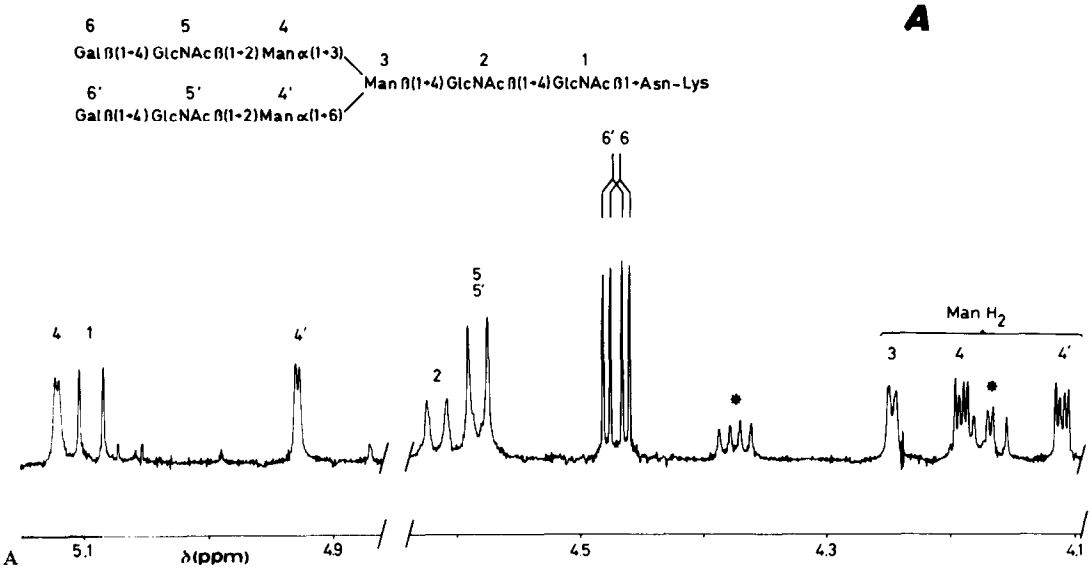

Fig.1A

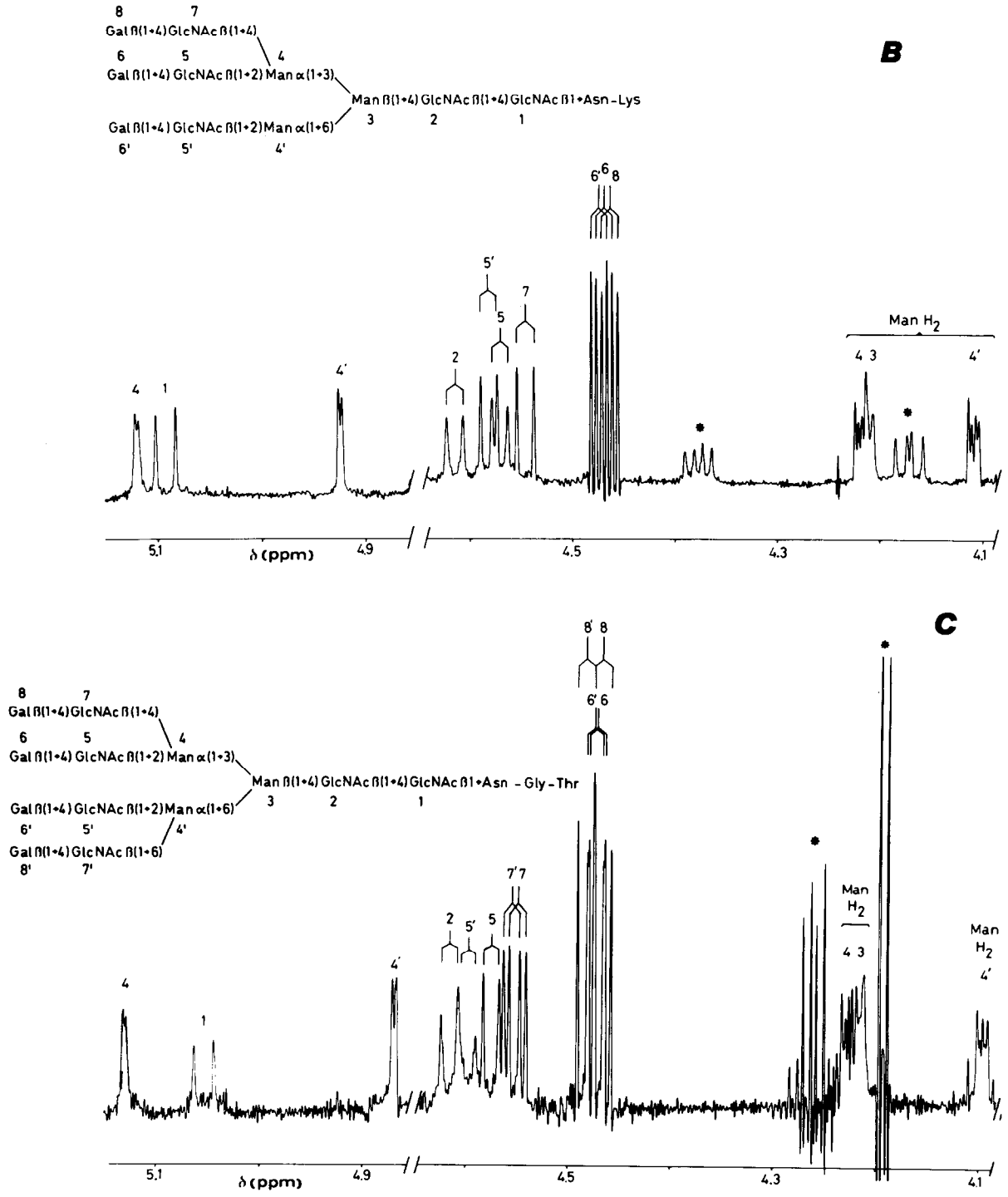

Fig.1. Structural reporter group regions of the resolution enhanced 500-MHz ^{1}H-NMR spectra of bi- (A), tri- (B) and tetra-antennary (C) asparagine-bound carbohydrate chains of the *N*-acetyllactosamine type. The numbers in the spectra refer to the corresponding residues in the structures. Resonance signals stemming from protons linked to C-α of an amino acid of the peptide moiety, are indicated by an asterisk. The HOD-line, as well as the H-1 signal of mannose 3, which at a sample temperature of 300 K is hidden under this line, are left out from the spectra.

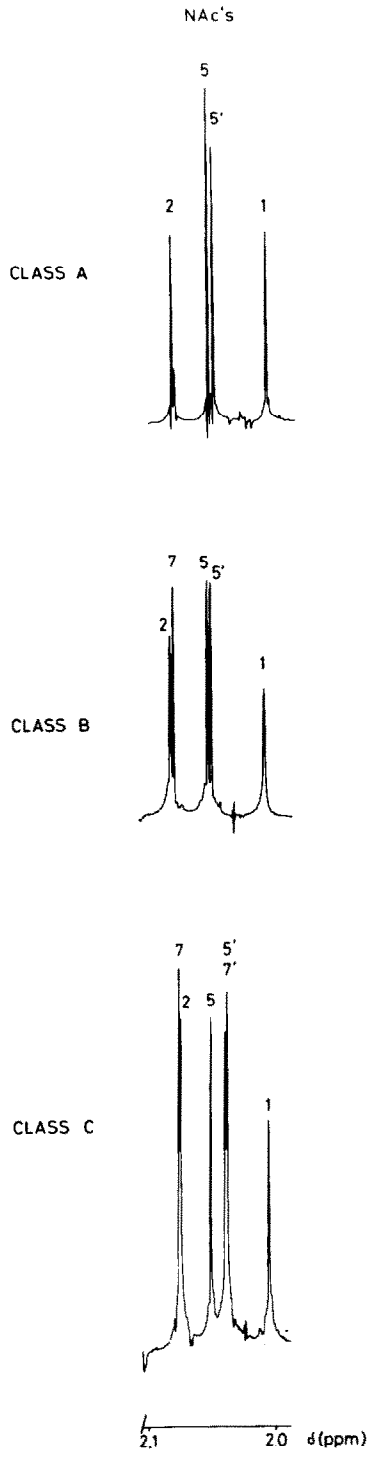

Fig.2. Characteristic patterns of the *N*-acetyl protons in the resolution enhanced 500-MHz ^{1}H-NMR spectra for the classes A, B and C carbohydrate chains. The numbers in the figure refer to the corresponding GlcNAc-residues.

coupling constants of vicinal axial-equatorial ($J_{1,2}$ = 0.8 Hz) and di-equatorial protons ($J_{1,2}$ = 1.6 Hz) causes distinct patterns for the resonance signals of the H-2 of β-(**3**) and α-linked (**4** and **4′**) mannose residues, respectively. These characteristic patterns provide us with an independent criterion to distinguish α- and β-linked mannose residues from each other.

In the 500-MHz spectrum of the bi-antennary glycopeptide (fig.1A) the signals of the anomeric protons of the galactose residues (**6** and **6′**) are interwoven but well resolved ($\Delta\delta$ = 0.006 ppm ~ 3 Hz). The lower field doublet at δ = 4.473 ppm is attributed to Gal **6′** and the higher field doublet at δ = 4.467 ppm to Gal **6**. These assignments could be made by comparison with the spectrum of the enzymically prepared monosialo analogue of this bi-antennary glycopeptide, having the sialic acid residue α(2→6) linked to Gal **6** (upper branch) [7]. Attachment of sialic acid to C-6 of Gal **6** gives rise to an upfield shift of the H-1 signal of this residue [8], whereas the chemical shift of the H-1 of the unsubstituted Gal **6′** remains unchanged compared with the asialo compound. The above mentioned interpretation is affirmed by the 500-MHz ^{1}H-NMR spectrum of the sialo-glycopeptide mixture from horse pancreatic ribonuclease [9]. The major component contains sialic acid α(2→6) linked to Gal **6** as well as α(2→3) to Gal **6′**. This study enabled the discovery of a minor component in the mixture bearing an α(2→3) linked sialic acid residue in the upper branch and an α(2→6) linked sialic acid residue in the lower branch (although the alternative that two different contaminants are involved cannot be fully excluded). This conclusion could be drawn since, besides the Gal H-1 doublets at δ = 4.549 ppm and δ = 4.444 ppm, additional doublets of equally low intensity are found at δ = 4.545 ppm and δ = 4.447 ppm, which were ascribed to the presence of the above-mentioned isomeric compound. This shows that for the set of H-1 signals of C-3 substituted galactose residues (at δ ~ 4.54 ppm) as well as for the set of H-1 signals of C-6 substituted galactose residues (δ ~ 4.44 ppm) the lowest field doublet belongs to the H-1 of the lower branch galactose (**6′**).

In the 500-MHz ^{1}H-NMR spectrum of the tri-antennary glycopeptide (fig.1B) an additional doublet at δ = 4.462 ppm is observed besides the doublets at δ = 4.473 ppm and δ = 4.468 ppm, belonging to the H-1's of Gal **6′** and **6**, respectively. Apparently, the chemical shifts of the latter protons are not altered by the presence of the third branch. By consequence

the highest field doublet must be ascribed to the H-1 of Gal 8.

Analogously in the 500-MHz spectrum of the tetra-antennary compound (fig.1C) a fourth doublet manifests itself at δ = 4.481 ppm, which is ascribed to the H-1 of Gal 8′. The chemical shift values of the H-1's of the other galactose residues (**6′**, **6** and **8**) remain essentially unchanged in comparison with those of the tri-antennary glycopeptide.

In accordance with the expected mobility of terminal residues, the line widths of the doublets of the galactose H-1's are very small, especially in comparison to those of the glucosamine H-1's.

In the spectrum of the bi-antennary compound it is remarkable that the anomeric proton signals of the peripheral *N*-acetyl glucosamine residues **5** and **5′** coincide even at 500 MHz. The doublet of H-1 of GlcNAc **2** shows considerably broader lines than the doublets of the other GlcNAc residues **1**, **5** and **5′**. This phenomenon is observed in all glycopeptides having an intact rigid core structure of trimannosyl-*N*,*N*′-diacetyl-chitobiose (unpublished results).

In the 500-MHz spectrum of the tri-antennary glycopeptide the doublet of H-1 of GlcNAc **7** at δ = 4.545 ppm shows significantly sharper lines than those of the H-1's of GlcNac **5** and **5′**. Attachment of the third branch to C-4 of Man **4** causes a shift decrement of the signal of the H-1 of GlcNAc **5** ($\Delta\delta$ = –0.012 ppm). This is a revision of the tentative assignment given before [2]. This extension results also in considerable line broadening of this doublet. The chemical shifts and line widths of the H-1 signals of GlcNAc **2** and **5′** are unaffected.

In the 500-MHz spectrum of the tetra-antennary glycopeptide the signals of the H-1's of GlcNAc **7** and **7′** appear as relatively sharp doublets in the high-field part of the GlcNAc anomeric region at δ = 4.553 ppm and δ = 4.547 ppm, respectively. The chemical shift values and the line widths of the H-1 signals of GlcNAc **2** and **5** are unaltered in comparison with those in the spectrum of the tri-antennary glycopeptide. However, the chemical shift and also the line width of the H-1 signal of GlcNAc **5′** are considerably affected by the presence of the additional *N*-acetyllactosamine unit bound to C-6 of Man **4′**. In this case a shift increment ($\Delta\delta$ = +0.016 ppm) is observed for the H-1 of GlcNAc **5′** upon substitution. The difference in line width between the H-1 signals of GlcNAc **5** and **5′** in the spectrum of the tetra-antennary compound suggests that an *N*-acetyllactosamine unit linked to C-6 of a mannose residue reduces the flexibility of the branch to which it is attached, much more profoundly than a similar unit bound to C-4.

Another characteristic feature of the 500-MHz ^{1}H-NMR spectra is found in the resonance patterns of the *N*-acetyl protons of the GlcNAc residues (fig.2A,B,C). For each of the three classes at 500 MHz all *N*-acetyl signals are well resolved.

For the bi-antennary glycopeptide all *N*-acetyl signals could unequivocally be assigned. The signals of GlcNAc **5** (δ = 2.050 ppm) and **5′** (δ = 2.046 ppm) could be distinguished on the basis of the spectrum of the monosialo upper branch analogue of the bi-antenna [7]: the signal at 2.046 ppm remains at the same position; that of GlcNAc **5** undergoes a shift increment to 2.069 ppm due to the attachment of sialic acid to C-6 of Gal **6**. The lines of the *N*-acetyl protons of the residues **1** and **2** are broader compared with those of **5** and **5′**, again caused by the lesser mobility of the rigid core of the glycopeptide in comparison with the peripheral residues.

In the spectrum of the tri-antennary glycopeptide the occurrence of the additional branch causes only slight differences in chemical shifts of the *N*-acetyl signals of residues **1**, **2**, **5** and **5′**. The additional signal at δ = 2.075 ppm belongs to GlcNAc **7**. An argument in favour of this assignment can be derived from the line width: the internal GlcNAc **2** may give rise to a broader *N*-acetyl signal (δ = 2.078 ppm).

The resonance positions of the protons of GlcNAc **1** are influenced by the structure of the peptide moiety, e.g. in the tetra-antennary-asn, gly, thr structure H-1 of GlcNAc **1** resonates at 5.053 ppm, in bi- and tri-antennary-asn, lys compounds at 5.093 ppm. A slight effect is observed for the *N*-acetyl signal of GlcNAc **1** (see table 1). The *N*-acetyl signals of GlcNAc **2** and **7** in the spectrum of the tetra-antennary compound can be assigned on the basis of the line widths. A tentative assignment of the remaining *N*-acetyl signals of the tetra-antenna is given in table 1. More reference compounds are necessary to arrive at a definite identification of these remaining *N*-acetyl signals. The excellent resolution of the *N*-acetyl signals is in any case suitable for the determination of the number of constituting GlcNAc residues.

4. Concluding remarks

The 500-MHz ^{1}H-NMR spectra of the complex

carbohydrate structures presented in this study, show a strikingly higher resolution than the corresponding 360-MHz spectra. This large improvement is due to the stronger magnetic field, whereas also a more sophisticated resolution enhancement computer program contributes. The observation that significant differences occur in the line widths of the anomeric signals turned out to be of great value in the spectral interpretation.

Resolution in combination with line widths allowed the assignment of all individual galactose and *N*-acetyl glucosamine H1's in bi-, tri- and tetra-antennary glycopeptides. A better insight is obtained in the patterns of the H-2's of the mannose residues. For the bi- and tri-antennary structures all *N*-acetyl signals could be identified.

This gain in spectral resolution is of the utmost importance for the characterization of compounds which can be conceived as extended structures of bi-, tri- and tetra-antennary glycopeptides. It can be expected that in particular the location and type of linkage of fucose and sialic acid residues can be elucidated. Furthermore structural investigation of reducing oligosaccharides representing partial structures of these complex compounds will be facilitated.

Acknowledgements

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), the Netherlands Foundation for Cancer Research (KWF, grant UUKC-OC 79-13), the Centre National de la Recherche Scientifique (Laboratoire Associé No. 217: Biologie physicochimique et moléculaire des glucides libres et conjugués) and the National Institute of General Medical Sciences (GM-10374).

References

[1] Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., Fournet, B., Strecker, G., Spik, G., Montreuil, J., Schmid, K. and Binette, J. P. (1978) FEBS Lett. 89, 149–152.

[2] Fournet, B., Montreuil, J., Strecker, G., Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., Binette, J. P. and Schmid, K. (1978) Biochemistry 17, 5206–5214.

[3] Montreuil, J. and Vliegenthart, J. F. G. (1979) in: Glycoconjugate Research, Proc. IVth Int. Symp. Glycoconjugates (Gregory, J. D. and Jeanloz, R. W. eds) vol. I, pp. 35–78, Academic Press, New York.

[4] Vliegenthart, J. F. G., van Halbeek, H. and Dorland, L. (1980) in: 27th International Congress of Pure and Applied Chemistry, Helsinki 1979 (Varmavuori, A. ed) in press, Pergamon, Oxford.

[5] Ernst, R. R. (1966) in: Advances in Magnetic Resonance (Waugh, J. S. ed) vol. 2, pp. 1–135, Academic Press, New York.

[6] Schmid, K., Nimberg, R. B., Kimura, A., Yamaguchi, H. and Binette, J. P. (1977) Biochim. Biophys. Acta 492, 291–302.

[7] Van den Eijnden, D. H., Joziasse, D. H., Dorland, L., van Halbeek, H., Vliegenthart, J. F. G. and Schmid, K. (1980) Biochem. Biophys. Res. Commun. 92, 839–845.

[8] Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., Strecker, G., Michalski, J.-C., Fournet, B., Spik, G. and Montreuil, J. (1978) Eur. J. Biochem. 87, 323–329.

[9] Schut, B. L., Dorland, L., Haverkamp, J., Vliegenthart, J. F. G. and Fournet, B. (1978) Biochem. Biophys. Res. Commun. 82, 1223–1228.