SEMINAR

APPLICATION OF NMR TO THE STRUCTURAL ELUCIDATION OF COMPLEX CARBOHYDRATES*

L. DORLAND and J. F. G. VLIEGENTHART

Department Bio-Organic Chemistry, State University,
Croesestraat 79, 3522 AD Utrecht,
The Netherlands

The structural analysis of the carbohydrate chains of glycoproteins is usually carried out by a combination of chemical and enzymic methods. Some frequently used techniques are: periodate oxidation, Smith degradation, permethylation analysis, acetolysis, alkaline degradation (β-elimination), and sequential degradation with glycosidases. Despite the high degree of sophistication reached by these methods, it is evident that still quite a number of uncertainties remain even with regard to the monosaccharide composition. For example, it is sometimes difficult to estimate accurately the number of mannose and *N*-acetyl-hexosamine residues.

During the last few years high resolution ^{1}H-NMR spectroscopy has obtained a firm position among the techniques available for structural studies of biomolecules. In the structure elucidation of carbohydrates

*Lecture delivered at Les Houches by L. Dorland.

R. Balian et al., eds.
Les Houches, Session XXXIII, 1979 – Membranes et Communication Intercellulaire / Membranes and Intercellular Communication
©North-Holland Publishing Company, 1981

high resolution ^{1}H-NMR spectra give valuable information about qualitative and quantitative aspects of the carbohydrate structure. For relatively large compounds the spectra are not completely interpretable. However, already a partial interpretation of the spectra can furnish relevant structural information. The structural reporter groups which are useful for this aim are those that comprise the signals of (a) the anomeric protons; (b) special non-anomeric protons, resonating outside of the poorly resolved bulk, e.g. the H-2 protons of mannose- and the H-3 protons of sialic acid residues; and (c) protons of substituents like *N*-acetyl, *N*-glycolyl and *O*-acetyl groups.

In the following, a few examples will be presented of what in principle can be done with 360 MHz ^{1}H-NMR spectroscopy.

Figure 1 gives the structure for the carbohydrate chain of human serotransferrin. Jamieson [1] proposed on the basis of periodate oxidation, partial acid hydrolysis, methylation analysis, and enzymic degradations the upper formula of fig. 1, whereas Spik [2] arrived at a more symmetric structure using essentially the same techniques. The principle differences are the following: Jamieson found 4 mannose residues and suggested GlcNAc→Asn to be directly attached to mannose; Spik found only 3 mannose residues and she proposed the occurrence of a *N*-acetylchitobiose unit linked to asparagine.

In fig. 2 the overall 360 MHz ^{1}H-NMR spectrum of the asialo serotransferrin glycopeptide [3] is given. Simple integration of the

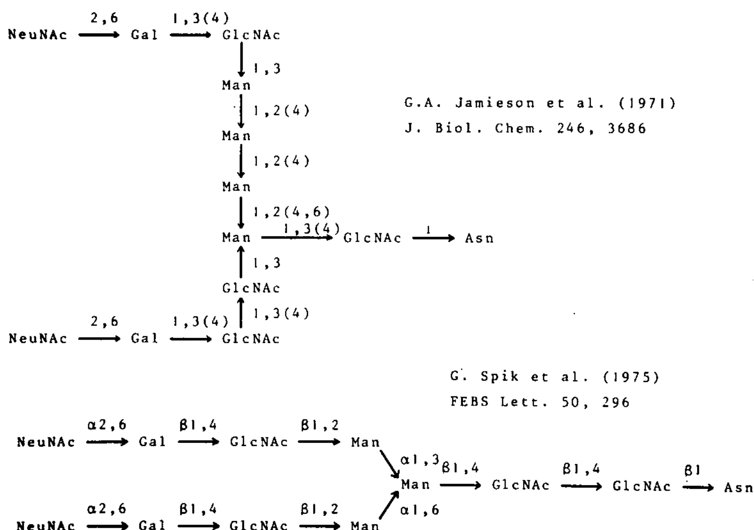

Fig. 1. Proposed structures for the carbohydrate chain of human serotransferrin.

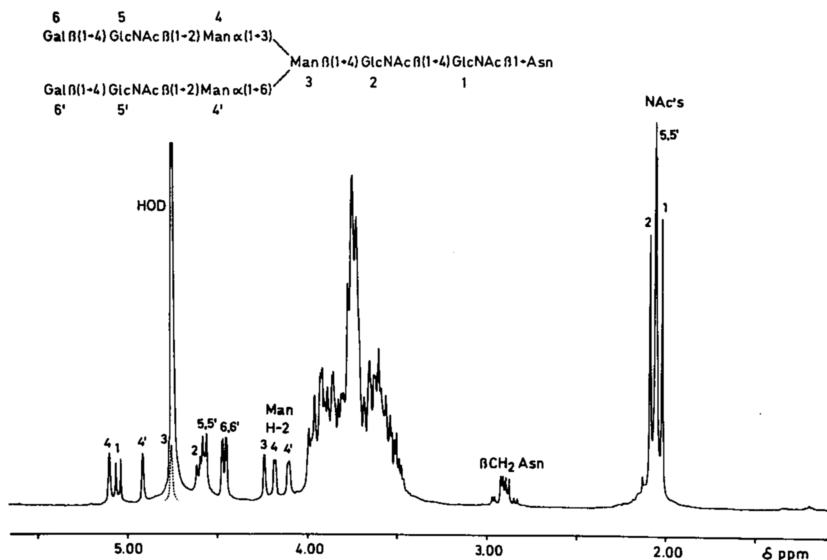

Fig. 2. 360 MHz ^{1}H-NMR spectrum of the asialo-serotransferrin glycopeptide in D_2O.

mannose H-2 and H-1 protons shows immediately that only 3 mannose residues are present in this structure, which excludes the structure given by Jamieson. In fact the proposal of Spik is correct, but before considering this in more detail, I want to mention that this report is limited to complex glyco-chains, coupled via an *N*-glycosidic linkage to asparagine in glycoproteins. According to the carbohydrate composition, these can be divided into two groups having in common a pentasaccharide core, which consists of a *N*-acetylchitobiose–Asn moiety, linked to a mannotriose branching unit. In the first family of the Asn glycans the pentasaccharide is substituted by one or more *N*-acetyl lactosamine units. These structures can be further extended by sialic acid and/or fucose residues in terminal positions. In the second family the pentasaccharide bears oligomannose chains. The total number of mannose residues in these type of compounds can vary considerably. Both groups of structures are presented in fig. 3.

In figs. 4–6 a few structures of the lactosamine type are presented to define the compounds which have been investigated. Figure 4 presents the extension of the pentasaccharide core with two *N*-acetyl lactosamine units. This is defined as the biantennary structure [4]. Figure 5 gives a further extension with a *N*-acetyl lactosamine moiety affording the triantennary structure, and fig. 6 shows the tetraantennary structure [4]. Compounds also occur which represent partial structures of the afore-

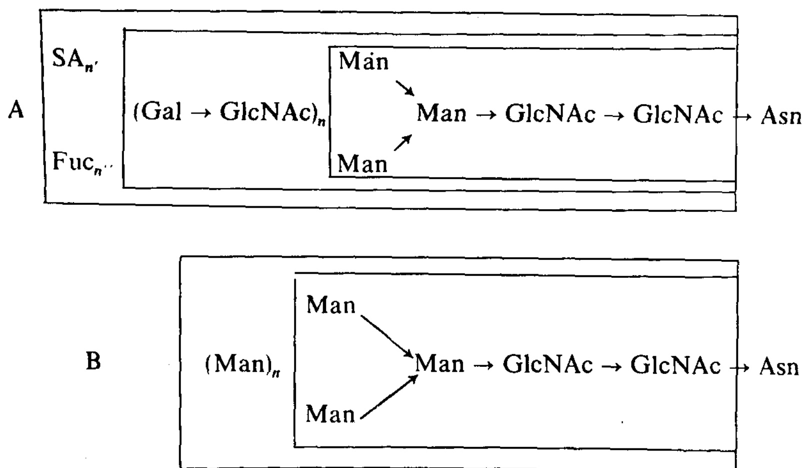

Fig. 3. Two types of *N*-glycosidic carbohydrate structures occurring in glycoproteins.

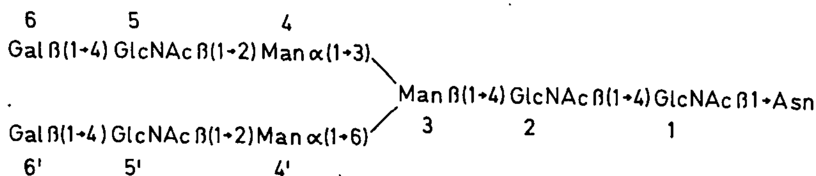

Fig. 4. Biantennary structure.

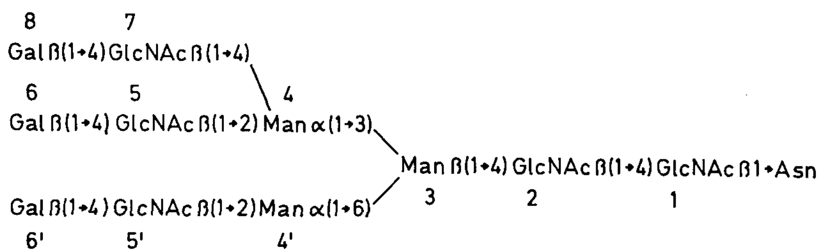

Fig. 5. Triantennary structure.

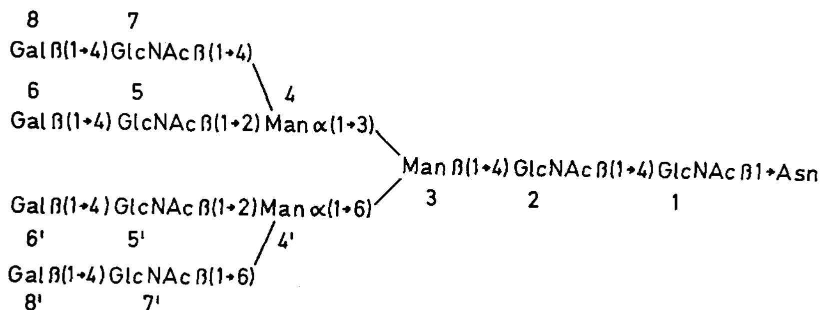

Fig. 6. Tetraantennary structure.

mentioned classes of carbohydrate chains. In particular from glycoproteins which have a large heterogeneity, these partial structures can be obtained.

The 360 MHz ^{1}H-NMR spectrum of the general structural element GlcNAcβ→1 Asn could completely be interpreted [5]. The coupling constant $J_{1,2}=9.8$ Hz is indicative of a β-glycosidic bond of the GlcNAc residue. The $^4C_1$ (D) chair conformation of the pyranosyl ring of the GlcNAc residue could be deduced from the vicinal coupling constants $J_{HH'}$ using an adapted Karplus equation. Another frequently occurring structural element in glycoproteins is Fucα(1→6)GlcNAcβ1 →Asn. The 360 MHz ^{1}H-NMR spectrum of this compound is also completely interpretable. Characteristic for the fucose residue are the resonances of the anomeric proton ($\delta=4.90$ ppm; $J_{1,2}=3.7$ Hz) and the protons of the CH_3 group ($\delta=1.21$ ppm; $J_{5,6}=6.6$ Hz). The coupling constant of 3.7 Hz is indicative of an α-glycosidic bond of the fucose residue. The ring conformations of both sugar residues were determined: for GlcNAc the $^4C_1$ (D) chair and for Fuc the $^1C_4$ (L) chair [5]. The attachment of fucose to position 6 of GlcNAc gives rise to changes in the chemical shifts for H-4, 5, and 6 of GlcNAc when compared with the data of GlcNAcβ1→Asn. Also the value of the geminal coupling constant $J_{6,6'}$ of GlcNAc is changed ($-12.7 \rightarrow -11.4$ Hz).

As soon as the number of monosaccharides increases the pattern of the non-anomeric protons becomes far too complex to be analysed in full detail.

An example of a complex Asn-glycopeptide is the asialo-glycopeptide obtained from human serotransferrin, which spectrum is given in fig. 2. The resonances of this glycopeptide can be divided into signals from the following groups of protons:

(a) Anomeric protons (9); at this temperature, one is hidden under the HOD line. This signal can be visualized by recording the spectrum at another temperature.
(b) The H-2 protons of mannose residues.
(c) The remaining non-anomeric protons.
(d) β-CH_2 protons of asparagine.
(e) CH_3 protons of the *N*-acetyl groups.

It is impossible to interpret the bulk of the non-anomeric protons. However, many structural details of this compound are reflected in the chemical shifts of the anomeric protons and the H-2 protons of the mannose residues. Integration showed that 9 anomeric protons are present, which is in accordance with the proposed number of monosaccharide units. An extra check for the number of amino sugar residues

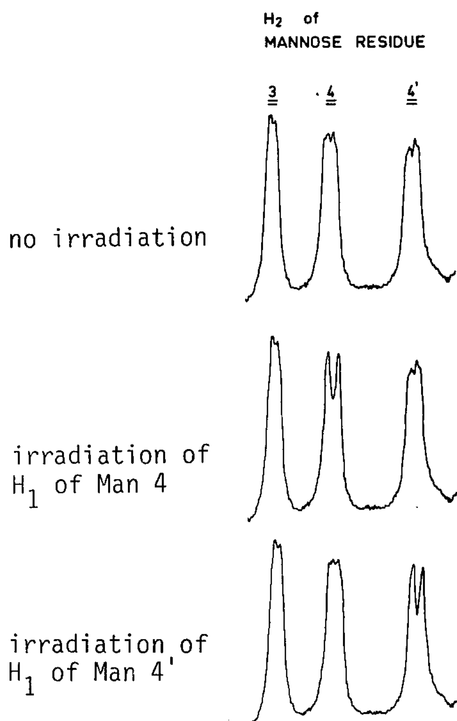

Fig. 7. Assignment of the mannose H-2 protons by selective irradiation.

can be obtained from integration of the *N*-acetyl signals. We arrived at this interpretation on the basis of a wide variety of partial structures, each containing a characteristic part of this molecule. The H-2 protons of the mannose residues could be assigned by selective irradiation of the H-1 protons of the mannose residues (fig. 7). From the NMR spectrum of this glycopeptide and the spectra of the various reference compounds some conclusions can be drawn:

(1) The chemical shifts of the various anomeric protons in the intact glycopeptide and the partial structures thereof occur at characteristic positions.

(2) The primary structure and the type of the glycosidic linkages are reflected by the chemical shifts and the coupling constants of the anomeric protons of the various monomers. The total NMR spectrum can be used as a fingerprint.

(3) The mannotriose branching core, surrounded by GlcNAc residues can be recognized on the basis of the pattern of the mannose H-2 proton resonances.

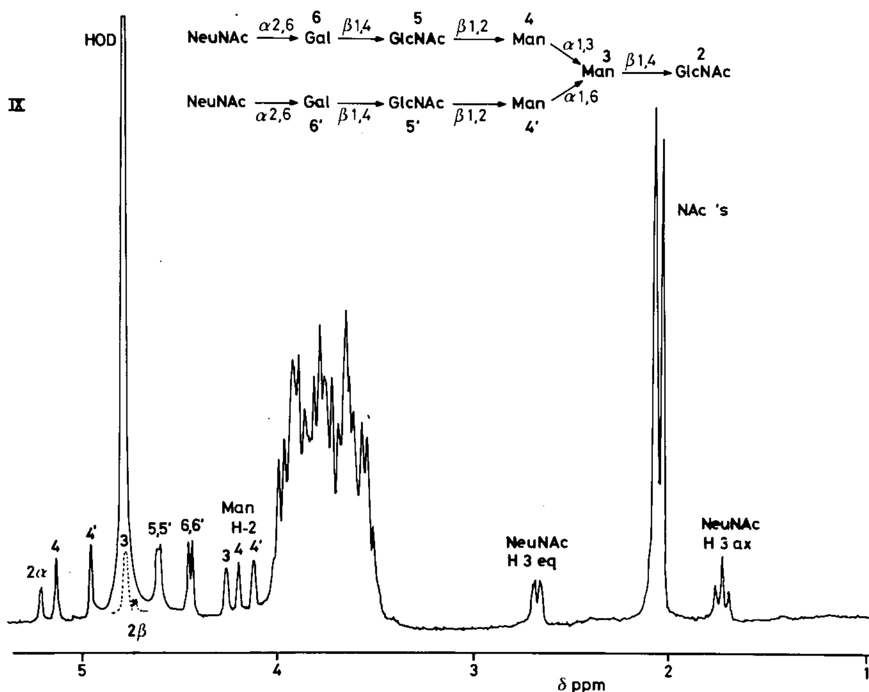

Fig. 8. 360 MHz ^{1}H-NMR spectrum of sialo-oligosaccharide IX in D_2O.

The occurrence of sialic acid at terminal positions of the biantennary structure has some remarkable influences on the spectrum, depending on the type of glycosidic linkage. First $\alpha(2\rightarrow6)$ linked sialic acid, as present in the oligosaccharide IX, isolated from urine of sialidosis patients [6]. Figure 8 gives the 360 MHz ^{1}H-NMR spectrum of this sialo-oligosaccharide and its structure. Additional signals in the spectrum of the sialo-compound are those of the H-3_{eg} and H-3_{ax} protons of sialic acid. The introduction of sialic acid at position 6 of the galactose residues gives rise to a few small but significant shift increments for some anomeric protons. There is a downfield shift of H-1 of Man 4 and 4′ and GlcNAc 5 and 5′, whereas an upfield shift is present for the anomeric protons of Gal 6 and 6′. When only one sialic acid residue occurs in $\alpha(2\rightarrow6)$ linkage to Gal, then its location in the biantenna can be inferred from the spectrum. For this purpose the chemical shifts of the anomeric protons of the mannose residues 4 and 4′ are used.

Sialic acid also occurs in $\alpha(2\rightarrow3)$ linkage to Gal. Such a sialic acid residue can easily be recognized from the chemical shifts of the H-3_{eq}

Table 1

Linkage	Chemical shift of H-3	
	Equatorial	Axial
NeuAcα(2→3)Gal	2.76	1.80
NeuAcα(2→6)Gal	2.67	1.72

and H-3_{ax} protons. These values together with those for α(2→6) linked sialic acid are given in table 1. The only H-1 that undergoes a significant shift on attachment of sialic acid to position 3 of galactose is that of galactose itself. In fig. 9, the spectrum is presented of a sialo-oligosaccharide having the biantennary structure with α(2→3) linked sialic acid residues in both branches. In contrast to α(2→6) linked sialic acid it is, in a monosialo biantenna, not possible to distinguish whether α(2→3)-linked sialic acid is attached to the upper or the lower branch because the anomeric protons of Gal 6 and 6′ have the same chemical shift.

An interesting glycoprotein is α_1-acid glycoprotein which shows a large heterogeneity in the carbohydrate structures [7]. Bi-, tri-, and tetraantennary structures occur in this protein. Furthermore, in the tri- and tetraantennary structures fucose can be present in an α(1→3)

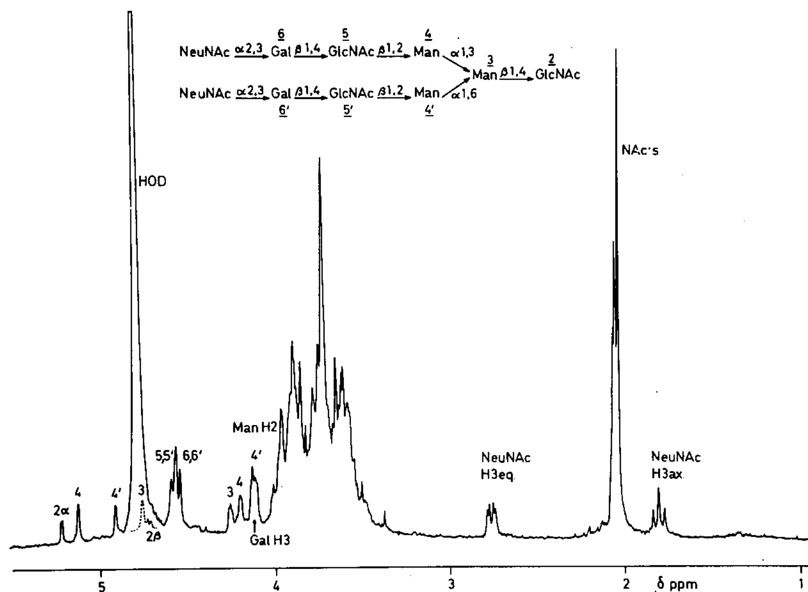

Fig. 9. 360 MHz ^{1}H-NMR spectrum of sialo-oligosaccharide VII in D_2O.

Table 2

	Chemical shift of					
	H-1 of mannose			H-2 of mannose		
Structure	3	4	4′	3	4	4′
Biantenna	4.764	5.121	4.928	4.247	4.189	4.110
Triantenna	4.757	5.119	4.924	4.215	4.215	4.109
Tetraantenna	4.754	5.127	4.866	4.215	4.215	4.092

linkage to GlcNAc. The determination of the antenna type can be carried out by high resolution proton NMR spectroscopy [8]. The substitution pattern of the mannotriose branching core is reflected in the chemical shifts of the H-1 and H-2 protons of the mannose residues as shown in table 2.

Another interesting glycoprotein is conalbumin or ovotransferrin [9]. The carbohydrate moiety differs in several respects from sero- and lactotransferrin. For example, it does not contain galactose and sialic acid. The 360 MHz ^{1}H-NMR spectrum and the carbohydrate structure are given in fig. 10.

The interpretation was made on guidance of some ^{1}H-NMR spectra of partial structures of this glycopeptide. On the basis of the spectral data of the partial structures we discovered that the shift increments

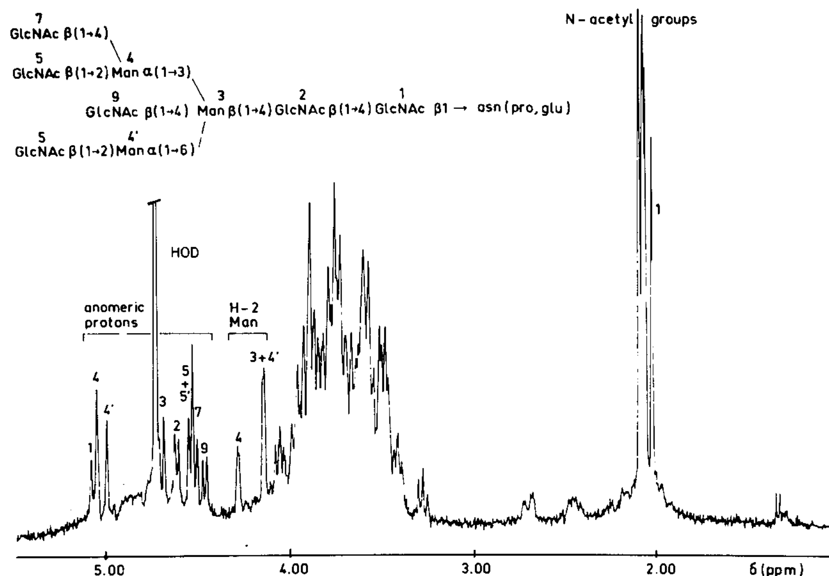

Fig. 10. 360 MHz ^{1}H-NMR spectrum of a glycopeptide from ovotransferrin.

	δ H_1 of residue									δ H_2 of residue		
	1	2	3	4	4'	5	5'	7	9	3	4	4'
predicted chemical shift	5.072	4.616	4.688	5.062	5.004	4.555	4.555	4.522	4.471	4.150	4.275	4.151
observed chemical shift	5.079	4.621	4.685	5.062	5.001	4.543	4.543	4.516	4.468	4.143	4.284	4.143

Fig. 11. Predicted and observed chemical shifts for the glycopeptide of ovotransferrin.

which result from making compounds more complex by the addition of a monosaccharide are, in general, additive. This opened the possibility to predict the spectra of compounds which are closely related to the set of reference compounds. In fig. 11 the comparison is shown of the predicted and observed values for the characteristic protons of this glycopeptide. The agreement is excellent.

To summarize: high resolution ^{1}H-NMR spectroscopy is a powerful technique for the structure determination of the carbohydrate chains of glycoproteins.

Acknowledgements

Thanks are due to Dr. J. Montreuil and co-workers, Lilly, France and to Dr. K. Schmid and co-workers, Boston, USA.

These investigations were supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO).

References

[1] G. A. Jamieson, M. Jett and S. L. DeBernardo, J. Biol. Chem. 246 (1971) 3686.
[2] G. Spik, B. Bayard, B. Fournet, G. Strecker, S. Bouquelet and J. Montreuil, FEBS Lett. 50 (1975) 296.
[3] L. Dorland, J. Haverkamp, B. L. Schut, J. F. G. Vliegenthart, G. Spik, G. Strecker, B. Fournet and J. Montreuil, FEBS Lett. 77 (1977) 15.
[4] J. Montreuil, Pure Appl. Chem. 42 (1975) 431.
[5] L. Dorland, B. L. Schut, J. F. G. Vliegenthart, G. Strecker, B. Fournet, G. Spik and J. Montreuil, Eur. J. Biochem. 73 (1977) 93.
[6] L. Dorland, J. Haverkamp, J. F. G. Vliegenthart, G. Strecker, J.-C. Michalski, B. Fournet, G. Spik and J. Montreuil, Eur. J. Biochem. 87 (1978) 323.
[7] B. Fournet, J. Montreuil, G. Strecker, L. Dorland, J. Haverkamp, J. F. G. Vliegenthart, J. P. Binette and K. Schmid, Biochemistry 17 (1978) 5206.
[8] L. Dorland, J. Haverkamp, J. F. G. Vliegenthart, B. Fournet, G. Strecker, G. Spik, J. Montreuil, K. Schmid and J. P. Binette, FEBS Lett. 89 (1978) 149.
[9] L. Dorland, J. Haverkamp, J. F. G. Vliegenthart, G. Spik, B. Fournet and J. Montreuil, Eur. J. Biochem. 100 (1979) 569 .