

To further characterize the asparagine-linked oligosaccharides of procollagen type I, exhaustive proteolytic digestion of the purified carboxyl-terminal propeptide was carried out [14]. High-resolution (500 MHz) $^1\text{H-NMR}$ spectroscopy was performed on the purified glycopeptide preparation to elucidate the primary sequence of the carbohydrate units. Here, we report the results of this study which demonstrate the oligosaccharides of the carboxyl-propeptide to be a mixture of high-mannose type components displaying microheterogeneity with respect to the number of mannose residues.

Materials and Methods

Materials

$^2\text{H}_2\text{O}$ (99.96 atom% ^2H) and 4-vinyl pyridine were purchased from Aldrich Chemical Co. Papain (two-times crystallized) and pronase (type XIV) were obtained from Sigma Chemical Co and Sephadex gels were products of Pharmacia Fine Chemicals. The commercial sources of the other chemicals and materials have been documented earlier [14]. Chemicals were reagent grade or purer.

Preparation of the carboxyl-terminal propeptides

The propeptide fraction was purified from the culture medium of 17-day old chick embryo calvaria (3 times 24 h incubation periods) as previously described [14]. Final purification of the propeptide was achieved by Sephadex G-150 gel filtration (1.1×145 cm) chromatography using 0.2 M $(\text{NH}_4)\text{HCO}_3$ as the eluent.

The purified propeptides were subsequently dissolved in 0.1 M Tris-HCl (pH 8.5)/8.0 M urea/1.0 mM EDTA at a concentration of approx. 6 mg per ml and reduced by addition of dithiothreitol at a twenty-fold molar excess over protein disulfide concentration. After 3 h at room temperature, the reaction was terminated by the addition of 4-vinyl pyridine (3-fold molar excess over the thiol reagent). After 15 min the resulting propeptide preparation (C1 and C2) was desalted employing Sephadex G-75 chromatography (1.4×120 cm) in 0.2 M $(\text{NH}_4)\text{HCO}_3$ and then stored in a lyophilized form at -20°C . The purity of the propeptide was evaluated by SDS polyacrylamide gel electrophoresis as described previously [15].

Isolation of the glycopeptides from the carboxyl-terminal propeptide

Approx. 25 mg of the purified, reduced polypeptide mixture were dissolved in 1.0 ml of 0.28 M cysteine-HCl (pH 8.0), containing 0.124 M NaCN/0.2 M Na_2EDTA , and then digested for 24 h at 60°C with 3.0 mg of papain added in four equal aliquots at 0, 3, 9 and 20 h. A small amount of toluene was added at 9 and at 20 h to inhibit bacterial contamination. Next, the digestion mixture (pH ≈ 8.5) was lowered to 40°C and three 1.0 mg aliquots of pronase were added at intervals of 4 h. Five additional 0.2 mg aliquots of pronase were added at equal time-intervals over the next 48 h.

The resulting glycopeptide mixture was purified by sequential gel filtration chromatography, first using Sephadex G-50 (Fine) and then Sephadex G-25. The elution profiles were monitored by absorbance at 230 nm and for neutral hexoses [16]. The column elution profiles together with the experimental details are given in Fig. 1.

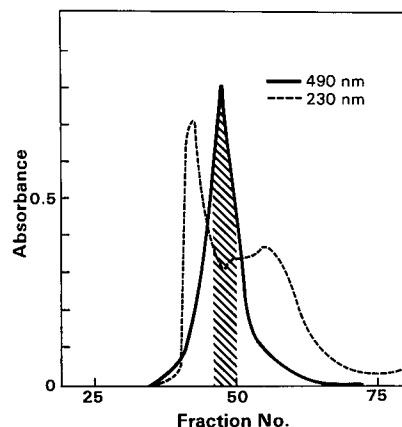


Fig. 1. Chromatography of the glycopeptides derived from the carboxyl-terminal propeptide of Type I procollagen. The glycopeptide mixture of the papain-pronase digest of the propeptide was first fractionated on a Sephadex G-50 column (1.1×150 cm) equilibrated with 50 mM $(\text{NH}_4)\text{HCO}_3$ (chromatogram not shown). The resulting glycopeptide fraction was then applied to a Sephadex G-25 column (0.9×150 cm) which had also been equilibrated with 50 mM $(\text{NH}_4)\text{HCO}_3$. Fractions 46–50 were pooled, lyophilized and subjected to NMR spectroscopy and amino acid and carbohydrate analyses as described in Methods. Columns were run at a flow rate of 5 ml per h and 1.2 ml fractions were collected.

Carbohydrate analyses

Quantitative sugar analyses were performed as follows. An aliquot of the glycopeptide preparation (0.1 mg) was solvolyzed by 1 M HCl in methanol for 24 h at 85 °C. Gas-liquid chromatography of the trimethylsilyl derivatives of the resulting methylglycosides was conducted on a Varian 3700 apparatus equipped with a WCOT CPSi15 fused silica capillary column (25 m × 0.3 mm) using flame-ionization detection. Oven temperature was programmed from 130 to 220 °C (2°/min); N₂ was used as carrier gas (flow rate 1.5 ml/min).

500 MHz ¹H-NMR spectroscopy

Prior to NMR spectroscopic analysis, the glycopeptide preparation (1.1 mg) was repeatedly treated with ²H₂O at room temperature, with intermediate lyophilization. Finally, the sample was redissolved in 400 μl ²H₂O. 500 MHz ¹H-NMR spectroscopy was performed on a Bruker WM-500 spectrometer (SON hf-NMR facility, Department of Biophysics, University of Nijmegen, The Netherlands) operating in the Fourier transform mode and equipped with a Bruker Aspect-2000 computer, as described previously [17,18]. Resolution enhancement of the spectrum was achieved by Lorentzian-to-Gaussian transformation. The probe temperature was 27.0 ± 0.1 °C. Chemical shifts (δ) are expressed in ppm downfield from internal sodium 4,4-dimethyl-4-silapentane-1-sulfonate (DSS), but were actually measured by reference to internal acetone (δ = 2.225 in ²H₂O at 27 °C) with an accuracy of 0.002 ppm.

Results and Discussion

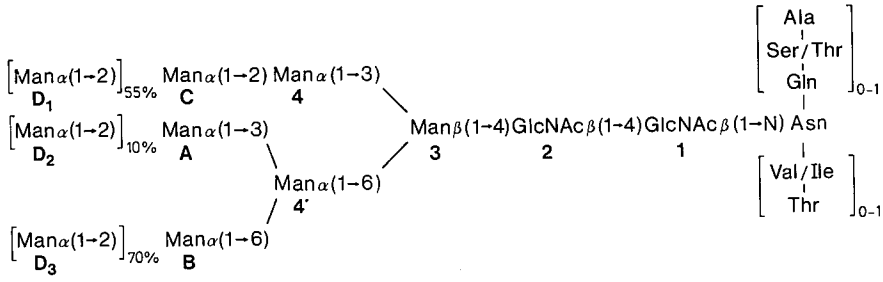
500 MHz ¹H-NMR spectroscopy was employed to elucidate the primary carbohydrate structure of the procollagen glycopeptide fraction. The overall spectrum is presented in Fig. 2A along with the expanded resolution-enhanced structural reporter group regions (Fig. 2B). Relevant NMR data of the oligomannoside-type components of the glycopeptide mixture are listed in Table I. The signals of the spectrum were assigned by using the ¹H-NMR data of the reference glycopeptides Man₆(GlcNAc)₂Asn isolated from hen ovalbumin [18], Man₈(GlcNAc)₂Asn derived from bovine

lactotransferrin [18,19] and Man₉(GlcNAc)₂Asn from soybean agglutinin [18,20] which were also acquired at 500 MHz (Table I).

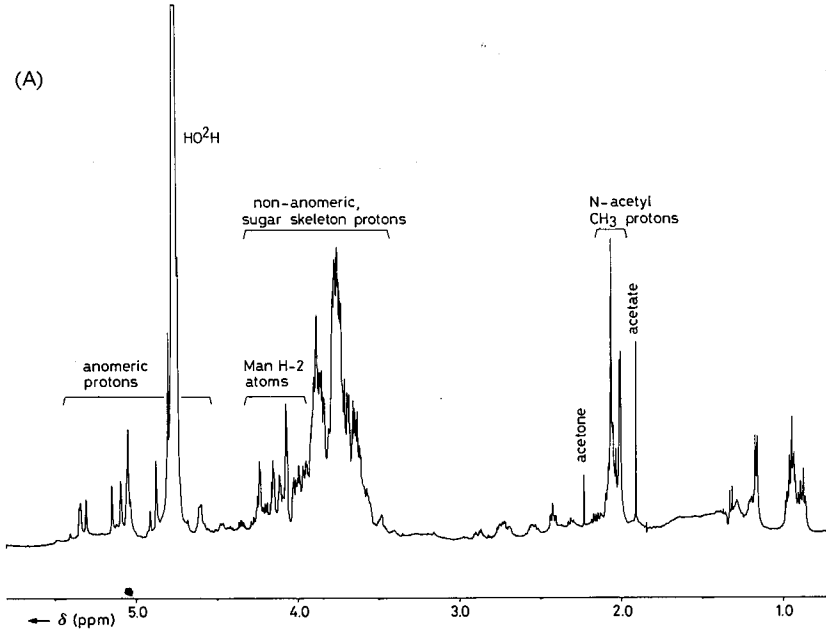
From the comparison of the spectrum of the glycopeptide mixture obtained from the procollagen carboxyl-extension with those of the reference glycopeptides [18], it may be inferred that all of the constituents have in common Man residues **3**, **4**, **4'**, **A**, **B**, and **C** and the *N,N'*-diacetylchitobiose unit (for designation of monosaccharide residues, see Fig. 2). Therefore, the smallest component of the mixture contains at least six Man residues and is identical to Man₆(GlcNAc)₂Asn derived from hen ovalbumin (see Table I). The *N*-acetyl signal of GlcNAc-1 and also its H-1 signal are split into two or even more signals each; this reflects the heterogeneity of the peptide moiety. The amino acid composition of this moiety was found to be consistent with the previously reported amino acid sequences of the carboxyl-terminus of procollagen [11,12]. The peptide moiety contained Asn, Glu, Thr, Val, Gly and Pro in a molar ratio of 1.0 : 0.57 : 0.43 : 0.35 : 0.27 and 0.23. Traces of other amino acids were also found. The chemical shift of the *N*-acetyl signal of GlcNAc-2 (δ = 2.059) is characteristic for the extension of the pentasaccharide core with Man residues only [18].

Besides the β-linked Man-3, clearly characterized by its H-2 signal at δ = 4.232 (see Fig. 2B and Table I), only α-linked Man residues occur in the peripheral part of the glycan chains. This can be deduced from the chemical shifts of their H-1 signals in combination with their *J*_{1,2} values [18]. The set of chemical shifts of H-1 (δ = 4.867) and H-2 (δ = 4.146) of Man-4' is indicative of disubstitution of this residue at C-3 and C-6 by Man-A and -B, respectively [18,20]. The H-1 atom of Man-B gives rise to two doublets at δ = 4.906 (terminal B) and at δ = 5.144 (B substituted with D₃ at C-2) having an intensity ratio of 3 : 7. From these values it may be concluded that 70% of the Man-B residues in the glycopeptide sample bear an α(1 → 2)-linked Man-D₃ (δ_{H-1} = 5.041), whereas in the remaining part Man-B occupies a terminal position in the chain.

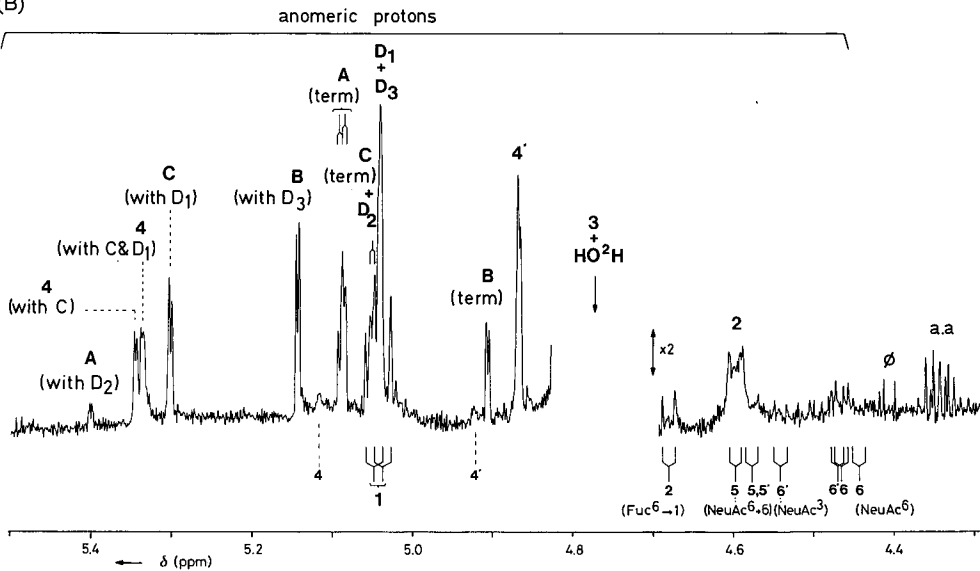
The H-1 atom of Man-A resonates in doublets at δ = 5.09 and δ = 5.401 in the intensity ratio of 9 : 1. The latter chemical shift value points to the



(A)



(B)



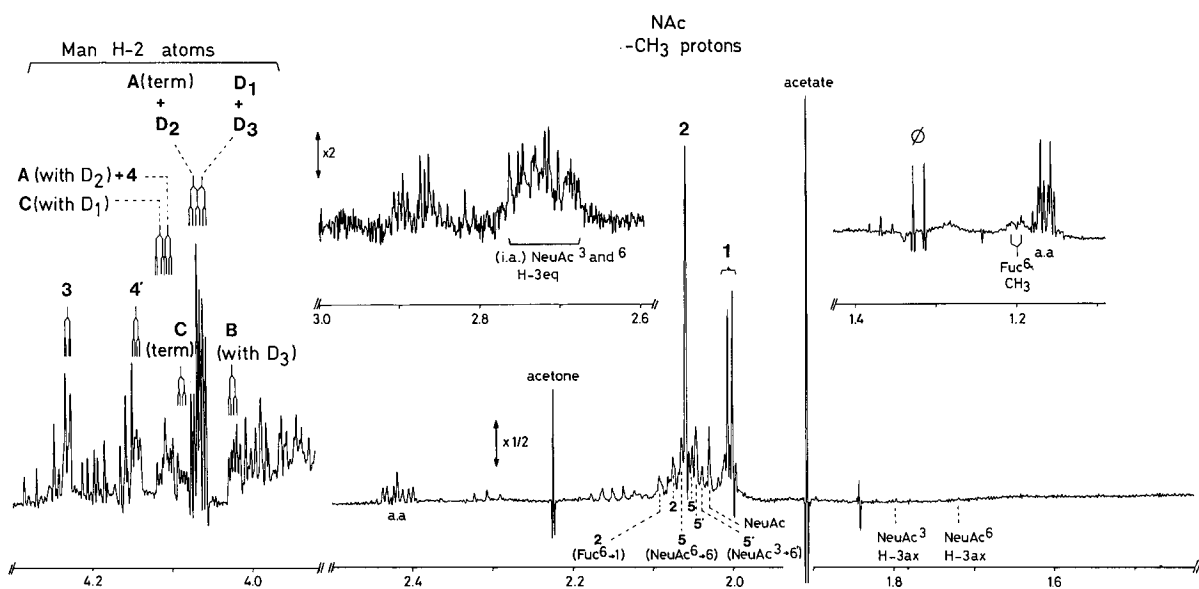
presence of Man- D_2 in $\alpha(1 \rightarrow 2)$ -linkage to Man-A; therefore, in 10% of the structures in the glycopeptide mixture, Man-A is substituted by Man- D_2 ($\delta_{H-1} = 5.051$). Thus, the remaining 90% of the structures contain Man-A in terminal, non-reducing position. Interestingly, the H-1 signal for the terminal Man-A is doubled ($\delta = 5.086$ and $\delta = 5.091$) (see Fig. 2B). This doubling may be explained in terms of heterogeneity of the peptide moiety (compare Refs. 18 and 21). The latter statement is in complete accord with the model which we proposed for the favoured conformation of oligomannoside-type carbohydrate chains in solution [18,20,22]; Man-A is in relatively close spatial proximity of the chitobiose core, thereby making the chemical shifts of its protons susceptible to the structure of the peptide moiety.

For H-1 of Man-4 in the upper branch, two doublets are observed at $\delta = 5.345$ and $\delta = 5.337$ in the ratio of 45:55. The approximate value of the chemical shift for this proton ($\delta \approx 5.34$), in

conjunction with both the absence of an H-1 signal at $\delta = 5.11$ as well as the singularity of the Man-3 H-2 signal, indicate that in all components of the mixture Man-4 bears the $\alpha(1 \rightarrow 2)$ -linked Man-C [18-22]. In turn, H-1 of Man-C gives rise to two doublets at $\delta = 5.302$ (C substituted with D_1 at C-2) and $\delta = 5.051$ (terminal C). These features for the H-1 signals of Man-4 and Man-C can be explained by the presence of Man- D_1 in 55% of the structures in the glycopeptide mixture. The presence of Man- D_1 ($\delta_{H-1} = 5.041$) causes a downfield shift of the H-1 signal of Man-C ($\Delta\delta = 0.251$ ppm) and a slight upfield shift of H-1 of Man-4 ($\Delta\delta = -0.008$ ppm) as compared to the Man-4-Man-C branch without Man- D_1 (see Man $_g$ - and Man $_f$ -compounds, Table I and compare Refs. 18, 19 and 22).

From the $^1\text{H-NMR}$ data discussed in detail above, the following essential aspects of the structure of the carbohydrate units of the procollagen carboxyl-extension can be deduced. The carbo-

Fig. 2. (A) Overall 500 MHz $^1\text{H-NMR}$ spectrum of the glycopeptide preparation derived from the carboxyl-terminal peptide of procollagen from chick calvaria, recorded in $^2\text{H}_2\text{O}$ at 27°C . (B) Expanded structural-reporter-group regions of spectrum (A) after resolution enhancement. The numbers and letters above the spectrum refer to the corresponding residues in the structure on top of the figure. Assignments below the spectrum refer to residues in a contaminating, diantennary *N*-acetylglucosamine-type glycopeptide (for complete structure, see footnote c to Table II). Relative-intensity scales of various parts of the spectrum in (B) differ from that of the α -anomeric region ($4.8 < \delta < 5.5$), as indicated. The HO 2 H-line, as well as the coinciding H-1 signal of Man-3, have been omitted from spectrum (B); their position is indicated by an arrow. (a.a. denotes signal from amino acid proton. Signals marked by \emptyset stem from a frequently occurring non-protein non-carbohydrate contaminant of unknown structure.)



hydrate components (greater than 95%) of the glycopeptide preparation derived from the carboxyl-terminal peptide of procollagen are of the oligomannoside type. (It should be noted that a contaminant (5–10%) glycopeptide was present, which possesses a diantennary structure of the *N*-acetylglucosamine type, as explained in the legend to Table II.) All glycopeptides contained at

least six Man residues; however, heterogeneity occurs in the carbohydrate moiety with respect to the presence of each of the Man-D residues. Consideration of chemical shifts, in conjunction with relative intensities, of the anomeric-proton signals in the ¹H-NMR spectrum leads to relative abundances of Man-D₁, -D₂ and -D₃ of 55%, 10% and 70%, respectively. These values are consistent with

TABLE I

¹H CHEMICAL SHIFTS OF STRUCTURAL-REPORTER GROUPS OF CONSTITUENT MONOSACCHARIDES FOR THE GLYCOPEPTIDE PREPARATION FROM THE CARBOXYL-TERMINAL PEPTIDE OF PROCOLLAGEN, TOGETHER WITH THOSE FOR SOME RELATED OLIGOMANNOSIDE-TYPE GLYCOPEPTIDES

n.d., value could not be determined.

Reporter group	Residue ^a	Chemical shift ^b in				
		Man _x (GlcNAc) ₂ Asn procollagen		Man ₆ (GlcNAc) ₂ Asn ovalbumin [18]	Man ₈ (GlcNAc) ₂ Asn lactotransferrin [18,19]	Man ₉ (GlcNAc) ₂ Asn soybean agglutinin [18,20]
		with D	without D			
H-1	GlcNAc-1'	{ 5.038 ^{c,d} 5.049 ^{c,d}		5.070	5.092	5.092
	GlcNAc-2	4.598 ^d		4.605	4.608	4.610
NAc	GlcNAc-1	{ 2.001 ^c 2.007 ^c		2.011	2.007	2.015
	GlcNAc-2	2.059		2.061	2.066	2.067
H-1	Man-3	4.77		4.770	4.77	4.77
	Man-4	5.337 (55) ^c	5.345 (45)	5.342	5.345	5.334
	Man-4'	4.867		4.871	4.868	4.869
	Man-A	5.401 (10)	{ 5.086 (90) ^c 5.091	5.095	5.401	5.404
	Man-B	5.144 (70)	4.906 (30)	4.908	5.141	5.143
	Man-C	5.302 (55)	5.051 (45)	5.052	5.059	5.308
	Man-D ₁	5.041 (55)	–	–	–	5.049
	Man-D ₂	5.051 (10)	–	–	5.059	5.061
	Man-D ₃	5.041 (70)	–	–	5.040	5.042
H-2	Man-3	4.232		4.229	4.228	4.228
	Man-4	4.104		4.114	4.10	4.098
	Man-4'	4.146		4.143	4.15	4.156
	Man-A	4.104	4.071	4.066	4.10	4.109
	Man-B	4.025	n.d.	3.990	4.02	4.023
	Man-C	4.113	4.087	4.066	4.07	4.109
	Man-D ₁	4.064	–	–	–	4.073 ^f
	Man-D ₂	4.071	–	–	4.07	4.073
	Man-D ₃	4.064	–	–	4.07	4.066 ^f

^a For complete structures and designation of monosaccharide residues, see Fig. 2.

^b Chemical shifts were acquired at 500 MHz for ²H₂O solutions at 27 °C; they are expressed in ppm downfield from internal DSS.

^c Multiplicity of signal is due to heterogeneity of the peptide moiety (see text).

^d Centering δ values between the two columns for the procollagen sample indicates that these chemical shifts are not sensitive to heterogeneity with regard to the Man-D residues. Their relative abundances were assumed to be 100%.

^e Numbers in parentheses indicate the centesimal abundance (±5%) of components with and without the Man-D residue in question.

^f Assignments are interchangeable.

nose unit is not well understood. Obviously, it can play a role in both the assembly and/or turnover of the procollagen molecule. Since the carboxyl-propeptide is isolated from the extracellular milieu, it is not certain what the structure and composition of the original post-translational carbohydrate unit is. Studies are now underway to isolate the intracellular product from the chick calvaria to clarify this point.

Acknowledgements

This investigation was supported by USPHS, NIH grants HL-19717, HL-13262 and GM-10374, by the Netherlands Foundation for Chemical Research (SON/ZWO) and by the Netherlands Foundation for Cancer Research (KWF), grant UUKC 83-13. We are grateful to Mr. George Crombie for carrying out the amino acid analyses.

References

- 1 Fietzek, P.P. and Kühn, K. (1976) *Int. Rev. Connect. Tissue Res.* 7, 1-60
- 2 Miller, E.J. (1976) *Mol. Cell Biochem.* 13, 165-192
- 3 Prockop, D.J., Berg, R.A., Kivirikko, K.I. and Uitto, J. (1976) in *Biochemistry of Collagen* (Ramachandran, G.N. and Reddi, A.H., eds.), pp. 163-273, Plenum, New York
- 4 Kivirikko, K.I. and Myllylä, R. (1979) *Int. Rev. Connect. Tissue Res.* 8, 23-72
- 5 Clark, C.C. and Kefalides, N.A. (1976) *Proc. Natl. Acad. Sci. U.S.A.* 73, 34-38
- 6 Duksin, D. and Bornstein, P. (1977) *J. Biol. Chem.* 252, 955-962
- 7 Olsen, B.R., Guzman, N.A., Engel, J., Condit, C. and Aase, S. (1977) *Biochemistry* 16, 3030-3036
- 8 Anttinen, H.O., Karinen, A., Ryhänen, L. and Kivirikko, K.I. (1978) *FEBS Lett.* 87, 222-226
- 9 Clark, C.C. and Kefalides, N.A. (1978) *J. Biol. Chem.* 253, 47-51
- 10 Clark, C.C. (1979) *J. Biol. Chem.* 254, 10798-10802
- 11 Showalter, A.M., Pesciotta, D.M., Eikenberry, E.F., Yamamoto, T., Pastan, I., DeCrombrugge, B., Fietzek, P.P. and Olsen, B.R. (1980) *FEBS Lett.* 111, 61-65
- 12 Pesciotta, D.M., Dickson, L.A., Showalter, A.M., Eikenberry, E.F., DeCrombrugge, B., Fietzek, P.P. and Olsen, B.R. (1981) *FEBS Lett.* 125, 170-174
- 12 (a) Cunningham, L.W. and Frederiksen, D.W. (eds.) (1982) *Methods Enzymol.* 82A, 1006
- 13 Fuller, F. and Boedtker, H. (1981) *Biochemistry* 20, 996-1996
- 14 Kosen, P.A. and Franzblau, C. (1982) *Biochemistry* 21, 4278-4284
- 15 Laemmli, U.K. (1970) *Nature* 227, 680-685
- 16 Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A. and Smith, F. (1956) *Anal. Chem.* 28, 350-356
- 17 Vliegthart, J.F.G., Van Halbeek, H. and Dorland, L. (1981) *Pure Appl. Chem.* 53, 45-77
- 18 Vliegthart, J.F.G., Dorland, L. and Van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem.* 41, 209-374
- 19 Van Halbeek, H., Dorland, L., Vliegthart, J.F.G., Spik, G., Chéron, A. and Montreuil, J. (1981) *Biochim. Biophys. Acta* 675, 293-296
- 20 Dorland, L., Van Halbeek, H., Vliegthart, J.F.G., Lis, H. and Sharon, N. (1981) *J. Biol. Chem.* 256, 7708-7711
- 21 Van Halbeek, H., Dorland, L., Vliegthart, J.F.G., Jouanneau, J. and Bourrillon, R. (1981) *Biochem. Biophys. Res. Commun.* 99, 886-892
- 22 Van Halbeek, H., Dorland, L., Veldink, G.A., Vliegthart, J.F.G., Strecker, G., Michalski, J.-C., Montreuil, J. and Hull, W.E. (1980) *FEBS Lett.* 121, 71-77
- 23 Hubbard, S.C. and Ivatt, R.J. (1981) *Annu. Rev. Biochem.* 50, 555-584