Communication

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 256, No. 15, Issue of August 10, pp. 7708-7711, 1981
Printed in U.S.A.

Primary Structure of the Carbohydrate Chain of Soybean Agglutinin

A REINVESTIGATION BY HIGH RESOLUTION ^{1}H NMR SPECTROSCOPY*

(Received for publication, February 2, 1981, and in revised form, April 20, 1981)

Lambertus Dorland, Herman van Halbeek, and Johannes F. G. Vliegenthart

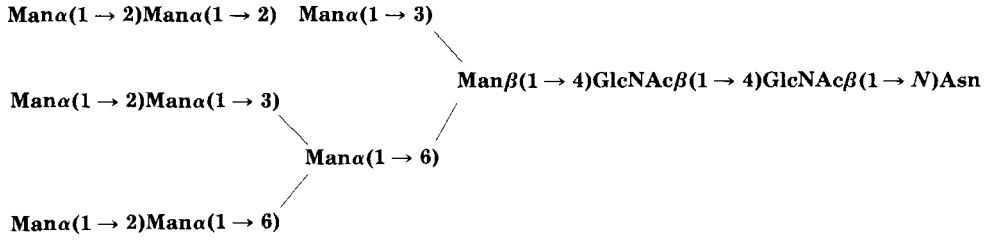
From the Department of Bio-Organic Chemistry, University of Utrecht, Croesestraat 79, NL-3522 AD Utrecht, The Netherlands

Halina Lis and Nathan Sharon

From the Department of Biophysics, The Weizmann Institute of Science, Rehovoth, Israel

In an earlier report (Lis, H., and Sharon, N. (1978) *J. Biol. Chem.* 253, 3468-3476) it was concluded that the plant glycoprotein soybean agglutinin contains two types of *N*-glycosidically linked carbohydrate units, both consisting of mannose and *N*-acetylglucosamine, in a molar ratio of 9:2. Both were described to have the same core structure, Manα(1 → 6)[Manα(1 → 3)]Manβ(1 → 4)GlcNAcβ(1 → 4)GlcNAc, but different patterns of branching.

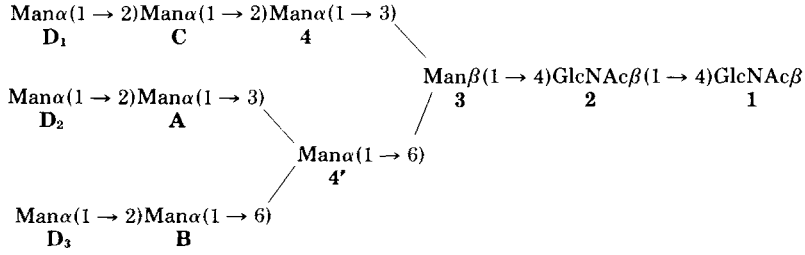
Here we present data showing that the asparagine-linked carbohydrate moiety of soybean agglutinin is homogeneous and possesses the following structure:



This conclusion is based on high resolution ^{1}H NMR spectroscopy at 500 MHz of the isolated glycopeptide, and at 360 MHz of the oligosaccharide $Man_9GlcNAc$ obtained after digestion of the crude soybean agglutinin glycopeptide by endo-β-*N*-acetylglucosaminidase H. The revised structure is identical in all respects with that of the high mannose *N*-glycosidic units of porcine thyroglobulin, of bovine lactotransferrin, and of the glycoprotein from Chinese hamster ovary cell membranes.

Glycoproteins are ubiquitous in nature, *e.g.* as enzymes, hormones, lectins, immunoglobulins, and structural elements. To understand the role of the carbohydrate units of these compounds in the biochemical processes wherein they are involved, the primary structures of these units have to be elucidated.

Despite the diversity of glycoproteins, their carbohydrate components have many common structural features (1-3). In particular, the oligosaccharide chains *N*-glycosidically linked to an asparagine residue of the protein backbone appear to possess a virtually invariant core structure of a pentasaccharide, $Man_3GlcNAc_2$.[1] For the oligomannoside-type structures, containing no other sugars but only additional mannoses linked to the core region, nearly all investigations led to the following structure in its entirety (see below) or to partial structures thereof, lacking one or more of the α(1 → 2)-linked mannoses.



One of the few exceptions hitherto reported (1-4) was the carbohydrate structure of the lectin soybean agglutinin. After isolation and purification of the glycopeptide from a pronase digest, the structure of its sugar moiety was characterized by

* 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) and by Grant UUKC-OC 79-13 from the Netherlands Foundation for Cancer Research (KWF). 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.

[1] All sugars are of the D-configuration.

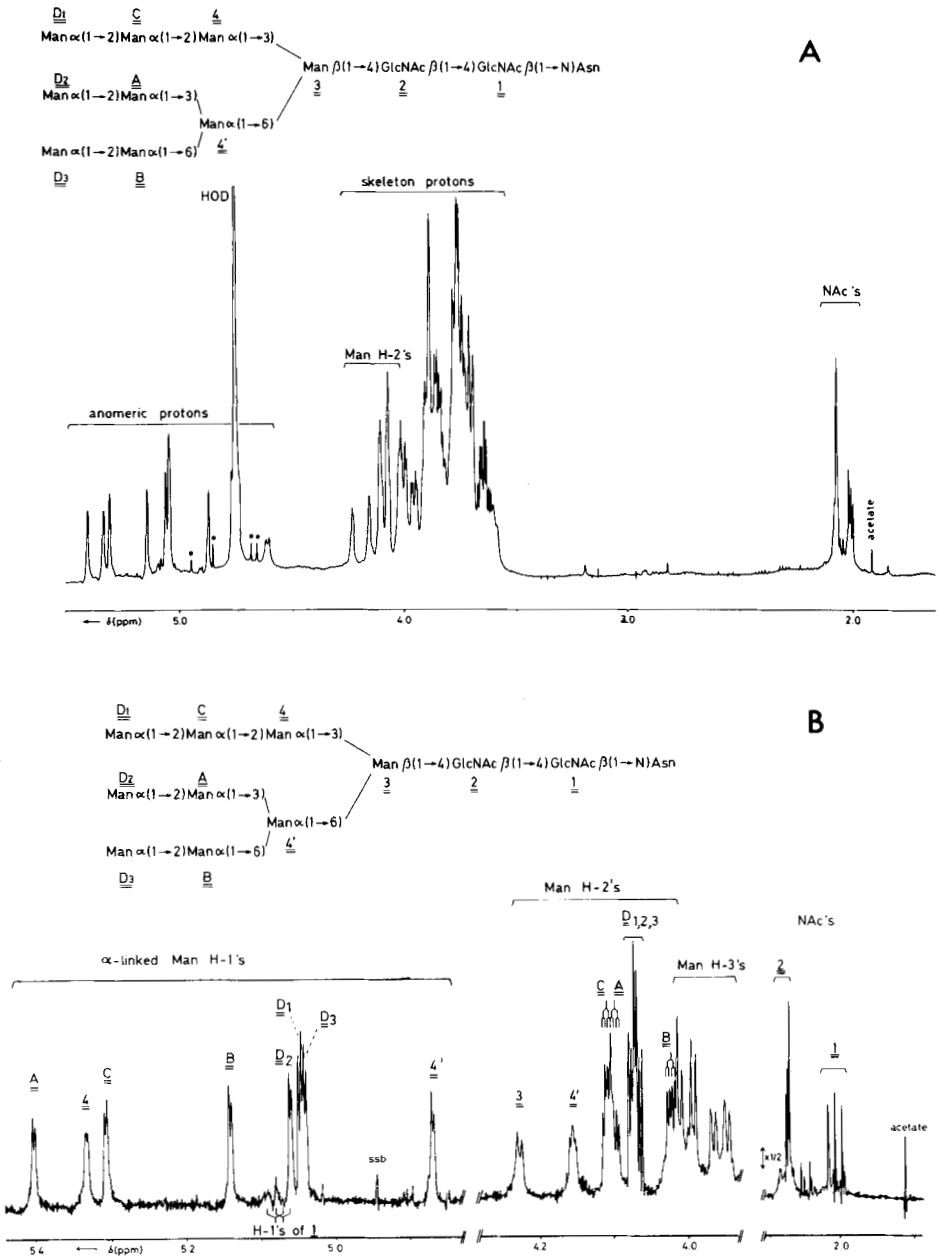


FIG. 1. **The 500-MHz ¹H NMR spectrum of $Man_9GlcNAc_2Asn$.** *A*, overall 500-MHz ¹H NMR spectrum of glycopeptide $Man_9GlcNAc_2Asn$ from soybean agglutinin in D_2O at 300 K together with its structure. *Asterisks*, spinning side bands. *B*, structural reporter group regions of the resolution-enhanced 500-MHz ¹H NMR spectrum of $Man_9GlcNAc_2Asn$ in D_2O at 300 K. The *numbers* and *letters* in the spectrum refer to the corresponding residues in the structure.

chemical and enzymic methods. Mainly on the basis of methylation analysis, it was proposed that two types of carbohydrate moieties occur in SBA.[2] One contains two branches of only $\alpha(1 \rightarrow 2)$-linked mannoses attached to the core pentasaccharide, and another one has three branches, having the second branching point on the $\alpha(1 \rightarrow 3)$-linked mannose residue of the core (Man 4) (4). Since SBA is one of the very few, rather well characterized plant glycoproteins, the question arose whether these structures would be specific for plants, implicating another way of biosynthesis and/or processing of the carbohydrate chain than generally accepted to occur in animals (*e.g.* Ref. 5).

During the last few years high resolution ¹H NMR spectroscopy has proved to be a powerful method for the determination of primary structures of carbohydrate chains of a wide variety of glycoproteins (6). Recently, considerable progress has been made in applying this technique for elucidation of oligomannoside-type structures (7–12). Therefore, it was decided to reinvestigate the carbohydrate moiety of SBA by means of high resolution ¹H NMR spectroscopy.

Solutions in D_2O (2 mM) of the glycopeptide, $Man_9GlcNAc_2Asn$ (4), and of the corresponding decasaccharide, $Man_9GlcNAc$ (prepared by digestion of the crude glycopeptide with endo-β-*N*-acetylglucosaminidase H, followed by purification by ion exchange chromatography on a column (0.9 × 60 cm) of Dowex X2 in pyridine/acetate buffer, pH 3.4) were adjusted to pD ≃ 7. Deuterium-exchanged samples were prepared by a 5-fold lyophilization of the solutions, finally using 99.96 atom % deuterated D_2O (Aldrich). The 500-MHz ¹H NMR spectra of the glycopeptide were recorded on a Bruker WM-500 spectrometer, operating in the pulsed Fourier transform mode at probe temperatures of 300 K or 330 K, and equipped with a Bruker Aspect-2000 computer. The 360-MHz ¹H NMR spectrum of the oligosaccharide was recorded at 300 K on a Bruker HX-360 spectrometer, equipped with a Bruker BNC-12 computer. Resolution enhancement was achieved by Lorentzian to Gaussian transformation from quadrature phase detection, followed by complex Fourier transformation (13). Chemical shifts at 300 K are expressed in parts per million downfield from internal sodium 2,2-dimethyl-2-silapentane-5-

[2] The abbreviation used is: SBA, soybean agglutinin.

TABLE I

¹H chemical shifts of structural reporter groups of constituent monosaccharides for the carbohydrate moiety of soybean agglutinin and for a urinary oligosaccharide from mannosidosis patients (11)

Reporter group	Residue[a]	$Man_9GlcNAc_2Asn$ from soybean agglutinin	GlcNAc anomer	$Man_9GlcNAc$ from	
				Soybean agglutinin	Mannosidosis
		Chemical shift[b]		Chemical shift[c]	Chemical shift[b]
H-1 of	1	5.092			
	2	4.610	α	5.229	5.231
			β	4.72[d]	4.714
NAc of	1	2.015[e]			
	2	2.067[e]	α	2.050	2.050
			β	2.047	2.046
H-1 of	3	4.77[d]	α	4.78[d]	4.776
			β		4.772
	4	5.334	α	5.333[f]	5.337
			β		5.335
	4′	4.869	α	4.866	4.872
			β	4.862	4.869
	A	5.404	α	5.397	5.398
			β	5.409	5.407
	B	5.143	α,β	5.141	5.142
	C	5.308	α,β	5.308	5.308
	D_1	5.049	α,β	5.043	5.048
	D_2	5.061	α	5.053	5.058
			β	5.059	5.063
	D_3	5.042	α,β	5.036	5.040
H-2 of	3	4.228	α	4.237	4.239
			β	4.226	4.229
	4	4.098	α,β	4.090	4.089
	4′	4.156	α	4.155	4.158
			β	4.149	4.155
	A	4.109	α	4.105	4.106
			β	4.094	4.103
	B	4.023	α,β	4.016	4.025
	C	4.109	α,β	4.105	4.109
	D_1	4.073[g]	α,β	4.064[g]	4.069[g]
	D_2	4.073[g]	α,β	4.072[g]	4.073[g]
	D_3	4.066[g]	α,β	4.064[g]	4.066[g]

[a] For complete structure and coding of monosaccharide residues see Fig. 1.

[b] Data acquired at 500 MHz, $T = 300$ K, pD ≃ 7.

[c] Data acquired at 360 MHz, $T = 300$ K, pD ≃ 7.

[d] Value cannot be determined more accurately (± 0.01 ppm) due to interference of the HOD line at $T = 300$ K.

[e] Signals stemming from the main glycopeptide (see text).

[f] If for a certain proton $|\Delta\delta_{\alpha-\beta}| \leq 0.002$ ppm, the δ value obtained for this proton at 360 MHz is necessarily an average value, since such a Δδ is not detectable at 360 MHz (*cf.* Ref. 12).

[g] Assignments may be interchanged.

sulfonate, but were actually measured by reference to internal acetone (δ = 2.225 ppm) with an accuracy of 0.002 ppm.

The 500-MHz ¹H NMR spectrum of the SBA glycopeptide, recorded in D_2O at 300 K, is depicted in Fig. 1*A*. The expanded structural reporter group (6, 12) regions are given in Fig. 1*B*. The NMR spectral features are summarized in Table I.

Integration of the anomeric region of the spectrum, recorded at 330 K, is consistent with the occurrence of 9 mannose and 2 *N*-acetylglucosamine residues. The chemical shifts of the H-1's of GlcNAc **1** and **2** in conjunction with those of H-1 and H-2 of Man **3** point to the presence of the core pentasaccharide, *i.e.*, Manβ(1 → 4)GlcNAcβ(1 → 4)GlcNAcβ(1 → *N*)Asn which is substituted at C-3 and C-6 of Man **3** by α-linked mannoses (**4** and **4′**, respectively) (6, 12).

Besides the above mentioned β-anomeric proton signals, 8 H-1 signals of α-linked mannoses are observed. On the basis of previous findings (9, 11, 12) the 3 anomeric proton signals at δ ≃ 5.05 ppm are assigned to 3 terminal α(1 → 2)-linked mannose residues, pointing to three branches in the glycan structure. Furthermore, the chemical shift of the relatively high field doublet of H-1 of Man **4′** (δ = 4.869 ppm) is indicative of a substitution of this residue at C-3 and C-6 by α-linked mannoses (**A** and **B**, respectively). This is evident from high resolution ¹H NMR spectroscopy of glycopeptides from ovalbumin (9), from the μ chain of human IgM (9), and of glycoasparagines isolated from the urine of a patient with Gaucher's disease (10). In addition, the set of chemical shift values of the structural reporter groups of the SBA glycopeptide (Table I) is almost identical with that of a glycopeptide, containing 9 mannose residues, isolated from bovine lactotransferrin (1, 14). The close resemblance of this set with that of the $Man_9GlcNAc$ oligosaccharide from urine of mannosidosis patients (11) prompted the preparation of the SBA oligosaccharide, $Man_9GlcNAc$ (see above). Indeed, the NMR spectra of the oligosaccharides, $Man_9GlcNAc$, obtained from SBA and from mannosidosis urine turned out to be identical. It is thus evident that the carbohydrate moiety of SBA is homogeneous, and possesses a single structure with three branches, having Man **4′** as the second branching point (see Fig. 1*B*).

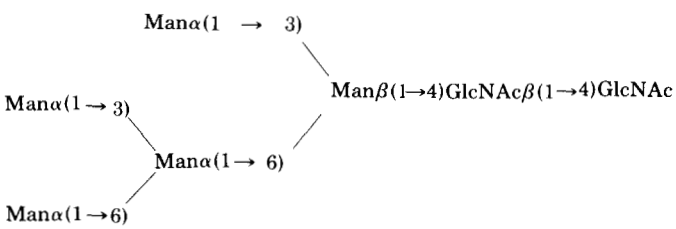
With regard to the 360-MHz ¹H NMR spectrum of the SBA oligosaccharide, it should be noted that this is in fact a superposition of the subspectra of the α- and β-anomer of this compound. These subspectra are different with respect to the chemical shifts of the structural reporter groups of the core residues (**2**, **3**, **4**, and **4′**) (see Table I), but also with respect to those of the Man residues **A** and $\mathbf{D_2}$ (11). This far-reaching influence of anomerization, which is due to the close spatial proximity of the **A**-$\mathbf{D_2}$ branch and the anomeric center of GlcNAc **2** (11), has been observed also by Cohen and Ballou (9) and by Atkinson (cited in Ref. 9), but has not been perceived as such. These authors assigned the signal, belonging to the H-1 of Man **A** in the β-anomer of oligosaccharides $Man_5GlcNAc$ and $Man_6GlcNAc$, as being derived from minor components of unspecified structures in mixtures of isomeric oligosaccharides (9). The report of Cohen and Ballou allows the deduction of an additional proof for the influence that the anomeric configuration of GlcNAc **2** exerts on the chemical environment of protons of other residues in the chain, since it describes that the above mentioned signal disappears upon reducing the oligosaccharide. From the spectra of $Man_5GlcNAc$ (Ref. 9, Fig. 8*a*) and of $Man_6GlcNAc$ (Ref. 9, Fig. 4), it can be derived that here the anomerization effect upon the H-1 of Man **A** is even larger than in the case when Man $\mathbf{D_2}$ is present ($|\Delta\delta_{\alpha-\beta}| \simeq 0.02$ ppm) (*cf.* Ref. 11 and Table I).

In the spectrum of the SBA glycopeptide (see Fig. 1*B*), sets of *N*-acetyl singlets are observed for GlcNAc **1** at δ = 2.015, 2.006, and 1.997 ppm and for GlcNAc **2** at δ = 2.067, 2.069, and 2.071 ppm, whereas in the spectrum of the oligosaccharide two signals are observed in the *N*-acetyl region at δ = 2.050 and 2.047 ppm, in the anomeric ratio (α:β = 2:1), belonging to GlcNAc **2**. This indicates that in the glycopeptide a heterogeneity exists which is restricted to the peptide moiety. As reported earlier (15–17), the chemical shifts of the structural reporter groups of GlcNAc **1** and **2** are sensitive to the extension of asparagine with other amino acids. Amino acid analysis of the SBA glycopeptide showed that besides Asn the following amino acids are present in the peptide backbone: Thr, Ser, Glu/Gln, Pro, Gly, Ala, Val, Ile, Leu, Phe, and Lys. Per mol of asparagine a total amount of 1.75 mol of these amino acids is found. The differences in the chemical shifts of the *N*-acetyl signals of GlcNAc **1** and of those of GlcNAc **2** may therefore be ascribed to the presence of small amounts of

a variety of contaminating glycopeptides, each with a different peptide moiety but all with the same carbohydrate unit.

The application of ^1H NMR spectroscopy for the structure elucidation of the carbohydrate unit of SBA leads to the conclusion that this part of the compound is homogeneous. It rules out the earlier suggested possibility of a structure with two branches only, whereas the second 3,6-disubstituted mannose residue turns out to be Man **4'**, $\alpha(1 \rightarrow 6)$-linked to the first one (Man **3**). These unambiguous assignments can be made thanks to the possibility of ^1H NMR spectroscopy to discriminate between Man **4** and **4'**.

The revised structure of the carbohydrate moiety of SBA is identical in all respects with that of the *N*-glycosidically linked $Man_9GlcNAc_2$ oligosaccharides, found in other glycoproteins such as porcine thyroglobulin (18) and bovine lactotransferrin (14) and in the glycopeptide from Chinese hamster ovary cell membranes (19). *N*-Glycosidically linked high mannose oligosaccharide units having the same branching points, although with a different number of $\alpha(1 \rightarrow 2)$-linked mannose residues, have been found in many other glycoproteins, including *Aspergillus* taka-amylase (20), ovalbumin (21, 22), yeast mannans (23), human IgM (24) and other immunoglobulins (25), ribonuclease B (26), and horse pancreatic ribonuclease.[3] It now appears that the same heptasaccharide,



with different degrees of substitution by $\alpha(1 \rightarrow 2)$-linked mannose residues, is of general occurrence in a wide variety of organisms from yeast to higher animals.

Acknowledgments—Thanks are due to Dr. W. E. Hull (Bruker Analytische Messtechnik, Rheinstetten, FRG) for recording the 500-MHz ^1H NMR spectra and to Dr. T. Muramatsu (Kobe, Japan) for providing us with endo-β-*N*-acetylglucosaminidase H.

REFERENCES

1. Montreuil, J. (1980) *Adv. Carbohydr. Chem. Biochem.* **37,** 157–223
2. Kornfeld, R., and Kornfeld, S. (1980) in *The Biochemistry of Glycoproteins and Proteoglycans* (Lennarz, W. J., ed) pp. 1–34, Plenum Press, New York
3. Sharon, N., and Lis, H. (1981) in *The Proteins* (Neurath, H., and Hill, R. L., eds) 3rd Ed, Vol. V, Academic Press, New York, in press
4. Lis, H., and Sharon, N. (1978) *J. Biol. Chem.* **253,** 3468–3476
5. Schachter, H., and Roseman, S. (1980) in *The Biochemistry of Glycoproteins and Proteoglycans* (Lennarz, W. J., ed) pp. 85–160, Plenum Press, New York
6. Vliegenthart, J. F. G., Van Halbeek, H., and Dorland, L. (1980) in *IUPAC 27th International Congress of Pure and Applied Chemistry, Helsinki 1979* (Varmavuori, A., ed) pp. 253–262, Pergamon Press, Oxford
7. Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Strecker, G., and Montreuil, J. (1980) *FEBS Proc. Meet.* Abstract 182
8. Narasimhan, S., Harpaz, N., Longmore, G., Carver, J. P., Grey, A. A., and Schachter, H. (1980) *J. Biol. Chem.* **255,** 4876–4884
9. Cohen, R. E., and Ballou, C. E. (1980) *Biochemistry* **19,** 4345–4358
10. Van Halbeek, H., Dorland, L., Veldink, G. A., Vliegenthart, J. F. G., Michalski, J.-C., Montreuil, J., Strecker, G., and Hull, W. E. (1980) *FEBS Lett.* **121,** 65–70
11. Van Halbeek, H., Dorland, L., Veldink, G. A., Vliegenthart, J. F. G., Strecker, G., Michalski, J.-C., Montreuil, J., and Hull, W. E. (1980) *FEBS Lett.* **121,** 71–77
12. Vliegenthart, J. F. G., Van Halbeek, H., and Dorland, L. (1981) *Pure Appl. Chem.* **53,** 45–77
13. Ernst, R. R. (1966) *Adv. Magn. Res.* **2,** 1–135
14. Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Spik, G., Chéron, A., and Montreuil, J. (1981) *Biochim. Biophys. Acta* **675,** 293–296
15. 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
16. Schmid, K., Binette, J. P., Dorland, L., Vliegenthart, J. F. G., Fournet, B., and Montreuil, J. (1979) *Biochim. Biophys. Acta* **581,** 356–359
17. 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
18. Ito, S., Yamashita, K., Spiro, R. G., and Kobata, A. (1977) *J. Biochem. (Tokyo)* **81,** 1621–1631
19. Li, E., and Kornfeld, S. (1979) *J. Biol. Chem.* **254,** 1600–1605
20. Minobe, S., Nakajima, H., Itoh, N., Funakoshi, I., and Yamashina, I. (1979) *J. Biochem. (Tokyo)* **86,** 1851–1854
21. Tai, T., Yamashita, K., Setsuko, I., and Kobata, A. (1977) *J. Biol. Chem.* **252,** 6687–6694
22. Conchie, J., and Strachan, I. (1978) *Carbohydr. Res.* **63,** 193–213
23. Cohen, R. E., Ballou, L., and Ballou, C. E. (1980) *J. Biol. Chem.* **255,** 7700–7707
24. Chapman, A., and Kornfeld, R. (1979) *J. Biol. Chem.* **254,** 816–828
25. Brenckle, R., and Kornfeld, R. (1980) *Arch. Biochem. Biophys.* **201,** 160–173
26. Liang, C.-J., Yamashita, K., and Kobata, A. (1980) *J. Biochem. (Tokyo)* **88,** 51–58

[3] B. L. Schut, personal communication.