

Determination of the Structure of Four Glycopeptides from Hen Ovalbumin Using 360-MHz Proton Magnetic Resonance Spectroscopy[†]

Jeremy P. Carver,* Arthur A. Grey, Francoise M. Winnik, John Hakimi, Costante Ceccarini,[‡] and Paul H. Atkinson

ABSTRACT: Four glycopeptides have been purified by Dowex and Bio-Gel P2 chromatography from Pronase digests of hen ovalbumin. The high-resolution proton magnetic resonance spectra of these glycopeptides and various products of their enzymatic digestion have been obtained at 360 MHz. By use of information derived from the spectra of a number of model

compounds, an unambiguous assignment of all C1-H and Man C2-H resonances in the spectra can be made. On this basis structures are proposed for the four glycopeptides which are identical with those structures previously deduced from destructive chemical methods.

As carbohydrate structures become implicated in biological recognition phenomena, so the need for rapid, nondestructive, and sensitive techniques for their characterization becomes essential. We, along with others, have been investigating the potential of high-field proton magnetic resonance spectroscopy for this purpose (Carver & Grey, 1981; Montreuil & Vliegenthart, 1979). In Atkinson et al. (1981), we show how NMR at 360 MHz can be used effectively in monitoring glycopeptide purification procedures. It becomes clear from that study that ¹H NMR provides a criterion of purity not previously available to glycoprotein biochemistry. In this paper we show how primary structure information can be deduced from the NMR spectra of the four homogeneous glycopeptides from ovalbumin, isolated by the methods previously described (Atkinson et al., 1981).

Cohen & Ballou (1980) and Van Halbeek et al. (1980) have interpreted the NMR spectra of several high mannose glycopeptides and oligosaccharides obtained at 40 and 27 °C, respectively. The assignments reported below for the corresponding compounds [glycopeptides E3 and D3 from ovalbumin; see Atkinson et al. (1981) for nomenclature] are essentially in agreement when the effects of temperature are taken into account.

Experimental Procedures

Materials

The ovalbumin glycopeptides E3, D3, C3B, and A3 were prepared, carbohydrate composition determined, and NMR spectra obtained as described in Atkinson et al. (1981).

Endoglycosidase H. Endoglycosidase H was purified from *Streptomyces griseus* by the method of Tarentino & Maley (1974). Glycopeptides (3–10 μmol) were incubated at 37 °C for 15 h with 0.009 unit of enzyme in a final volume of 400–1000 μL. The oligosaccharide products were recovered

as described in Atkinson et al. (1981).

β-Galactosidase. β-Galactosidase was purified from Jack bean meal by the method of Snaith & Levvy (1968); 1.82 μmol of A3 was incubated at 37 °C for 18 h with 11 units of enzyme in a final volume of 2.35 mL of 0.05 M sodium citrate-phosphate buffer, pH 3.5. The glycopeptide was recovered as described in Atkinson et al. (1981).

β-N-Acetylglucosaminidase. β-N-Acetylglucosaminidase (fraction B3A) was purified from octopus (C. Ceccarini, unpublished results). The enzyme was free from contamination with α-fucosidase, β-galactosidase, and α- or β-mannosidase, as judged by assays with the appropriate *p*-nitrophenyl derivatives. Narrowly pooled C3B (20 mg) was incubated at 37 °C with 10 units of enzyme in a final volume of 0.85 mL of 0.07 M citrate-phosphate buffer. An additional 10.8 units of enzyme was then added and the sample further incubated at 37 °C for 22.5 h. At this point a third addition of 10.8 units was made and the sample incubated at 37 °C for 4 days. The products were isolated by chromatography on G-25 and Bio-Gel P2 as described in Atkinson et al. (1981).

α-Mannosidase. α-Mannosidase was prepared from Jack bean meal by the method of Snaith & Levvy (1968). Complete removal of the α-Man residues in E3 and D3 was obtained with freshly prepared enzyme. Five Man residues were removed when 10.2 mg of D3 was digested with 3.6 units of α-mannosidase for 21 h at 37 °C. Similarly, four residues were removed when 9.4 mg of E3 was digested with 3.6 units for 21 h. Corresponding oligosaccharides, prepared as above, yielded a disaccharide (see Results) when 6 μmol was incubated at 37 °C for 40 h with 20 units of freshly prepared α-mannosidase in a final volume of 1.15 mL. The products were separated on Bio-Gel P2 according to the procedures described in Atkinson et al. (1981).

Methods

Nuclear Magnetic Resonance. The various glycopeptides and oligosaccharides were prepared for examination by 360 MHz ¹H NMR spectroscopy according to the methods described in Atkinson et al. (1981). In brief, the samples were dissolved in H₂O, passed through a column of Chelex (Bio-Rad), lyophilized several times from 99.7% D₂O, and finally dissolved in 99.96% D₂O. Depending on the amounts of material available, the final volume was either 100 μL in a cylindrical microcell (Wilmad Glass Co., Inc.) or 400 μL in a 5-mm NMR tube. Chemical shifts are quoted relative to an internal standard of acetone (2.225 ppm relative to internal 4,4-dimethyl-4-silapentane-1-sulfonate). Samples were gen-

[†] From the Department of Medical Genetics and Toronto Biomedical NMR Centre, University of Toronto, Toronto, Ontario, Canada M5S 1A8 (J.P.C., A.A.G., and F.M.W.), Department of Developmental Biology and Cancer, Albert Einstein College of Medicine, Bronx, New York 10461 (J.H. and P.H.A.), and Department of Biology, Hunter College, New York, New York 10021 (C.C.). Received January 6, 1981. This work was supported by grants from the Medical Research Council of Canada (MT-3732 and MA-6499), the National Cancer Institute of Canada, and the National Institutes of Health (CA13402, GM24071, and CA13330). P.H.A. is an Established Investigator of the American Heart Association and J.H. is a postdoctoral fellow of the Leukemia Society of America.

[‡] Present address: Istituto di Embriologia Molecolare, Arco Felici, Napoli, Italy.

Table I: GLC-Deduced Composition of Ovalbumin Glycopeptides and Their Oligosaccharides Derived from Endoglycosidase H

sample	glycopeptide or oligosaccharide	mannose (nmol)	galactose (nmol)	GlcNAc (nmol)
A3	glycopeptide	98.3 ± 5.9 (5.0) ^a	17.7 ± 1.8 (0.9)	100.2 ± 6.9 (5.1)
	oligosaccharide	24.5 ± 3.4 (4.8)	5.3 ± 0.9 (1.0)	21.0 ± 3.5 (4.1)
B3 ^b	glycopeptide	176.3 ± 6.1 (4.3)	27.1 ± 4.0 (0.6)	210.9 ± 10.8 (5.1)
	oligosaccharide	nd	nd	nd
C3	glycopeptide	150.6 ± 5.3 (5.2)		110.3 ± 5.3 (3.8)
	oligosaccharide	58.7 ± 5.8 (5.2)		32.2 ± 2.5 (2.8)
D3	glycopeptide	24.7 ± 2.6 (6.4)		6.1 ± 1.2 (1.6)
	oligosaccharide	65.7 ± 7.7 (5.8)		12.8 ± 1.5 (1.2)
E3	glycopeptide	43.6 ± 3.2 (5.2)		14.9 ± 1.5 (1.8)
	oligosaccharide	61.3 ± 1.8 (5.1)		11.1 ± 1.1 (0.9)

^a Values in parentheses are normalized as described in the text; nd, not determined. GlcNAc to Asn ratios were determined by amino acid analysis (Atkinson et al., 1981). Expressing the overall composition relative to Asn yields 4.6:0.84:5.8:1.0, 5.1:0:4.3:1.0, 6.0:0:1.83:1.0, and 4.7:0:1.6:1.0 for the Man:Gal:GlcNAc:Asn ratios of A3, C3B, D3, and E3, respectively. ^b Glycopeptide B3 is not a homogeneous preparation and is therefore not discussed further in the text.

erally examined at two temperatures, room temperature and 60–90 °C, in order to move the residual HDO resonance so that all the resonances in the 5.5–4.0-ppm range could be observed. All spectra were collected in the Fourier transform mode. In general, the spectrometer conditions were 4-kHz sweep width, 4-s cycle time, 16K data points, and 16–512 scans. The nuclear Overhauser enhancement (NOE) experiments were performed, as described by Richarz & Wüthrich (1978), by J.-R. Brisson as part of his Ph.D. research and will be described in detail elsewhere.

Permethylation. Permethylation analysis was performed according to the procedure of Tai et al. (1975) with approximately 0.5 μmol of oligosaccharide derived by endoglycosidase H treatment of glycopeptides E3 and D3. For C3B, approximately 0.5 μmol of glycopeptide was used, and the analysis was performed according to the method of Hakomori (1964) as described in Narasimhan et al. (1980).

Results and Discussion

Hexose Composition. The anomeric hydrogen resonances of hexoses are expected to fall within a limited region of the NMR spectrum, from about 5.5 to 4.3 ppm. With a few exceptions, this region can be further divided according to the anomeric configuration of the glycosidic hydrogen: from about 5.5 to 4.9 ppm for α anomers and from about 4.9 to 4.3 ppm for β anomers.

However, without additional information, a given C1–H chemical shift and coupling constant can be interpreted as arising from any one of several different hexoses. Thus, recourse to other analytical methods is necessary in order to establish the hexose composition of a glycopeptide. In this paper we have used GLC to establish qualitatively which hexoses are present (Table I). Since under certain circumstances GLC underestimates the amount of GlcNAc present, we have not used these data to establish the quantity of each hexose in the glycopeptides. However, once it is known from the GLC analysis which hexoses are present in the glyco-

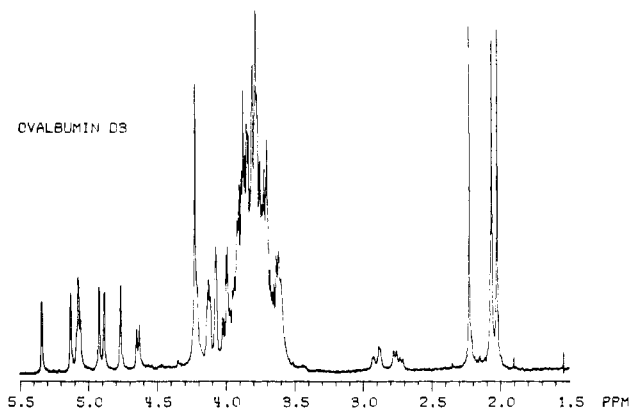


FIGURE 1: 360-MHz NMR spectrum of glycopeptide D3 at 78 °C.

peptide, the resonances of the anomeric hydrogens can be associated with one of the hexoses, in most cases. The quantitative composition can then be obtained by integration of the NMR spectrum. Results for individual glycopeptides are discussed in subsequent sections.

Proton Magnetic Resonance Spectra. A typical 360-MHz ¹H NMR spectrum of a high-mannose glycopeptide (D3) is shown in Figure 1. The anomeric hydrogen resonances are found from 4.5 to 5.5 ppm; the remaining hexose ring hydrogen resonances are clustered in the region 3.5–4.2 ppm. Additional resonances are found at about 2.7 ppm arising from the Asn β-methylene group while those at approximately 2 ppm are due to the N-acetyl groups of GlcNAc residues. The internal acetone reference gives the resonance at 2.225 ppm.

In general, the chemical shifts of the anomeric hydrogens of the hexoses reflect their linkage and substitution. This is particularly true for α-Man (Gorin et al., 1968, 1969) although less so for β-GlcNAc and β-Gal (Carver & Grey, 1981). Because of this sensitivity, it is possible in many cases to deduce the linkages present in a glycopeptide from the anomeric hydrogen chemical shift and coupling constants. When Man C2–H chemical shifts are also considered, it is possible to deduce the complete structure (Carver & Grey, 1981; Dorland et al., 1977a).

In Figure 2 the region of resonances from the anomeric hydrogens of the four glycopeptides are shown. The spectra obtained at high temperature have been used because this moves the resonance from residual HDO to a position outside the region of glycopeptide anomeric hydrogen resonances. Throughout the following, chemical shifts will refer to the spectra obtained at high temperatures (as shown in the figures) unless specifically stated otherwise. Chemical shift values for both temperatures are given in Table II since many literature values are quoted only at room temperature.

A. Structure of E3. Part of the spectrum of E3 at 76 °C is shown in Figure 2. The anomeric hydrogen resonances from 5.124 to 4.636 ppm occur with intensity ratios 2:1:1:1:1:1, giving a total of seven anomeric hydrogens. The composition of E3 from GLC analysis corresponds to 5.2 Man and 1.8 GlcNAc residues, assuming seven hexoses per molecule (based on the NMR integration results, molecular weight, and molar ratios to Asn; Table I). Thus, the seven anomeric hydrogen resonances observed must be distributed among five Man and two GlcNAc residues. The presence of the latter is confirmed by the appearance of two singlets arising from N-acetyl groups (2.063 and 2.021 ppm, each with an intensity of three hydrogens). The two 8-Hz doublets at 5.068 and 4.636 ppm are therefore assigned to β-GlcNAc. The former chemical shift corresponds to that found for the model compound, β-GlcNAc(1,4)Asn [C1–H, 5.090 ppm, doublet, J = 8–9 Hz;

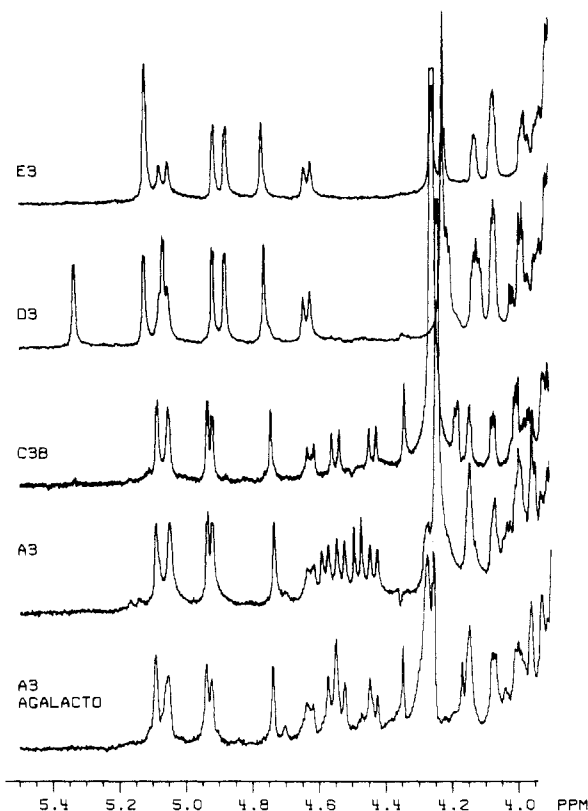


FIGURE 2: Region of the 360-MHz NMR spectrum containing the anomeric and C2 hydrogen resonances for the glycopeptides at $77 \pm 1^\circ\text{C}$.

NAc, 2.020 ppm, singlet; 50°C ; Tanaka & Yamashina (1973); Dorland et al. (1977b)].

The cleavage of E3 by endoglycosidase H suggests the presence of a chitobiosyl core structure (Tarentino & Maley, 1974; Tai et al., 1975). The resonance at 4.636 ppm is consistent with the chemical shift of the β -GlcNAc(1,4) of such a structure. After endoglycosidase H treatment, β -GlcNAc-(1,4)Asn (identified by NMR) is released from the glycopeptide converting the β -GlcNAc(1,4) in the core to a reducing terminal hexose. The loss of resonances at 5.068 and 4.636 ppm (see Figure 3) is consistent with this reaction scheme, as is the appearance of anomeric resonances at 5.244 and 4.742 ppm, corresponding to the α and β forms of the reducing GlcNAc. The latter resonances appear as singlets due to deuteration at C-2 (Salo et al., 1976). The *N*-acetyl resonances at 2.063 and 2.021 ppm have been replaced by a single resonance at 2.051 ppm with a concomitant reduction in intensity by three hydrogens. The GLC analysis of the oligosaccharide composition (Table I) also shows the loss of one GlcNAc residue.

One additional effect is that the resonance at 5.124 ppm in the glycopeptides is split into two in the corresponding oligosaccharide (Figures 2 and 3). These two components integrate to a single hydrogen. For this reason and others to be discussed elsewhere, we believe that these resonances belong to the α and β forms of the oligosaccharide.

The C1-H resonances of the five Man residues have characteristic coupling constants ($J_{1,2}$) of less than 2 Hz (Kammerling et al., 1975) and are found at 5.124 (2 H), 4.918, 4.880, and 4.772 ppm. Decoupling these resonances enables the chemical shifts of the corresponding C2 hydrogens to be determined. Actual linkages can be assigned by consideration of the individual C1-H and C2-H chemical shifts (4.5–5.5 and 4.0–4.3 ppm, respectively). Such an analysis together with the results of enzyme treatment permits the structure to be deduced.

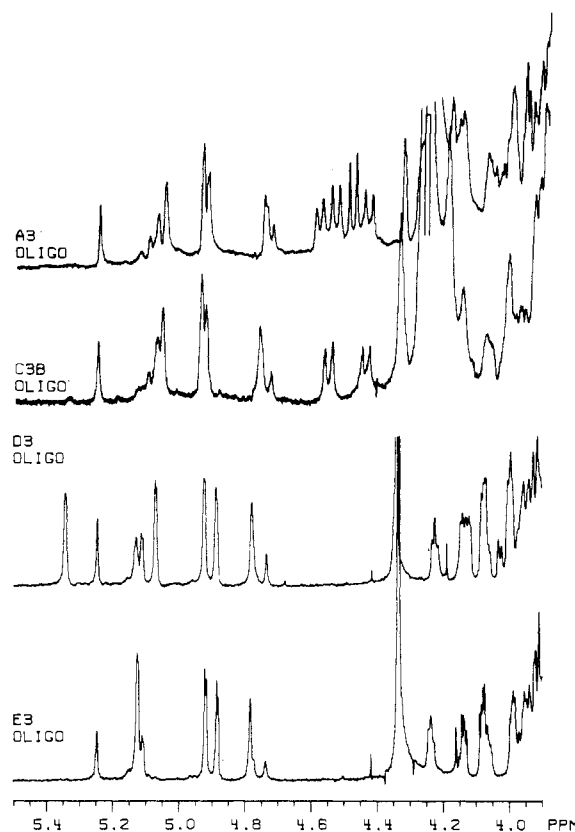


FIGURE 3: Region of the 360-MHz NMR spectrum containing the anomeric and C2 hydrogen resonances for the oligosaccharides obtained by endoglycosidase H digestion of the corresponding glycopeptides. The spectra were obtained at $77 \pm 1^\circ\text{C}$.

When the product from endoglycosidase H treatment of E3 was exhaustively digested with α -mannosidase, four out of the five Man residues were released, leaving β -Man(1,4)GlcNAc as the final product. Thus, the reducing terminal structure must be β -Man(1,4) β -GlcNAc(1,4) β -GlcNAc(1,4)Asn. The results of the methylation analysis (Table III) indicate that E3 contains three terminal and two 3,6-disubstituted Man as well as two 4-substituted GlcNAc residues. A comparison of the room temperature chemical shifts of the anomeric and C2 hydrogens of E3 (Table II) with the general chemical shift ranges given in Carver & Grey (1981) suggests that of the five Man residues there are two terminal α (1,3), one terminal α (1,6), one substituted α (1,6), and one substituted β (1,4) linked unit. Thus, the β -Man(1,4) and one of the α -Man(1,6) residues are 3,6-disubstituted. Combining these results, one may write the structure of E3 as that shown as I in Table IV. This structure is identical with that found by chemical methods (Tai et al., 1975).

B. Structure of D3. Part of the spectrum of D3 at 78°C is shown in Figure 2. The ratio of intensities for the resonances in the region from 5.336 to 4.638 ppm is 1:1:2:1:1:1:1, giving a total of eight anomeric hydrogens by integration. The GLC analysis (Table I) yields 6.4 Man and 1.6 GlcNAc residues on the assumption of a total of eight hexoses per molecule. The underestimation of the GlcNAc content by this method appears to arise from incomplete hydrolysis of the chitobiosyl core, since the GLC composition of the corresponding endoglycosidase H derived oligosaccharide is in agreement with the ^1H NMR data. When the same reasoning as for E3 is used, it can be concluded that D3 contains six Man and two β -GlcNAc residues. As before, the presence of the latter two residues is confirmed by the singlets at 2.063 and 2.022 ppm with intensities corresponding each to three hydrogens.

Table II: C1-H/C2-H Chemical Shift Pairs for Ovalbumin Glycopeptides at 24 and 77 °C^{c,d}

		E3	D3	C3B	A3	A3 agalacto
mannosyl ^d						
β-4i	HT	4.772/4.224	4.765/4.215	4.745/4.183	4.735/4.143	4.737/4.146
	RT	4.8 ^a /4.255	4.772/4.234	4.8 ^a /4.188	4.8 ^a /4.157	a/4.158
α-3i,t	HT	5.124/4.074	5.336/4.119	5.052/4.239	5.049/4.272	5.057/a
	RT	5.092/4.074	5.348/4.114	5.038/4.255	5.042/4.290	5.046/4.290
α-6i	HT	4.880/4.131	4.883/4.126	4.923/4.142	4.920/4.143	4.921/4.136
	RT	4.873/4.145	4.871/4.146	4.918/4.162	4.921/4.157	4.915/4.158
α-2t	HT		5.069/4.074			
	RT		5.053/4.063			
α-3t	HT	5.124/4.074	5.126/4.074	5.089/4.071	5.090/4.068	5.090/4.073
	RT	5.092/4.067	5.089/4.063	5.055/4.061	5.042/4.058	5.046/4.058
α-6t	HT	4.918/3.981	4.920/3.985	4.935/4.000	4.935/3.998	4.936/4.004
	RT	4.908/3.983	4.909/3.988	4.929/4.005	4.927/4.005	4.926/4.002
others						
βGN(1,2)	HT			4.552 (8.2) ^c	4.534 (8.2)	4.534 (9.0)
	RT			4.532 (8.2)	4.520 (8.3)	4.516 (8.3)
βGN(1,4)	HT				4.582 (7.4)	4.559 (9.2)
	RT				4.535 (7.8)	4.516 (8.3)
bis GN	HT			4.440 (8.3)	4.430 (8.2)	4.435 (7.7)
	RT			4.410 (8.3)	4.413 (8.3)	4.409 (8.2)
core GN	HT	4.636 (7.6)	4.638 (7.6)	4.627 (7.4)	4.625 (7.4)	4.628 (7.5)
	RT	4.603 (8.0)	4.601 (7.4)	4.588 (7.2)	4.590 (7.8)	4.591 (7.5)
Asn GN	HT	5.068 (9.3)	5.068 (9.0)	5.07 (b)	5.07 (b)	5.07 (b)
	RT	5.073 (10)	5.07 (b)	5.07 (b)	5.07 (b)	5.069 (10)
Gal	HT				4.483 (7.7)	
	RT				4.471 (7.3)	
N-acetyl region						
arm GN	HT			2.059	2.060	2.065
	RT			2.049	2.049	2.049
arm GN	HT				2.072	2.080
	RT				2.079	2.083
bis GN	HT			2.069	2.060	2.065
	RT			2.066	2.063	2.063
core GN	HT	2.063	2.063	2.060	2.060	2.065
	RT	2.061	2.059	2.057	2.058	2.058
Asn GN	HT	2.021	2.022	2.021	2.020	2.020
	RT	2.010	2.011	2.009	2.010	2.010

^a Resonance obscured by the residual HDO signal. ^b Resonance obscured by other signals. ^c Numbers in parentheses are the $J_{1,2}$ coupling constants in Hz. ^d The number refers to the linkage; i indicates internal (i.e., substituted) and t indicates terminal (i.e., not substituted). Thus, 2t means a terminal α-Man(1,2) residue. HT = 75–78 °C, RT = 23–25 °C. bis = bisecting; see text.

Table III: Methylation Analysis of Glycopeptides and Oligosaccharides

ovalbumin compd ^{a,d}	partially O-methylated acetate derivatives (molar ratios) ^b								
	neutral alditols							2- <i>N</i> -methylaceta- midohexitols ^c	
	2,3,4,6- tetra- <i>O</i> - methyl	3,4,6-tri- <i>O</i> -methyl	2,4,6-tri- <i>O</i> -methyl	2,3,6-tri- <i>O</i> -methyl	2,4-di- <i>O</i> -methyl	3,6-di- <i>O</i> - methyl	2- <i>O</i> - methyl		
	3,4,6-tri- <i>O</i> -methyl	2,4,6-tri- <i>O</i> -methyl	2,3,6-tri- <i>O</i> -methyl	2,4-di- <i>O</i> -methyl	3,6-di- <i>O</i> - methyl	2- <i>O</i> - methyl	3,4,6-tri- <i>O</i> -methyl	3,6-di- <i>O</i> -methyl	
E3	+	0	0	0	+	0	0	nd	0
D3	2.7	1.0	0	0	0.8	0	0	0	0.6
C3B	1.3	1.0	0	0	0.9	0	0.7	1.0	1.1

^a Glycopeptides were permethylated and prepared for analysis as described in the text. Quantitation of the partially methylated derivatives for E3 and D3 proved difficult: + indicates present but not quantitated; nd means not determined. ^b These derivatives were identified by their characteristic fragmentation patterns; reduction was carried out with sodium borohydride. Molar ratios are expressed relative to the 3,4,6-tri-O-methyl neutral alditol. ^c For C3B, only the relative ratio of the two 2-(N-methylacetamido)hexitol derivatives was determined. ^d E3 and D3 were examined as endoglycosidase H derived oligosaccharides; C3B, as the intact glycopeptide.

The spectra of E3 and D3 are very similar. Removal of one Man residue from the D3 oligosaccharide with α-mannosidase gives a compound with a spectrum identical with that of the E3 oligosaccharide. Thus, D3 is α-Man-E3. The presence of this α-Man causes (i) a reduction in the intensity of the 5.124 ppm resonance (terminal α(1,3)Man) from two hydrogens in E3 to one hydrogen in D3 and (ii) the appearance of C1-H/C2-H pairs (5.336/4.119 ppm and 5.069/4.074 ppm; 77 °C; Figure 2) in the spectrum of D3 due to a α-Man(1,2)α-Man(1,3) fragment. The pair at 5.069/4.074 ppm must correspond to the terminal α-Man(1,2) and the pair at 5.336/4.119 ppm to the 2-O-substituted α-Man(1,3).

Although the two terminal α-Man(1,3) residues of E3 have equivalent C1-H/C2-H chemical shifts at 76 °C, this is not true at room temperature. Thus, from a comparison of the values for E3 and D3 at 23 °C, it can be concluded that it is the residue with anomeric and C2 hydrogen chemical shifts of 5.092/4.074 ppm that is substituted in going from E3 (Man₅) to D3 (Man₆). However, from the C1-H/C2-H chemical shifts alone, there is no way of assigning to which of the terminal α-Man(1,3) residues these resonances belong. Thus, the α-Man(1,2) in D3 could reside on either of the terminal α-Man(1,3) units of E3. Methylation analysis cannot distinguish between these possibilities either, although the

Table IV: Structures of the Ovalbumin Glycopeptides and Model Compounds

I, E3	$ \begin{array}{c} \text{Mana}(1,6) \\ \diagup \quad \diagdown \\ \text{Mana}(1,3) \quad \text{Mana}(1,6) \\ \quad \quad \quad \diagup \quad \diagdown \\ \quad \quad \quad \text{Man}\beta(1,4)\text{GlcNAc}\beta(1,4)\text{GlcNAc}\beta(1,)\text{Asn} \end{array} $
II, D3	$ \begin{array}{c} \text{Mana}(1,6) \\ \diagup \quad \diagdown \\ \text{Mana}(1,3) \quad \text{Mana}(1,6) \\ \quad \quad \quad \diagup \quad \diagdown \\ \quad \quad \quad \text{Man}\beta(1,4)\text{GlcNAc}\beta(1,4)\text{GlcNAc}\beta(1,)\text{Asn} \end{array} $
III, C3B	$ \begin{array}{c} \text{Mana}(1,2) \\ \diagup \quad \diagdown \\ \text{Mana}(1,6) \quad \text{Mana}(1,6) \\ \diagup \quad \diagdown \quad \diagup \quad \diagdown \\ \text{GlcNAc}\beta(1,2) \quad \text{GlcNAc}\beta(1,4) \quad \text{Man}\beta(1,4)\text{GlcNAc}\beta(1,4)\text{GlcNAc}\beta(1,)\text{Asn} \end{array} $
IV, A3	$ \begin{array}{c} \text{GlcNAc}\beta(1,2) \\ \diagup \quad \diagdown \\ \text{Mana}(1,6) \quad \text{Mana}(1,6) \\ \diagup \quad \diagdown \quad \diagup \quad \diagdown \\ \text{GlcNAc}\beta(1,2) \quad \text{GlcNAc}\beta(1,4) \quad \text{Man}\beta(1,4)\text{GlcNAc}\beta(1,4)\text{GlcNAc}\beta(1,)\text{Asn} \end{array} $
V	$ \begin{array}{c} \text{GlcNAc}\beta(1,2)\text{Mana}(1,6) \\ \diagup \quad \diagdown \\ \text{GlcNAc}\beta(1,4) \quad \text{Man}\beta(1,4)\text{GlcNAc}\beta(1,4)\text{GlcNAc}\beta(1,)\text{Asn} \end{array} $
VI	$ \begin{array}{c} \text{GlcNAc}\beta(1,2)\text{Mana}(1,3) \\ \diagup \quad \diagdown \\ \text{GlcNAc}\beta(1,4) \quad \text{Man}\beta(1,4)\text{GlcNAc}\beta(1,4)\text{GlcNAc}\beta(1,)\text{Asn} \end{array} $

results given in Table III are in complete agreement with the conclusions drawn from the NMR data. However, the terminal α -Man(1,3) residues are distinguishable in that one is linked to an α -Man(1,3) while the other is linked to a β -Man(1,4). If it were possible to show that the 2-substituted Man is close in space to one or other of the above Man residues, then the structure would be defined. Fortunately, an NMR technique exists which detects the proximity of two hydrogens in space regardless of where they are located in the chemical structure. This technique is called nuclear Overhauser enhancement (NOE; see Methods). When the anomeric hydrogen of the 2-substituted Man is irradiated, an NOE is seen on the C2-H of the β -Man(1,4) (J.-R. Brisson, unpublished results). Thus the correct structure is that shown as II in Table IV.

Tai et al. (1977) resolved this question for the corresponding preparation of D3 by acetolysis and found the same structure as that deduced from the NOE experiment. These authors also showed that this compound, in contrast to E3, was resistant to endoglycosidase D and concluded that the α -Man(1,2) substitution on the lower α -Man(1,3) arm was the cause. As expected, our preparation of D3 is also resistant to endoglycosidase D under conditions which cleave E3.

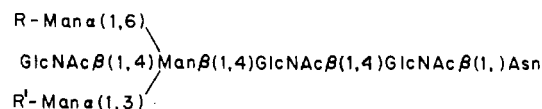
C. Structure of C3B. The region of anomeric hydrogen resonances in the spectrum of C3B (Figure 2) shows a pattern significantly different from that of the two previous glycopeptides. A total of nine hydrogens are found by integration of the resonances between 5.1 and 4.4 ppm. The composition of C3B from GLC analysis (Table I) corresponds to 5.2 Man and 3.8 GlcNAc residues, assuming nine hexoses per molecule. The anomeric resonances at 5.089, 5.052, 4.935, 4.923, and 4.745 ppm correspond to the five Man residues.

Three doublets (each corresponding to one hydrogen by integration) with coupling constants of approximately 8 Hz are clearly visible at 4.672, 4.522, and 4.440 ppm. The anomeric resonance of the ninth residue is buried in the

5.10–5.05-ppm region since this area corresponds to three hydrogens. Integration of the 2-ppm region indicates the presence of about 12 hydrogens which must arise from the four GlcNAc residues. Because of the 8-Hz coupling constants these must be β -linked. The hidden resonance at about 5.07 ppm can be assigned to the anomeric hydrogen resonance of a GlcNAc β (1,4)Asn (see above).

When C3B was treated with a mixture of exo- and endo- β -N-acetylglucosaminidase isolated from octopus (C. Cecarini, unpublished experiments), an oligosaccharide was obtained, the spectrum of which, when compared with that of the product from endoglycosidase H alone (Figure 3), lacked the 8-Hz doublets at 4.552 and 4.440 ppm (77 °C). No significant removal of Man residues was observed; thus, the GlcNAc residues giving rise to these doublets must be unsubstituted. From the methylation analysis (Table III) there are approximately equal amounts of terminal and 4-substituted GlcNAc residues. Although the quantitation of GlcNAc and terminal Man from methylation analysis is not in exact agreement with the ^1H NMR results, the former is more susceptible to error from the presence of contaminants and poor yields of the GlcNAc derivatives (Akhrem et al., 1979). As a result, although the methylation analysis is essential for the determination of the substitution pattern of the hexoses, ^1H NMR is still preferable for quantitation, as was the case for GLC (see above).

As with the previous compounds, the resonances at about 5.07 and 4.588 ppm (24 °C) can be assigned to the GlcNAc residues of the endoglycosidase H sensitive core chitobiosyl structure. The resonances at 2.057 and 2.009 ppm also arise from these 4-substituted residues (Table II). The other two GlcNAc residues with anomeric resonances (24 °C) at 4.410 and 4.532 ppm (*N*-acetyls: 2.066 and 2.049 ppm) are therefore terminal, on the basis of the methylation analysis and the effects of enzyme treatment. The 4.532-ppm resonance falls within the range found for terminal β -GlcNAc(1,2) residues; however, other linkages cannot be excluded with certainty on the basis of NMR alone. However, the 4.410-ppm resonance falls well outside of this region and closer to that found for β -GlcNAc(1,4) residues linked to the β -Man(1,4) of a trimannosyl core structure [i.e., a "bisecting GlcNAc"; see Carver & Grey (1981)]. A typical example is the structure shown as V in Table IV. The chemical shift found by Dorland et al. (1979) for the anomeric proton of the corresponding residue was 4.471 ppm. The results of the methylation analysis (Table III) show that there is a 3,4,6-trisubstituted Man, consistent with the presence of a bisecting GlcNAc on the β -Man of a trimannosyl core. Comparison of the observed chemical shifts for the Man residues (Table II) with the ranges in Carver & Grey (1981) leads to the conclusion that the linkages present are one substituted β (1,4), one substituted α (1,6), one terminal α (1,6), one substituted α (1,3), and one terminal α (1,2) or α (1,3). As we shall see later, the terminal α (1,3) assignment is the correct one. These conclusions are consistent with those from the methylation analysis which show the presence of 3,6-disubstituted, 2-substituted, and terminal Man residues as well as the trisubstituted Man already mentioned. Thus a partial structure may be written as



where R and R' are some combination of α -Man(1,6), α -Man(1,2 or 1,3), and β -GlcNAc(1,3 or 1,2). Since the remaining substituted Man residues are 2-substituted and 3,6-

disubstituted, the terminal GlcNAc and two Man residues must occupy these positions. Clearly, the 6-position must be occupied by the terminal α -Man(1,6). The other two terminal residues can be assigned as α -Man(1,3) and β -GlcNAc(1,2), since the alternate possibility [α -Man(1,2) and β -GlcNAc(1,3)] would predict the existence of resonances in quite different regions of the spectrum.

Comparison of the Man C1-H/C2-H chemical shifts further supports the existence of a bisected structure. In particular, the internal α -Man(1,3), α (1,6), and internal β -Man(1,4) chemical shifts (24 °C; 5.038/4.255, 4.918/4.162, and 4.80/4.188 ppm) are similar to those found for structure V (Table IV), i.e., 5.062/4.250, 5.004/4.151, and 4.697/4.184 ppm, respectively. When the NOE technique was applied to C3B, irradiation of the 2-*O*-substituted Man anomeric hydrogen showed an effect on the C2-H of the β -Man(1,4), in agreement with structure III, Table IV. The latter is identical with that found by Tai et al. (1977).

D. Structure of A3. The spectrum of A3 is very similar to that of C3B (Figure 2) with a few exceptions (see below). The entire anomeric region (5.1–4.4 ppm) corresponds to 11 hydrogens by integration. The composition of A3, from GLC analysis, is $\text{Man}_{5,0}\text{GlcNAc}_{5,1}\text{Gal}_{0,9}$ on the basis of the total of 11 hexoses established by the NMR quantitation. The presence of five GlcNAc residues is confirmed by the results of the integration of the region around 2 ppm. Thus, A3 differs from C3B in the addition of one GlcNAc and one Gal residue. The presence of a chitobiosyl core structure is manifested by the resonances at approximately 5.07 and 4.590 ppm and the effects thereon of endoglycosidase H treatment (Figure 3).

The chemical shifts of the Man resonances (C1-H/C2-H) of A3 are identical with those of C3B except for the C2-H of the internal α -Man(1,3) (see below). Together with the resonance at 4.413 ppm ($J = 8.3$ Hz, 25 °C), these indicate the existence of a "bisected" structure similar to C3B with, however, the additional Gal and GlcNAc linked to an α -Man(1,3). The manner in which these residues are linked can be deduced from a consideration of enzymatic and NMR data, described below.

The region of anomeric hydrogen resonances in the spectrum of A3 contains two additional 8-Hz doublets at 4.520 and 4.471 ppm (24 °C) when compared to that of C3B. These doublets arise from the anomeric hydrogens of the additional GlcNAc and Gal residues. When A3 is treated with β -galactosidase, one Gal residue is removed and the spectrum of the product (Figure 2) shows a loss of the 4.471-ppm doublet and a shift in the 4.535-ppm doublet resonance to 4.516 ppm (24 °C). The anomeric hydrogen resonances of β -GlcNAc(1,2) residues have been shown to shift by -0.02 ppm when a β (1,4)Gal substitution is removed (Carver & Grey, 1981). The perturbation in the chemical shift of the GlcNAc anomeric hydrogen resonance observed upon removal of Gal in A3 is -0.019 ppm, suggesting that the Gal was linked β (1,4) to the GlcNAc. Thus, it can be concluded that the difference between C3B and A3 is in the addition of a *N*-acetyllactosamine arm to the internal α (1,3) Man. The question remains as to where on the Man the arm is attached.

In their study of the structure of the ovotransferrin glycopeptide (see structure VI of Table IV), Dorland et al. (1979) found the internal α -Man(1,3) residue to have a C2-H chemical shift of 4.280 ppm. This is 0.03 ppm to low field of that found for the corresponding hydrogen (4.250 ppm) in the Sandhoff structure (V, Table IV). This increment arises from the additional β (1,4) substitution by GlcNAc in the former structure. A similar comparison between the chemical

shifts of the corresponding hydrogens in C3B and A3 reveals a downfield shift of 0.035 ppm at 24 °C (4.255 and 4.290 ppm, respectively). This observation suggests that the internal α -Man(1,3) of A3 has an additional 4-substitution by a β -GlcNAc compared to C3B. Thus, the extra lactosamine arm is linked to either the 2 or 4 position.

When the NOE technique was applied to the C2-H hydrogen resonance at 4.290 ppm (24 °C), i.e., that assigned to the internal α -Man(1,3), NOE effects were observed on the resonances assigned to the corresponding anomeric hydrogen at 5.042 ppm and to the anomeric hydrogen resonance at 4.520 ppm. Because the distance to the anomeric hydrogen of the β -GlcNAc(1,4) is too great, an NOE effect can only be observed to the anomeric hydrogen of the β -GlcNAc(1,2) substituent. Thus, the 4.520-ppm resonance can be assigned to such a residue and the 4.535-ppm resonance (24 °C) to the β -GlcNAc(1,4). Since the latter resonance is the one which shifts upon removal of the terminal Gal, the extra lactosamine arm of A3 must be linked to the 4 position of the α -Man(1,3). Thus, we can assign the structure shown as IV (in Table IV), which is identical with that found by Yamashita et al. (1978).

A separate NOE experiment in which the C1-H resonance belonging to the core β -Man(1,4) (3,4,6-trisubstituted) was irradiated shows effects on the corresponding C2-H resonance at 4.157 ppm as well as on the C2-H resonance of the 2,4-disubstituted α -Man(1,3) at 4.290 ppm. These results can only be interpreted by placing the 2,4-disubstituted α -Man(1,3) as shown in structure IV, Table IV. The observation of endoglycosidase D resistance for both A3 and the compound isolated by Yamashita et al. (1978) supports this conclusion.

It should be noted that the C1-H resonances of the arm GlcNAc residues of C3B, A3, and agalacto-A3 are somewhat anomalous. In particular, C3B and A3 have resonances at 4.532 and 4.535 ppm (24 °C) which one might argue should be assigned to the terminal β (1,2) residues present in both structures. However, as the NOE experiment unequivocally demonstrates, the former is a terminal β -GlcNAc(1,2) and the latter is a β -GlcNAc(1,4) substituted at position 4 by β -Gal. This clearly indicates the caution that must be exercised with GlcNAc residues in making these assignments simply by analogy. In fact, the terminal β -GlcNAc(1,2) of A3 has a C1-H resonance at 4.520 ppm (compared to 4.532 ppm in C3B). The difference in these chemical shifts for identically linked GlcNAc residues must reflect the different interactions that they experience from the surrounding residues and therefore contains information about the three-dimensional structure.

Conclusions

The NMR technique provides a complete determination of the primary structure of the ovalbumin glycopeptides when composition has been determined by an independent method, such as GLC, and suitable models are available for comparisons of chemical shifts. The structures deduced a priori by the NMR method are the same as those found by Kobata and colleagues (Tai et al., 1975, 1977; Yamashita et al., 1978) by chemical methods. In the above analysis it has been shown that two precautions must be taken in the interpretation of the NMR data of glycopeptides: i.e., the spectra of the unknown and the model must be compared at the same temperature and the identity of GlcNAc anomeric hydrogen chemical shifts cannot be taken as an indication of identical linkages.

The chemical shifts and assignments in Table II are the first to be reported for hybrid structures. These have been included as part of a data base of 75 compounds representing nine

classes of glycopeptide structure from which a standard set of NMR parameters for the determination of primary structure have been deduced (Carver & Grey, 1981).

Some information concerning the secondary structure of these complex molecules in solution may also be obtained by NMR by use of nuclear Overhauser effects. Previously, in their studies on the conformation of the blood group oligosaccharides, Lemieux et al. (1980) showed that the use of the NOE effect yielded important information on the three-dimensional structure. In the above we have used a similar approach to deduce sequence information which could not be derived otherwise except by acetolysis studies.

Acknowledgments

We gratefully acknowledge the use, by permission of Dr. Alan McLaughlin, of the 360-MHz spectrometer at Brookhaven National Laboratories under the auspices of the U.S. Department of Energy and the use, by permission of Dr. John Markley, of the 360-MHz spectrometer at Purdue University, under the auspices of NIH Division of Research Resources. The authors gratefully acknowledge the expert assistance of Alan Lee in recording the spectra at the Toronto Biomedical NMR Centre and thank Jean-Robert Brisson for performing the NOE experiments. We also wish to thank Joyce Tsang-Lee for her excellent technical assistance and Pamela Stanley for her constructive criticism of the manuscript. Finally, we acknowledge the help of Jim Etchison, who provided the first samples of ovalbumin glycopeptides.

References

- Akhrem, A. A., Avvakumov, G. V., Sidorova, I. V., & Strel'chyonok, O. A. (1979) *J. Chromatogr.* **180**, 69–80.
Atkinson, P. H., Grey, A. A., Carver, J. P., Hakimi, J., & Ceccarini, C. (1981) *Biochemistry* **20**, 3979–3986.
Carver, J. P., & Grey, A. A. (1981) *Biochemistry* (following paper in this issue).
Cohen, R. E., & Ballou, C. E. (1980) *Biochemistry* **19**, 4345–4358.
Dorland, L., Haverkamp, J., Schut, B. L., Vliegthart, J. F. G., Spik, G., Strecker, G., Fournet, B., & Montreuil, J. (1977a) *FEBS Lett.* **77**, 15–20.

- Dorland, L., Schut, B. L., Vliegthart, J. F. G., Strecker, G., Fournet, B., Spik, G., & Montreuil, J. (1977b) *Eur. J. Biochem.* **73**, 93–97.
Dorland, L., Haverkamp, J., Vliegthart, J. F. G., Spik, G., Fournet, B., & Montreuil, J. (1979) *Eur. J. Biochem.* **100**, 569–574.
Gorin, P. A. J., Mazurek, M., & Spencer, J. F. T. (1968) *Can. J. Chem.* **46**, 2305–2310.
Gorin, P. A. J., Spencer, J. F. T., & Bhattacharjee, S. S. (1969) *Can. J. Chem.* **47**, 1499–1505.
Hakomori, S. (1964) *J. Biochem. (Tokyo)* **55**, 205–208.
Kammerling, J. P., Gerwid, G. J., Vliegthart, J. F. G., & Clamp, J. R. (1975) *Biochem. J.* **151**, 491–495.
Lemieux, R. U., Bock, K., Delbaere, L. T. J., Koto, S., & Rao, V. S. (1980) *Can. J. Chem.* **58**, 631–653.
Montreuil, J., & Vliegthart, J. F. G. (1979) in *Glycoconjugate Research* (Gregory, J. D., & Jeanloz, R. W., Eds.) Vol. I, pp 35–78, Academic Press, New York.
Narasimhan, S., Harpaz, N., Longmore, G., Carver, J. P., Grey, A. A., & Schachter, H. (1980) *J. Biol. Chem.* **255**, 4876–4884.
Richarz, R., & Wüthrich, K. (1978) *J. Magn. Reson.* **30**, 147–150.
Salo, W. L., Hamari, M., & Hallcher, L. (1976) *Carbohydr. Res.* **50**, 287–291.
Snaith, S. M., & Levvy, G. A. (1968) *Biochem. J.* **110**, 663–670.
Tai, T., Yamashita, K., Ogata-Arakawa, M., Koide, N., Muramatsu, T., Iwashita, S., Inoue, Y., & Kobata, A. (1975) *J. Biol. Chem.* **250**, 8569–8575.
Tai, T., Yamashita, K., Ito, S., & Kobata, A. (1977) *J. Biol. Chem.* **252**, 6687–6694.
Tanaka, M., & Yamashina, I. (1973) *Carbohydr. Res.* **27**, 175–183.
Tarentino, A. L., & Maley, F. (1974) *J. Biol. Chem.* **249**, 811–817.
Van Halbeek, H., Dorland, L., Veldink, G. A., Vliegthart, J. F. G., Michalski, J.-C., Montreuil, J., Strecker, G., & Hull, W. E. (1980) *FEBS Lett.* **121**, 65–70.
Yamashita, K., Tachibana, Y., & Kobata, A. (1978) *J. Biol. Chem.* **253**, 3862–3869.