Carbohydrate Research, 98 (1981) 132–138
Elsevier Scientific Publishing Company, Amsterdam – Printed in The Netherlands

Note

^{13}C-N.m.r.-spectral study of D-galactopyranosyl and 2-acetamido-2-deoxy-D-galactopyranosyl glycopeptides relevant to glycoproteins

KILIAN DILL*,
Department of Chemistry, Clemson University, Clemson, SC 29631 (U.S.A.)
BERNARD FERRARI, JEAN M. LACOMBE, AND ANDRÉ A. PAVIA
Laboratoire de Chemie Bioorganique, Faculté de Science, 33, rue Louis Pasteur, Avignon (France)
(Received June 23rd 1981; accepted for publication, July 14th, 1981)

Oligosaccharide chains of glycoproteins are attached to the peptide backbone by various covalent linkages[1–3]. Some of the more common linkages are those in 2-*N*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-L-asparagine (β-GlcNAc→Asn), 3-*O*-(2-acetamido-2-deoxy-α-D-galactopyranosyl)-L-serine (α-GalNAc→Ser), 3-*O*-(2-acetamido-2-deoxy-α-D-galactopyranosyl)-L-threonine (α-GalNAc→Thr), and 5-(β-D-galactopyranosyloxy)-L-lysine (β-Gal→Hyl). Other, less-common linkage-pairs are 4-(L-arabinofuranosyloxy)-L-proline (Ara→Hyp), 3-*O*-D-mannopyranosyl-L-serine (Man→Ser), 3-*O*-D-mannopyranosyl-L-threonine (Man→Thr), and 3-*O*-D-galactopyranosyl-L-serine (Gal→Ser). These glycopeptide linkages exhibit different chemical reactivities, and are impossible to deduce without destruction of the glycoprotein[1–4]. Even so, both chemical and enzymic methods are subject to considerable uncertainty (see refs. 3–6, and references cited therein).

In the past few years, natural-abundance, carbon-13 nuclear magnetic resonance (^{13}C-n.m.r.) spectroscopy has been used to gain dynamic and structural information about the carbohydrate residues of large glycoproteins[7–9]. Application of this technique to the structural study of glycoproteins seems promising, but there are some temporary limitations. One of them is the lack, for relevant, model compounds, of ^{13}C-n.m.r. data needed in order to make specific assignments for the spectra of glycoproteins. Most notably missing from the literature are ^{13}C-n.m.r. data for glycosylated amino acids, although two papers have recently been published on this topic[10,11]. Any ^{13}C-n.m.r. data on glycosylated amino acids may eventually permit application of this technique to intact glycoproteins, in order to (*a*) gain knowledge about the type of carbohydrate–protein linkage present, (*b*) quantitate the average carbohydrate chain-length, and (*c*) possibly, even gain insight into protein–carbohydrate interactions[8].

*To whom correspondence should be addressed.

We now present ^{13}C-n.m.r. data for the important glycopeptides 3-*O*-(2-acetamido-2-deoxy-α-D-galactopyranosyl)-*N*-acetyl-L-serine (α-GalNAc→SerNAc), 3-*O*-(2-acetamido-2-deoxy-α-D-galactopyranosyl)-*N*-acetyl-L-threonine (α-GalNAc→ThrNAc), α-GalNAc→Thr, 3-*O*-α,β-D-galactopyranosyl-L-serine (α,β-Gal→Ser), and 3-*O*-α,β-D-galactopyranosyl-L-threonine (α,β-Gal→Thr). α-GalNAc→Ser,Thr is a common carbohydrate–protein linkage found in many glycoproteins[2], and Gal→Ser and Gal→Thr linkages have been detected in plant cell-walls and earthworm-cuticle collagen[2,12]. Unlike other glycosylated amino acids that have been studied by ^{13}C-n.m.r., spectroscopy, some of these model compounds exhibit a significant dependence of chemical shift on pH for some of the carbohydrate carbon atoms. Moreover, unusual α-pyranose-C-1 (α*p*1) and C^{γ}Thr chemical-shifts are observed for α,β-D-galactose linked to threonine.

EXPERIMENTAL

Materials. — α-GalNAc→SerNAc, α-GalNAc→ThrNAc, α-GalNAc→Thr, α,β-Gal→Ser, and α,β-Gal→Thr were synthesized as previously described[13,14]. The samples to be examined by n.m.r. spectroscopy were prepared by passing an aqueous solution of the glycopeptide through a short column of Chelex-100 (H^+) ion-exchange resin. The effluent was freeze-dried, and the residue dissolved in deionized, distilled water just before use. Adjustments in the pH of the sample were made with M NaOH or HCl, using a Radiometer PHM63, digital pH-meter.

Methods. — In the case of α,β-Gal→Ser, α,β-Gal→Thr, and α-GalNAc→Thr, identification of C^{α} and C^{β} was made on the basis of the pH-titration behavior[10,11]. For α-GalNAc→SerNAc and α-GalNAc→ThrNAc, C^{α} and C^{β} were identified not only by pH-dependence (when one was observed), but also by noise-modulated, off-resonance proton-decoupling and single-frequency, off-resonance proton-decoupling.

Carbon-13 n.m.r. spectra were recorded with a JEOL-FX90Q instrument operating at 21 kG in the F.t. mode, using quadrature detection. Samples (~1.5 mL) were contained in 10-mm tubes, with a 5-mm tube containing D_2O inserted concentrically to serve as a field-frequency lock. The probe temperature was maintained at ~25° for all samples. For ^{13}C excitation, 90° radio-frequency pulses of 29 μs were used, and the carrier frequency was set at ~90 p.p.m. downfield from the ^{13}C resonance of Me_4Si. A spectral window of 5.5 kHz was used for most samples. Fully proton-decoupled spectra were obtained when the noise-modulated, ^1H irradiation, having a bandwidth of 1.0 kHz, was centered ~4 p.p.m. downfield from Me_4Si. Spectra with ^{13}C–^1H coupling were obtained by using a proton-decoupling technique the reverse of that used for n.O.e. measurements[15].

Chemical shifts are given relative to a trace of internal 1,4-dioxane (added only when chemical shifts were determined), whose chemical shift was taken to be 67.86 p.p.m. downfield from Me_4Si.

RESULTS AND DISCUSSION

Fig. 1 shows the proton-decoupled, natural-abundance, ^{13}C-n.m.r. spectra of the aliphatic region of α-GalNAc→SerNAc and α-GalNAc→ThrNAc (except for

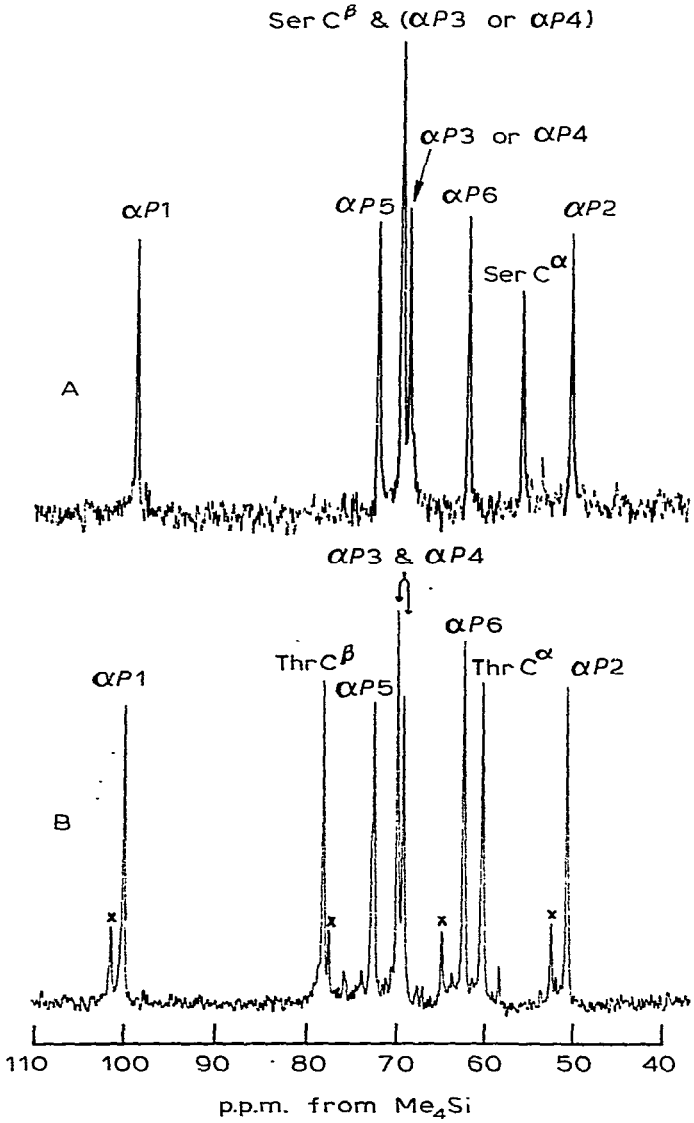

Fig. 1. Aliphatic carbon-atom region (except C^{γ} of threonine and -CH_3 of acetyl groups) of the proton-decoupled, natural-abundance, ^{13}C-n.m.r. spectra of α-GalNAc→SerNAc and α-GalNAc→ThrNAc. [Spectra were recorded with a spectral window of 5.50 kHz, and line broadening of 1.2 Hz was applied during the processing. *αp* refers to pyranose anomer in the Figure. The number following each *αp* designates the carbohydrate carbon atom. (A) 71mM α-GalNAc→SerNAc in H_2O, pH 6.8, at ~25°; 20,000 accumulations, with a recycle time of 2 s. (B) 133mM α-GalNAc→ThrNAc, pH 6.51; 10,000 accumulations, with a recycle time of 2 s.]

TABLE I

CARBON-13 N.M.R. CHEMICAL-SHIFT DATA[a] FOR THE β-GLYCOSYLAMINO ACIDS Gal→Ser AND Gal→Thr

Carbon atom	*β-Gal→Ser*[b]	*β-Gal→Thr*[c]
βp1	103.8 (0.5)	101.6
βp2	72.0	72.0
βp3	73.9	74.0
βp4	70.0	70.1
βp5	76.5	76.5
βp6	62.4	62.6
Ser C^{α}	56.2 (2.6)	
Ser C^{β}	69.2 (6.0)	
Thr C^{α}		61.0 (3.4)
Thr C^{β}		74.5 (4.6)
Thr C^{γ}		18.2

[a]Chemical shifts for these compounds are given at neutral pH (5.5 → 7.5). Estimated precision for the chemical shifts is ±0.05 p.p.m. The numbers in parentheses next to some of the chemical shifts indicate the pH-dependence (in p.p.m.) of these resonances when going from the cationic to the anionic form of the amino acid. All other resonances exhibit no, or <0.2 p.p.m., pH-dependence. The spectra of samples at low (<2.0) and high (>11.0) pH values were typically recorded for ~4,000 accumulations, using a 1-s recycle time, and a 4.00-kHz window. [b]125mM sample in H_2O, pH 6.26; 15,000 accumulations, with a recycle time of 2 s. [c]150mM sample in H_2O, pH 7.50; 20,000 accumulations, with a recycle time of 2 s.

TABLE II

CARBON-13 N.M.R. CHEMICAL-SHIFT DATA[a] FOR VARIOUS α-GLYCOSYLAMINO ACIDS

Carbon atom	*α-Gal→Ser*[b]	*α-GalNAc→SerNAc*[c]	*α-Gal→Thr*[d]	*α-GalNAc→ThrNAc*[e]	*α-GalNAc→Thr*[f]
αp1	100.3 (0.2)	99.0	101.2 (0.4)	100.1	100.4 (0.6)
αp2	69.7 (0.4)	51.2	70.0 (0.2)	51.3	51.3 (0.4)
αp3 & *αp4*	70.7 (0.5, 0.0)	69.9[g] (0.0), 69.2 (0.4)	70.7 (0.8, 0.2)	70.0 (0.2), 69.3 (0.7)	70.0 (0.0), 69.1 (0.7)
αp5	72.7 (0.4)	72.5	72.7	72.6	72.8 (0.2)
αp6	62.7	62.5	62.6	62.7	62.7
Ser C^{α}	56.3 (2.6)	54.6 (2.0)			
Ser C^{β}	67.7 (5.6)	69.9[g]			
Thr C^{α}			60.8 (2.7)	60.6 (2.1)	60.6 (2.6)
Thr C^{β}			75.6 (4.9)	78.2 (1.4)	76.0 (4.8)
Thr C^{γ}			19.9 (0.4)	19.4	19.5 (0.2)
Ac CH_3		23.4		23.7 (0.2), 23.4 (0.2)	23.6 (0.2)

[a]Chemical shifts for these compounds are given at neutral pH (5.5→7.5). Estimated precision for the chemical shifts is ±0.05 p.p.m. The numbers in parentheses next to some of the chemical shifts indicate the pH dependence (in p.p.m.) of these resonances when going from the cationic to the anionic form of the amino acid. All other resonances exhibit no, or less than 0.2 p.p.m., pH-dependence. See footnote *a* to Table I. [b]158mM sample in H_2O, pH 5.54; 15,000 accumulations, with a recycle time of 2 s. [c]Same as in Fig. 1A. [d]125mM sample in H_2O, pH 6.2; 25,000 accumulations, with a recycle time of 2 s. [e]Same as in Fig. 1B. [f]32mM sample in H_2O, pH 5.5; 20,000 accumulations, with a recycle time of 1 s. [g]Overlap of resonances, see Fig. 1.

C^{γ} of Thr, and the acetyl methyl groups). Tables I and II give the ^{13}C-n.m.r.-spectral data for the various derivatives of serine and threonine glycosylated with α-GalNAc, α-Gal, and β-Gal. The β-Gal glycosylated Ser and Thr were readily differentiated from the α-Gal and α-GalNAc model compounds on the basis of the coupling constants ($^1J_{CH}$) of the anomeric carbon atom. All β-Gal model compounds exhibited an anomeric $^1J_{CH}$ value of 158 Hz, whereas, for the α-Gal and α-GalNAc derivatives, the anomeric $^1J_{CH}$ value was 171 Hz. These data are in agreement with work previously published on coupling constants for α- and β-galactopyranose[8,16,17].

The assignments of the resonances to specific carbohydrate and amino acid carbon atoms were straightforward. All of the signals for anomeric carbon atoms occurred in the neighborhood of 100 p.p.m. downfield from Me_4Si, as expected for *O*-glycosylated amino acids[7-11]. The assignments of *αp*2–*αp*6 and *βp*2–*βp*6 of the carbohydrates were based on published ^{13}C-n.m.r. spectral data for glycoproteins[8], monosaccharides[18,19], and methyl glycosides[18,19]. The only carbohydrate carbon atoms that could not be assigned on a one-to-one basis were *αp*3 and *αp*4 of α-Gal and α-GalNAc (see Fig. 1 and Tables I and II). Most amino acid carbon atoms were readily assigned to specific resonances in the spectra, based on the literature values of chemical shifts for peptides containing Ser and Thr[20]. *O*-Glycosylation of serine and threonine results in a large, downfield shift (~8 p.p.m.) for C^{β}, as expected[10,11]. In some cases, assignments of C^{α} and C^{β} of Ser and Thr were based on the large pH-dependence of their chemical shifts when going from the cationic to the anionic form of the amino acid (see Tables I and II). The assignments for C^{β} of SerNAc and ThrNAc could not be unambiguously determined by this method; their assignments were based on various decoupling techniques (see Experimental section).

The glycopeptides discussed herein exhibit some interesting features not hitherto observed for other glycopeptides. There is little difference in the chemical shift of *αp*1 and *βp*1 of galactose linked to threonine. Contrary to this result, the chemical-shift difference between *αp*1 and *βp*1 of galactose which is either attached to serine or which has a glycosidic methyl group is 3.5 p.p.m. (see Tables I and II) and 4.4 p.p.m., respectively[18]. The unusual upfield-shift of the signal for the β-anomeric carbon atom of an *O*-β-glycosyl group attached to threonine had been observed by Pavia *et al.*[21], and has been rationalized as attributable to the *exo*-anomeric effect. The chemical shift for C^{γ} of β-Gal→Thr and α-Gal→Thr may prove to be a useful probe in determining the anomeric state of galactopyranose glycosylically attached to threonine. There is a chemical-shift difference of 1.7 p.p.m. between C^{γ} of β-Gal→Thr and α-Gal→Thr; this chemical-shift difference had also been observed by Pavia *et al.*[21] for other α- and β-glycosides of threonine.

Fig. 1 shows the ^{13}C-n.m.r. spectra of α-GalNAc→SerNAc and α-GalNAc→ThrNAc. We had previously shown that, as model compounds, the glycopeptides are most useful in the zwitterionic state[10,11]. The acetyl protecting groups on the α-amino nitrogen atom appear to affect the chemical shift of C^{α} and C^{β} only (see Table II for the chemical shifts of α-GalNAc→Ser,Thr). Therefore, the chemical shifts given for the carbohydrate carbon atoms of α-GalNAc→Ser,Thr may readily

be extrapolated to the spectra of glycoproteins. The chemical shifts of C^{α}, C^{β}, and C^{γ} of protected and unprotected threonine are given, and these can readily be used to extrapolate chemical shifts for unprotected α-GalNAc→Ser. The utility of our model compounds for assignments of resonances in the ^{13}C-n.m.r. spectra of glycoproteins is readily shown by the fact that chemical shifts for α-GalNAc→Thr (at neutral pH) compare favorably with the chemical shifts of β-Gal-(1→3)-α-GalNAc→Thr found in fish-antifreeze glycoproteins[8], except for the carbon atoms at, or near, the point of glycosylation (see later).

The coupling constants observed for the methyl carbon atoms of the acetyl groups attached to $\alpha p2$ of α-GalNAc→ThrNAc and α-GalNAc→SerNAc were both 128 Hz. This coupling constant has the exact value observed for GalNAc and derivatives of GalNAc (see ref. 8, and references cited therein), but is larger than the 117 Hz for α-GalNAc found in fish-antifreeze glycoproteins[8]. As mentioned[8], the smaller coupling-constants found for fish glycoproteins may have implications in regard to their structure–function relationship.

Some of the carbohydrate resonances of the glycopeptides exhibit significant, pH effects (>0.5 p.p.m.). These large effects have never been observed before for the glycopeptides[10,11]. Inspection of a model of α,β-Gal→Thr shed no light on this phenomenon. Moreover, these effects seem to be mainly localized in α-Gal→Ser, α-Gal→Thr, and unprotected α-GalNAc→Thr (especially carbon atom $\alpha p3$ or $\alpha p4$).

The data given herein have several important implications for the study of glycoproteins by ^{13}C-n.m.r. spectroscopy. (*i*) The anomeric region may readily be used to identify the anomeric carbon atoms of α,β-Gal and α,β-GalNAc glycosylated to serine in glycoproteins. However, care must be taken in attempting to identify α,β-Gal→Thr, as the chemical shifts of anomeric carbon atoms are approximately the same. The C^{γ} of Thr may be a better reporter group for the anomeric state of a galactopyranosyl group, but, in all practicality, this resonance, at ~19 p.p.m., will be in the region of the aliphatic carbon atoms of the protein (see spectra of proteins and glycoproteins, given in ref. 7). (*ii*) The β carbon atoms of serine and threonine may be used as reporter groups, in order to gain quantitative information as to the degree of glycosylation (due to a downfield shift of ~8 p.p.m.). (*iii*) For the most part, the chemical shifts of carbohydrate carbon atoms do not differ by more than 0.5 p.p.m. (as the pH is changed from <2.0 to >11.0); this suggests that peptide-bond formation will not change the chemical shifts of the majority of carbohydrate carbon atoms by more than 0.5 p.p.m.

Two factors must be taken into account when use is made of the data from model compounds to gain information about glycoproteins: (*i*) the chemical shifts of carbohydrate carbon atoms adjacent to a point of glycosylation will shift significantly[8], and (*ii*) protein folding may play a role in broadening[22] (and possibly shifting) the signals of carbon atoms of carbohydrates glycosylated to the peptide backbone of glycoproteins.

REFERENCES

1 E. F. WALBORG, JR. (Ed.), *Glycoproteins and Glycolipids in Disease Processes*, ACS, Washington, D.C., 1978, pp. 5–20.
2 N. SHARON, *Complex Carbohydrates, Their Chemistry, Biosynthesis, and Functions*, Addison-Wesley, Reading, Massachusetts, 1975, pp. 65–83.
3 R. D. MARSHALL AND A. NEUBERGER, *Adv. Carbohydr. Chem. Biochem.*, 25 (1970) 407–478.
4 A. NEUBERGER, A. GOTTSCHALK, R. D. MARSHALL, AND R. G. SPIRO, in A. GOTTSCHALK (Ed.), *Glycoproteins, Their Composition, Structure and Function*, Vol. 5A, Elsevier, Amsterdam, 1972, Chapt. 4, pp. 450–490.
5 G. F. SPRINGER, H. J. YANG, E. MBAWA, AND D. GROHLICH, *Naturwissenschaften*, 67 (1980) 473–474.
6 E. LISOWSKA AND M. DUK, *Carbohydr. Res.*, 79 (1980) 103–113.
7 K. DILL AND A. ALLERHAND, *J. Biol. Chem.*, 254 (1979) 4524–4531.
8 E. BERMAN, A. ALLERHAND, AND A. L. DEVRIES, *J. Biol. Chem.*, 255 (1980) 4407–4410.
9 E. BERMAN, D. E. WALTERS, AND A. ALLERHAND, *J. Biol. Chem.*, 256 (1981) 3853–3857.
10 K. DILL AND A. ALLERHAND, *FEBS Lett.*, 107 (1979) 26–29.
11 A. ALLERHAND, K. DILL, E. BERMAN, J. M. LACOMBE, AND A. A. PAVIA, *Carbohydr. Res.*, 97 (1981) 331–336.
12 L. MUIR AND Y. C. LEE, *J. Biol. Chem.*, 245 (1970) 502–509.
13 B. FERRARI AND A. A. PAVIA, *Carbohydr. Res.*, 79 (1980) c1–c7.
14 J. M. LACOMBE, A. A. PAVIA, AND J. M. ROCHEVILLE, *Can. J. Chem.*, 59 (1981) 473–481.
15 D. SHAW, *Fourier Transform NMR Spectroscopy*, Elsevier, Amsterdam, 1976.
16 K. BOCK, I. LUNDT, AND C. PEDERSEN, *Tetrahedron Lett.*, (1973) 1037–1040.
17 K. BOCK AND C. PEDERSEN, *J. Chem. Soc., Perkin Trans. 2*, (1974) 293–297.
18 P. A. J. GORIN AND M. MAZUREK, *Can. J. Chem.*, 53 (1975) 1212–1223.
19 N. YAMAOKA, T. USUI, H. SUGIYAMA, AND S. SETO, *Chem. Pharm. Bull.*, 22 (1974) 2196–2200.
20 K. RICHARZ AND K. WÜTHRICH, *Biopolymers*, 17 (1978) 2133–2141.
21 A. A. PAVIA, S. N. UNG-CHHUN, AND J. M. LACOMBE, *Nouv. J. Chim.*, 5 (1981) 101–108.
22 K. DILL AND A. ALLERHAND, unpublished results.