Vol. 122, No. 2, 1984 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 31, 1984 Pages 859-866

THE IDENTIFICATION OF TERMINAL α(1→3)-LINKED GALACTOSE IN *N*-ACETYLLACTOSAMINE TYPE OF GLYCOPEPTIDES BY MEANS OF 500-MHz ^{1}H-NMR SPECTROSCOPY

Lambertus Dorland[¶], Herman van Halbeek and Johannes F.G. Vliegenthart[Ψ]

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

Received June 18, 1984

Summary. 500-MHz ^{1}H-NMR spectroscopy was employed to study two *N*-acetyllactosamine-type glycopeptide fractions which were derived from a bovine thyroglobulin preparation (Cummings, R.D., and Kornfeld, S. (1982) *J. Biol. Chem. 257*, 11230-11234). By this method, their branches were found to be terminated either by NeuAc in α(2→6)-linkage or by Gal in α(1→3)-linkage. For the first time, the Galα(1→3)Galβ(1→4)GlcNAcβ(1→•) sequence is characterized by ^{1}H-NMR to occur in *N*-glycosidic carbohydrate chains of glycoproteins. Moreover, this approach made possible the branch localization of such a unit. Microheterogeneity with respect to the presence of α-linked Gal or NeuAc in terminal position of a certain branch in one of the preparations, could be adequately assessed in terms of structures by ^{1}H-NMR.

Introduction. In the course of a study by Dr. R.D. Cummings and Dr. S. Kornfeld (Washington University, St. Louis, U.S.A.) aimed at elucidation of the specific sugar sequences in *N*-glycosidic carbohydrate chains that are recognized by[1] E-PHA and L-PHA lectins from *Phaseolus vulgaris*, they prepared two glycopeptide samples from a bovine thyroglobulin preparation, designated I-1 and I-2 (1). By means of high-resolution ^{1}H-NMR spectroscopic investigation of these samples, we detected,

[¶]Present address: University Children's Hospital, Nieuwe Gracht 137, NL-3512 LK Utrecht, The Netherlands.
[Ψ]To whom correspondence should be addressed.
[1]Abbreviations used are: DSS, sodium 4,4-dimethyl-silapentane-1-sulfonate; E-PHA, erythroagglutinating phytohemagglutinin; L-PHA, leukoagglutinating phytohemagglutinin; NMR, nuclear magnetic resonance; WEFT, water-eliminating Fourier-transform (non-selective 180° pulse - delay - 90° pulse - acquisition).

0006-291X/84 $1.50

inter alia, the[2] Galα(1→3)Galβ(1→4)GlcNAcβ(1→•) moiety to occur in the constituent compounds (1). Here we report the detailed characterization of this structural element, including its branch location, in I-1 and I-2, by ^{1}H-NMR spectroscopy.

Materials and methods. The glycopeptide preparations I-1 and I-2 were derived from a bovine thyroglobulin preparation by Dr. R.D. Cummings and Dr. S. Kornfeld. Details have been described (1). Prior to ^{1}H-NMR analysis, the glycopeptide samples were repeatedly treated with D_2O (99.96 %D, Aldrich, Milwaukee, U.S.A.), with intermediate lyophilization. The pD of the solutions was adjusted to 7. ^{1}H-NMR spectroscopy was performed on a Bruker WM-500 spectrometer (SON hf-NMR facility, Department of Biophysical Chemistry, Nijmegen University, The Netherlands), operating at 500 MHz in the Fourier-transform mode, at a probe temperature of 27°C (2). For solvent-peak suppression, a WEFT-pulse sequence (3) was applied. Chemical shifts are given for neutral solutions at 27°C, in parts per million relative to internal sodium 4,4-dimethyl--4-silapentane-1-sulfonate (DSS); they were actually measured by reference to internal acetone (δ 2.225), with an accuracy of 0.002 ppm.

Results and discussion. Sugar analysis of I-1 and I-2, performed by gas-liquid chromatography of trimethylsilylated methylglycosides obtained on methanolysis (4), showed the presence of Fuc, Man, Gal, GlcNAc and NeuAc as constituent monosaccharides. The 500-MHz ^{1}H-NMR spectrum of glycopeptide sample I-2 is depicted in Fig. 1. The NMR-spectral features of the main components of I-1 and I-2, recognizable from the expanded resolution-enhanced spectra, are summarized in Table 1.

The NMR-spectral data for glycopeptide fraction I-1 point to a triantennary *N*-acetyllactosamine type of structure for the main (> 90%) component. This can be deduced from the chemical shifts of H-1 of Man-$\underline{4}$ (δ 5.130) and Man-$\underline{4}$' (δ 4.932), in combination with those of H-2 of Man-$\underline{3}$, Man-$\underline{4}$ (coinciding at δ 4.216) and of Man-$\underline{4}$' (δ 4.112) (2). A Fuc residue is

[2]All sugars mentioned possess the D-configuration, except Fuc, which is L-Fuc.

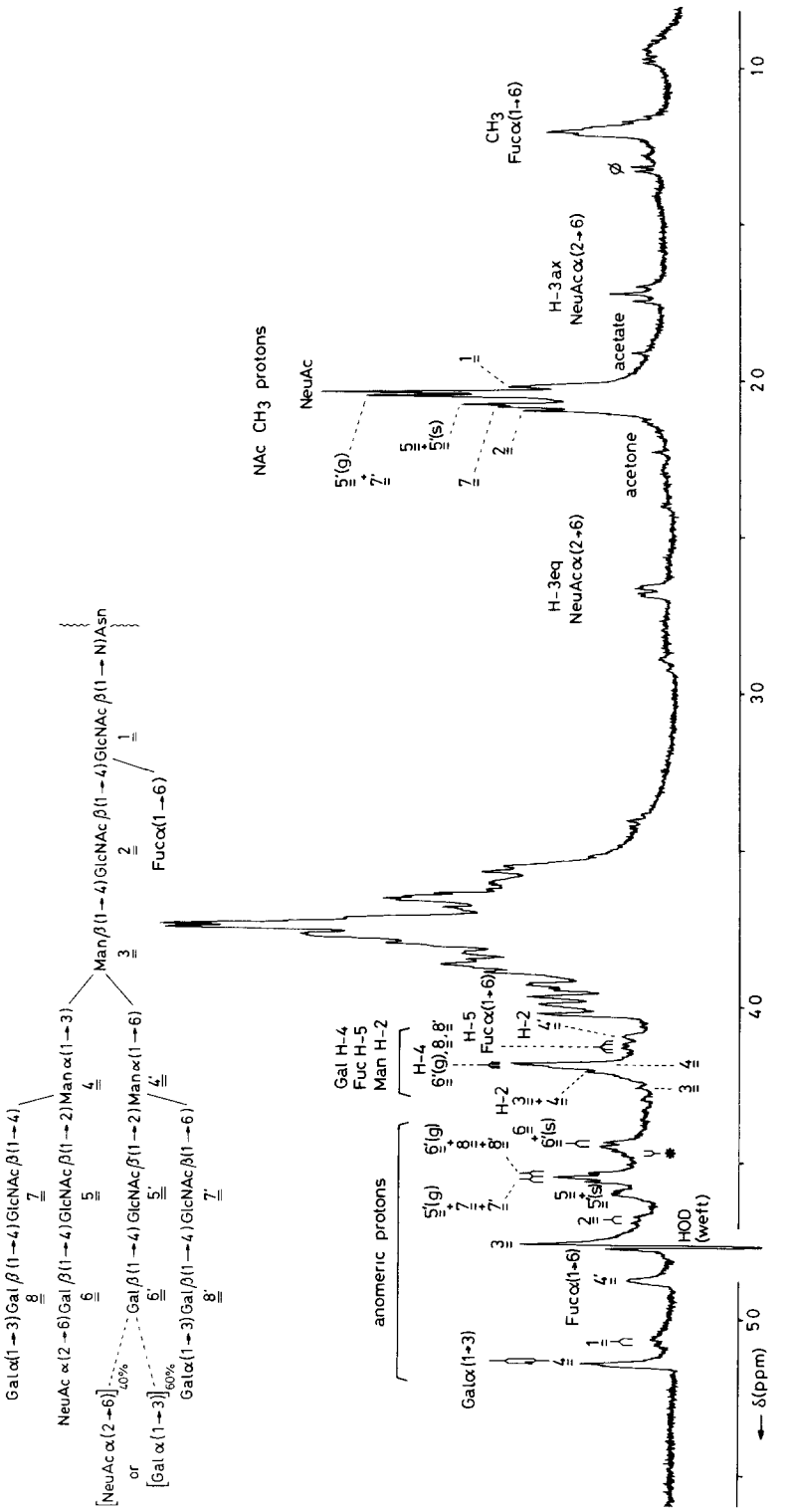
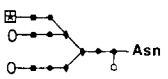
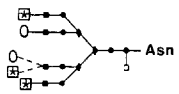


Fig. 1. *500-MHz ^{1}H-NMR spectrum (D_2O; pD 7; 27°C) of glycopeptide sample I-2 derived (1) from a bovine thyroglobulin preparation.*

The HOD-line was suppressed by a WEFT-pulse sequence. The numbers in the spectrum refer to the corresponding residues in the structural formula. The superscripts *g* and *s* refer to the galactosylated and the sialylated 5'-6' branch, respectively. The signal at δ 4.258, marked by H-2 of Man-3 below the spectrum, might indicate the presence of a small amount of tri'-antennary glycopeptide, that is, the analogue of I-2 missing the galactosylated 7-8 branch (see also footnote to Table 1). The signal at δ 4.47, marked by an asterisk, indicates that a very small number of β-Gal residues occur in terminal position, without NeuAc-α(2→6)- or Galα(1→3)-substitution.

Table 1

^{1}H chemical shifts of structural-reporter groups of constituent monosaccharides for the main components of glycopeptide preparations I-1 and I-2 derived (1) from bovine thyroglobulin

Reporter group	Residue[a]	Chemical shift[a,b] in I-1* 	Chemical shift[a,b] in I-2** 
H-1	GlcNAc-1	5.073[c] / 5.053	5.075
	GlcNAc-2	4.681	4.681
	Man-3	4.77	4.77
	Man-4	5.130	5.14[d]
	Man-4'	4.932	4.882
	GlcNAc-5	4.595	4.601
	GlcNAc-5'	4.602	4.59[s,g]
	Gal-6	4.441	4.443
	Gal-6'	4.447	4.443[s]/4.540[g]
	GlcNAc-7	4.539	4.556
	GlcNAc-7'	-	4.556
	Gal-8	4.539	4.540
	Gal-8'	-	4.540
	Galα(1→3)	5.145	5.148[e]
	Fucα(1→6)	4.874	4.88[f]
H-2	Man-3	4.216	4.208
	Man-4	4.216	4.208
	Man-4'	4.112	4.095
H-3ax	NeuAcα(2→6)	1.717[e]	1.720[e]
H-3eq	NeuAcα(2→6)	2.668[e]	2.672[e]
H-4	•→3)Galβ(1→4)	4.181	4.192[e]
H-5	Fucα(1→6)	4.12	4.126
	Galα(1→3)	4.193	4.192[e]
CH_3	Fucα(1→6)	1.20	1.208
NAc	GlcNAc-1	2.008[c] / 2.014	2.011[c] / 2.017
	GlcNAc-2	2.093	2.093
	GlcNAc-5	2.068	2.072
	GlcNAc-5'	2.066	2.070[s]/2.045[g]
	GlcNAc-7	2.077	2.080
	GlcNAc-7'	-	2.045
	NeuAcα(2→6)	2.029[h]	2.031[h]

[a] For comprehensive structure and numbering of monosaccharide residues, see Fig. 1.

α(1→6)-linked to GlcNAc-1; it is characterized by the chemical shifts of its own structural-reporter groups (H-1, H-5 and CH_3, see Table 1), and by the typical resonance positions of the structural reporters of GlcNAc-2 (δH-1 4.681; δNAc 2.093) (2). The β(1→2)-linked *N*-acetyllactosamine branches bear a NeuAc residue in α(2→6)-linkage to Gal-6 and -6', respectively. Evidence for this stems from the chemical shifts of H-3ax and H-3eq of the NeuAc residues together with those of H-1 of Man-4 and -4', and of the *N*-acetyl singlets of GlcNAc-5 and -5' (2). The β(1→4)-linked *N*-acetyllactosamine (7-8) branch is terminated by a Gal residue in α(1→3)-linkage to Gal-8. This can be inferred from the chemical shift of H-1 of the terminal Gal residue (δ 5.145) which, in combination with $J_{1,2}$ (≃ 3.5 Hz) indicates an α-type of linkage. The H-5 of this α-linked Gal is encountered at δ 4.193. The (1→3)-type of linkage is substantiated by the resonance position of Gal-8 H-4 (δ 4.181) (3). It should be mentioned that a Galα(1→4)Galβ(1→4)GlcNAcβ(1→•)

Table 1 *(continued)*

[b] Chemical shifts were acquired at 500 MHz for D_2O solutions at pD 7 and 27°C; they are expressed in ppm downfield from DSS. In the table-heading compounds are represented by schematic structures, according to (2):
◆— = Man; ●— = GlcNAc; ■— = Galβ(1→4); ○— = NeuAcα(2→6);
□— = Fuc; ⊠— = Galα(1→3).

[c] Signal is doubled due to heterogeneity of the peptide moiety.

[d] Value could not be determined more accurately (± 0.01 ppm), due to interference with the Galα(1→3) anomeric proton signal.

[e] Signal from more than one proton.

[f] Value could not be determined more accurately (± 0.01 ppm), due to interference with the Man-4' anomeric proton signal.

[h] Signal from more than one methyl group.

[s] Value for the sialylated 5'-6' branch.

[g] Value for the galactosylated 5'-6' branch.

---

* The relatively low signals at δ 4.254 and 4.193 may indicate the presence of a diantennary glycopeptide as a minor constituent (≃ 10%) of I-1; in this case, these signals are attributed to the Man-3 and Man-4 H-2 atoms, respectively. However, due to the low amount of this diantennary compound in the mixture, no detailed ^{1}H-NMR parameters could be derived for it, nor could the location of NeuAcα(2→6) and/or Galα(1→3) in a specific branch be inferred.

** The occurrence of a signal at δ 4.258, attributable to Man-3 H-2, in combination with the absence of a signal at δ 4.92 (H-1 of Man-4', if only β(1→2)-substituted), may suggest the presence of a small amount (≃ 15%) of a tri'-antennary analogue of the structure given in Fig. 1, in the mixture I-2. However, the signal at δ 4.258 might otherwise be assigned to the H-α of an amino acid in the peptide moiety.

moiety (5) is characterized by H-1 of the α(1→4)-linked Gal at[3] δ 4.95. The shift increments and decrements, caused by the introduction of Gal in α(1→3)-linkage upon the reporter groups of the *N*-acetyllactosamine unit 7-8, are in line with those described for the step from the disaccharide *N*-acetyllactosamine to the trisaccharide Galα(1→3)Galβ(1→4)GlcNAc (3): for H-1 of Gal-8, Δδ 0.072 ppm; for H-1 of GlcNAc-7, Δδ -0.006 ppm. However, the chemical shift of NAc of GlcNAc-7 is hardly or not affected by the introduction of the outer Gal, when compared to the asialo triantennary glycopeptide (GPII-5) derived from α_1-acid glycoprotein (2,6). The peptide portion of I-1 is rather heterogeneous, which comes to expression in the multiplicity of the H-1 and NAc signals of GlcNAc-1 (see Table 1).

Glycopeptide fraction I-2 contains a tetra-antennary compound of the *N*-acetyllactosamine type as the major (> 80%) constituent. All components of I-2 bear a Fuc residue in α(1→6)--linkage to GlcNAc-1. The kind of branching is evident from the chemical shift values of the Man H-2 signals (δH-2 of Man-3 and of Man-4 is 4.208; δH-2 of Man-4' is 4.095) (2). Gal-6 bears a NeuAc residue in α(2→6)-linkage, as can be concluded on the basis of the chemical shifts of H-3ax and H-3eq of NeuAc, and those of H-1 of Man-4 (δ ≃ 5.14) and the *N*-acetyl group of GlcNAc-5 (δ 2.072) (compare Ref. 2). The *N*-acetyllactosamine moieties 7-8 and 7'-8' each are substituted by Gal in α(1→3)--linkage to Gal-8 and -8', respectively. This can be inferred from the relatively intense, coinciding Gal H-1 signals at δ 5.148 (see Fig. 1), from the position of the α-Gal H-5 signals (δ ≃ 4.19) next to those of H-4 of the substituted Gal-8 and -8' (δ 4.192) and from the shift effect upon H-1 of Gal-8

[3]H. van Halbeek, L. Dorland, J.F.G. Vliegenthart, C. François-Gérard, G. Spik, and J. Montreuil: unpublished results.

and -8' ($\Delta\delta \simeq 0.07$ ppm, with respect to an asialo tetra-antennary glycopeptide (GPV-4), see Refs 2,6). The shift effects on H-1 and NAc signals of GlcNAc-7 and -7' are barely traceable. The remaining branch, 5'-6', shows microheterogeneity with respect to the type of termination. In $\simeq$ 40% of the 5'-6' branches in the mixture I-2, Gal-6' is substituted by NeuAc in $\alpha(2\rightarrow6)$-linkage. In this case, the *N*-acetyl signal of GlcNAc-5' is at δ 2.070, and H-1 of Gal-6' resonates at δ 4.443. The 5'-6' branches in the other components in the mixture contain Gal in $\alpha(1\rightarrow3)$-linkage to Gal-6' as is evident from H-1 of Gal-6' being observed at δ 4.540 and the *N*-acetyl signal of GlcNAc-5' at δ 2.045. The quantitative aspect of this result suffers from the inherent inaccuracy of integrating overlapping ^{1}H-NMR resonances.

The Gal$\alpha(1\rightarrow3)$Gal$\beta(1\rightarrow4)$GlcNAc$\beta(1\rightarrow\bullet)$ structural element has been reported to occur in the carbohydrate chains of membrane glycoproteins of Ehrlich ascites tumour cells (7,8) and also in those of Friend murine leukemia virus glycoproteins gp 69/71 (9). However, to our best knowledge, it has not been described for thyroid glycoproteins, as yet (see, for example, Refs. 10-13), nor has it been characterized by ^{1}H-NMR spectroscopy in glycoprotein carbohydrates.

It may be mentioned that the aforementioned trisaccharide unit has also been found to occur in glycolipids. In these compounds it was identified *i.a.* by ^{1}H-NMR spectroscopy, using dimethylsulfoxide-d_6 as a solvent (14-16). However, the NMR data published for glycolipids could not be used to identify unambiguously the Gal$\alpha(1\rightarrow3)$Gal$\beta(1\rightarrow4)$GlcNAc$\beta(1\rightarrow\bullet)$ moiety in I-1 and I-2, which is mainly due to the difference in the solvent used. Both sets of ^{1}H-NMR data are therefore of a unique value.

In conclusion, the primary structures of the bovine thyroglobulin glycopeptides I-1 and I-2 (1) have been determined in full detail by ^{1}H-NMR spectroscopy at 500 MHz. This includes the finding of the relatively scarcely found Galα(1→3)Galβ(1→4)-GlcNAcβ(1→•) branches, as well as the localization of such branches in tri- and tetra-antennary asparagine-linked oligosaccharides. The structure elucidation of I-2 is another example of the applicability of high-resolution ^{1}H-NMR spectroscopy in characterizing microheterogeneity.

Acknowledgements. Thanks are due to Mr G.J. Gerwig for performing carbohydrate analyses. This investigation was supported by the Netherlands Foundation for Chemical Research (SON/ZWO) and by grant UUKC 83-13 from the Netherlands Foundation for Cancer Research (KWF).

References

1. Cummings, R.D., and Kornfeld, S. (1982) *J. Biol. Chem. 257*, 11230-11234.
2. Vliegenthart, J.F.G., Dorland, L., and Van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem. 41*, 209-374.
3. Van Halbeek, H., Vliegenthart, J.F.G., Winterwerp, H., Blanken, W.M., and Van den Eijnden, D.H. (1983) *Biochem. Biophys. Res. Commun. 110*, 124-131.
4. Kamerling, J.P., and Vliegenthart, J.F.G. (1982) *Cell Biol. Monogr. 10*, 95-100.
5. François-Gérard, C., Pierce-Crétel, A., André, A., Dorland, L., Van Halbeek, H., Vliegenthart, J.F.G., Montreuil, J., and Spik, G. (1983) in *Proc. 7th Int. Symp. Glycoconj.*, Lund-Ronneby, Sweden (Chester, M.A., Heinegård, D., Lundblad, A., and Svensson, S., eds.) pp. 169-170.
6. 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.
7. Eckhardt, A.E., and Goldstein, I.J. (1983) *Biochemistry 22*, 5280-5289
8. Eckhardt, A.E., and Goldstein, I.J. (1983) *Biochemistry 22*, 5290-5297
9. Geyer, R., Dabrowski, U., Dabrowski, J., Egge, H., Geyer, H., Hunsmann, G., Schneider, J., and Stirm, S. (1983) in *Proc. 7th Int. Symp. Glycoconj.*, Lund-Ronneby, Sweden (Chester, M.A., Heinegård, D., Lundblad, A., and Svensson, S., eds.) pp. 292-293
10. Montreuil, J. (1982) in *Comprehensive Biochemistry* (Neuberger, A., and Van Deenen, L.L.M., eds.) Vol. 19B part II, pp. 1-188, Elsevier, Amsterdam.
11. Yamamoto, K., Tsuji, T., Irimura, T., and Osawa, T. (1981) *Biochem. J. 195*, 701-713.
12. Irimura, T., and Nicolson, G.L. (1983) *Carbohydr. Res. 115*, 209-220.
13. Van den Eijnden, D.H., Blanken, W.M., Winterwerp, H., and Schiphorst, W.E.C.M. (1983) *Eur. J. Biochem. 134*, 523-530.
14. Hanfland, P., Egge, H., Dabrowski, U., Kuhn, S., Roelcke, S., and Dabrowski, J. (1981) *Biochemistry 20*, 5310-5319.
15. Dabrowski, J., and Hanfland, P. (1982) *FEBS Lett. 142*, 138-142.
16. Dabrowski, J., Hanfland, P., and Egge, H. (1982) *Methods Enzymol. 83*, 69-86.