

# Rapid and simple approach for the NMR resonance assignment of the carbohydrate chains of an intact glycoprotein

## Application of gradient-enhanced natural abundance $^1\text{H}$ - $^{13}\text{C}$ HSQC and HSQC-TOCSY to the $\alpha$ -subunit of human chorionic gonadotropin

Tonny de Beer<sup>a</sup>, Carol W.E.M. van Zuylen<sup>a</sup>, Karl Hård<sup>b</sup>, Rolf Boelens<sup>b</sup>, Robert Kaptein<sup>b</sup>, Johann P. Kamerling<sup>a</sup>, Johannes F.G. Vliegenthart<sup>a,\*</sup>

*Bijvoet Center, Department of <sup>a</sup>Bio-Organic Chemistry and <sup>b</sup>NMR Spectroscopy, Utrecht University, Padualaan 8, NL-3584 CH Utrecht, The Netherlands*

Received 2 May 1994

### Abstract

The structure assessment of an intact glycoprotein in solution requires an extensive assignment of the carbohydrate NMR resonances. However, assignment of homonuclear spectra is very complicated because of the severe overlap of protein and carbohydrate signals. Application of pulsed field gradients allowed high quality natural abundance  $^1\text{H}$ - $^{13}\text{C}$  HSQC and HSQC-TOCSY spectra to be recorded of the  $\alpha$ -subunit of human chorionic gonadotropin. Most carbohydrate  $^1\text{H}$ - $^{13}\text{C}$  correlations appear in a distinct region between the aromatic region and the protein  $\text{C}^\alpha$ - $\text{H}^\alpha$  region. The enormous reduction in overlap led to fast and unambiguous assignment of the anomeric  $^1\text{H}$ - $^{13}\text{C}$  correlations. Subsequently, correlations of the monosaccharide skeleton atoms were readily assigned in the HSQC-TOCSY spectrum.

**Key words:** Glycoprotein;  $\alpha$ -subunit; Human chorionic gonadotropin; Pulsed field gradients; Nuclear magnetic resonance;  $^1\text{H}$ - $^{13}\text{C}$  correlation spectroscopy

### 1. Introduction

The functional significance of oligosaccharide chains of numerous glycoproteins in higher organisms is well established and has been comprehensively reviewed [1,2]. Despite the recent explosive development of multidimensional NMR techniques, not much is known about the structure of glycoproteins in solution. Determination of the structure of a glycoprotein requires an extensive assignment of the NMR resonances of the carbohydrate chains, in addition to those of the protein. The 'structural reporter group' concept has proven to be a very useful approach in the  $^1\text{H}$  NMR resonance assignment of oligosaccharide structures [3,4]. In homonuclear two dimensional (2D) spectra these structural reporter groups are used as starting points for the assignment of the resonances of the corresponding skeleton protons [5–8]. However, in homonuclear 2D spectra of intact glycoproteins, straightforward assignment of the oligosaccharide

resonances is complicated because the anomeric proton resonances coincide with the region containing most protein backbone  $\text{C}^\alpha$ -proton resonances [9–11]. Inspection of a database of NMR spectral data of proteins reveals that most  $\alpha$ -carbons are found within the  $^{13}\text{C}$  chemical shift range of 44–64 ppm [12,13], while most carbons of the carbohydrates resonate within the range of 61–105 ppm [14]. Taking advantage of this chemical shift difference, several groups have reported the partial  $^{13}\text{C}$  chemical shift assignment of carbohydrates linked to a glycoprotein by application of natural abundance one dimensional (1D)  $^{13}\text{C}$  NMR [15]. Recently, Medvedeva et al. [16] showed that the use of pulsed field gradients greatly improves the quality of heteronuclear spectra of proteins at the  $^{13}\text{C}$  natural abundance level. Here, we demonstrate that gradient-enhanced natural abundance  $^1\text{H}$ - $^{13}\text{C}$  HSQC and HSQC-TOCSY spectra of an intact glycoprotein greatly facilitate the NMR resonance assignment of the oligosaccharide chains. In the framework of our study of the structure of the  $\alpha$ -subunit of human chorionic gonadotropin ( $\alpha$ -hCG), we demonstrate the approach for this glycoprotein. hCG is a member of the family of heterodimeric glycoprotein hormones, and is involved in maintaining the corpus luteum in the early weeks of pregnancy [17,18]. The  $\alpha$ -subunit of 92 amino acids is N-glycosylated at Asn-52 and Asn-78, and contains either N-acetylglucosamine or hybrid type oligosaccharide chains [1,19–22].

\*Corresponding author. Fax: (31) (30) 540 980.

**Abbreviations:** NMR, nuclear magnetic resonance; HSQC,  $^1\text{H}$ -detected heteronuclear single quantum coherence spectroscopy; TOCSY, total correlation spectroscopy; INEPT, insensitive nuclei enhanced by polarization transfer; MLEV, M. Levitt; GARP, globally optimized alternating-phase rectangular pulses; hCG, human chorionic gonadotropin; Gal, D-galactose; Man, D-mannose; GlcNAc, N-acetyl-D-glucosamine; NeuAc, N-acetyl-D-neuraminic acid.

Most other carbohydrate NMR cross-peaks are observed in the region from 3.4 to 4.3 ppm ( $^1\text{H}$ ) and from 51 to 83 ppm ( $^{13}\text{C}$ ), indicated by the dotted box in

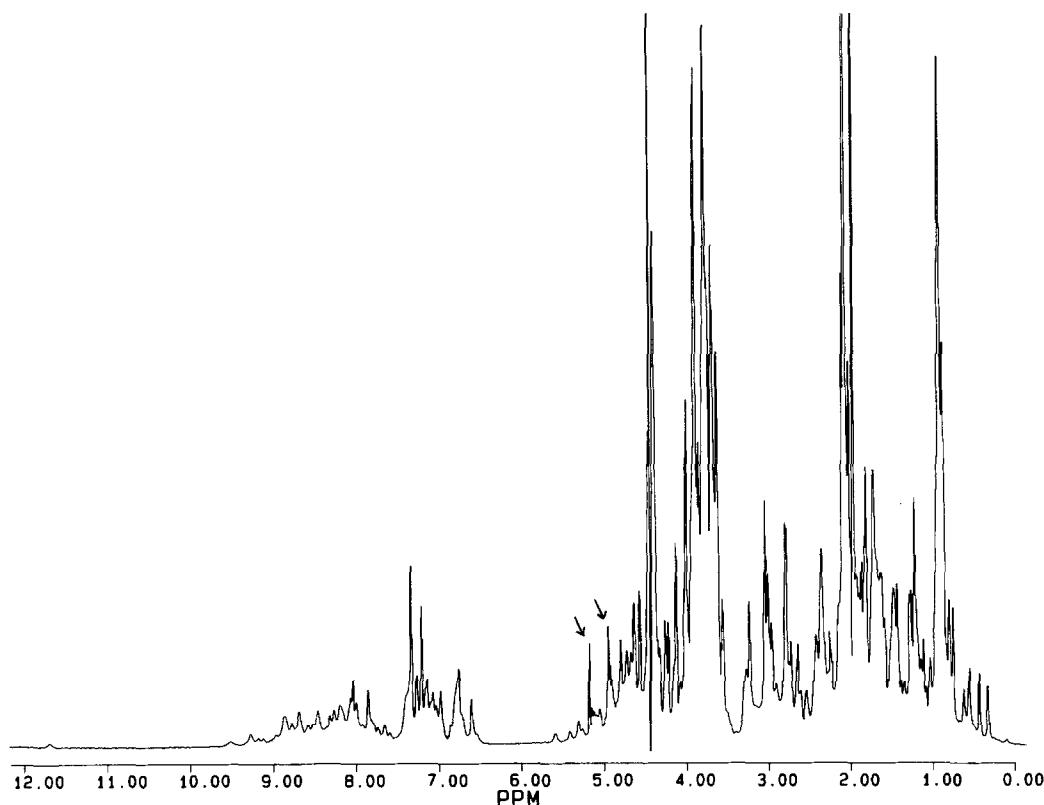


Fig. 1. 1D  $^1\text{H}$  NMR spectrum of 5 mM  $\alpha$ -hCG in 95% (v/v)  $\text{H}_2\text{O}/^2\text{H}_2\text{O}$ , 0.3 M NaCl, 1 mM  $\text{NaN}_3$ . The pH was adjusted to 5.1 by addition of dilute HCl. The experiment was performed on a Varian Unity plus 750 MHz spectrometer at a probe temperature of 338 K. The spectrum was recorded with a 1D NOE sequence with presaturation of the  $\text{H}_2\text{O}$  resonance during the relaxation delay. The  $^1\text{H}$  carrier frequency was set at 750.368 MHz. The spectral width was 12,000 Hz. In total 64 FIDs of 2,048 complex data points were acquired. The data were processed using the TRITON NMR software package (Bijvoet Center, Department of NMR Spectroscopy). In short, the time domain data were multiplied by a phase shifted squared sine bell window, and baseline corrected after Fourier transformation and zero filling to 4,096 data points with a fifth order polynomial fit. The Man-4 and the Man-4' anomeric resonances are indicated by an arrow.

Fig. 2A. To correlate the anomeric atoms with their respective skeleton atoms, a gradient-enhanced natural abundance 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC-TOCSY experiment was performed on the 5 mM  $\alpha$ -hCG sample. A high quality HSQC-TOCSY spectrum with sufficient resolution in both directions was obtained within a reasonable acquisition time (60 h). The part of the spectrum containing most carbohydrate NMR cross-peaks, and a corresponding part of the HSQC spectrum are shown in Fig. 3. To illustrate the assignment procedure, the assignment of a Gal residue, which is sialylated at C3, is described. On the H1 track through the anomeric  $^1\text{H}$ - $^{13}\text{C}$  correlation at  $\delta_{\text{H}}/\delta_{\text{C}} = 4.52/104.0$  (cross-peak a in Fig. 3) three relayed cross-peaks are observed, with  $^{13}\text{C}$  chemical shifts of 76.8 ppm (b), 70.5 ppm (c) and 68.8 ppm (d), respectively. On each of the  $^{13}\text{C}$  tracks through the cross-peaks b–d, a connectivity pattern of four protons is observed (H1–H4). Inspection of the HSQC spectrum then identifies the  $^1\text{H}$ - $^{13}\text{C}$  correlations of this Gal residue (e,f,g). The C3 (e) atom is readily assigned because of the downfield shift to 76.8 ppm, which is typical of a 3-substitution. The homonuclear  $^3J_{\text{HH}}$  coupling patterns

observed for the relayed cross-peaks on the  $^{13}\text{C}$  tracks, and spectral data of ( $\alpha$ 2-3)sialyllactose were used to discriminate between the H2/C2 (f) cross-peak and the H4/C4 (g) cross-peak [34].

Assignment of the NMR cross-peaks stemming from Man residues of an intact glycoprotein can be troublesome, because of the weak TOCSY transfer due to the small homonuclear coupling constants  $^3J_{12}$  and  $^3J_{23}$  [10,11]. To show the quality of the HSQC-TOCSY spectrum the assignment pathway of Man-4 is outlined. A weak cross-peak (i) on the H1 track through the anomeric correlation (h at  $\delta_{\text{H}}/\delta_{\text{C}} = 5.13/100.7$ ) and a cross-peak (j) on the  $^{13}\text{C}1$  track leads to assignment of the  $^1\text{H}$ - $^{13}\text{C}$  correlation of the H2/C2 pair (k at  $\delta_{\text{H}}/\delta_{\text{C}} = 4.17/78.0$ ). On the H2 track through this cross-peak three relayed cross-peaks (l–n) are observed. Inspection of the HSQC spectrum identifies the corresponding  $^1\text{H}$ - $^{13}\text{C}$  correlations on the respective  $^{13}\text{C}$  tracks (o–q), and comparison with chemical shift data of the Man-4 residue in a diantennary deca-saccharide [33] makes the assignment possible (H3/C3,p;H4/C4,q;H5/C5,o). The  $^1\text{H}$ - $^{13}\text{C}$  correlations of oligosaccharide hydroxymethyl groups appear

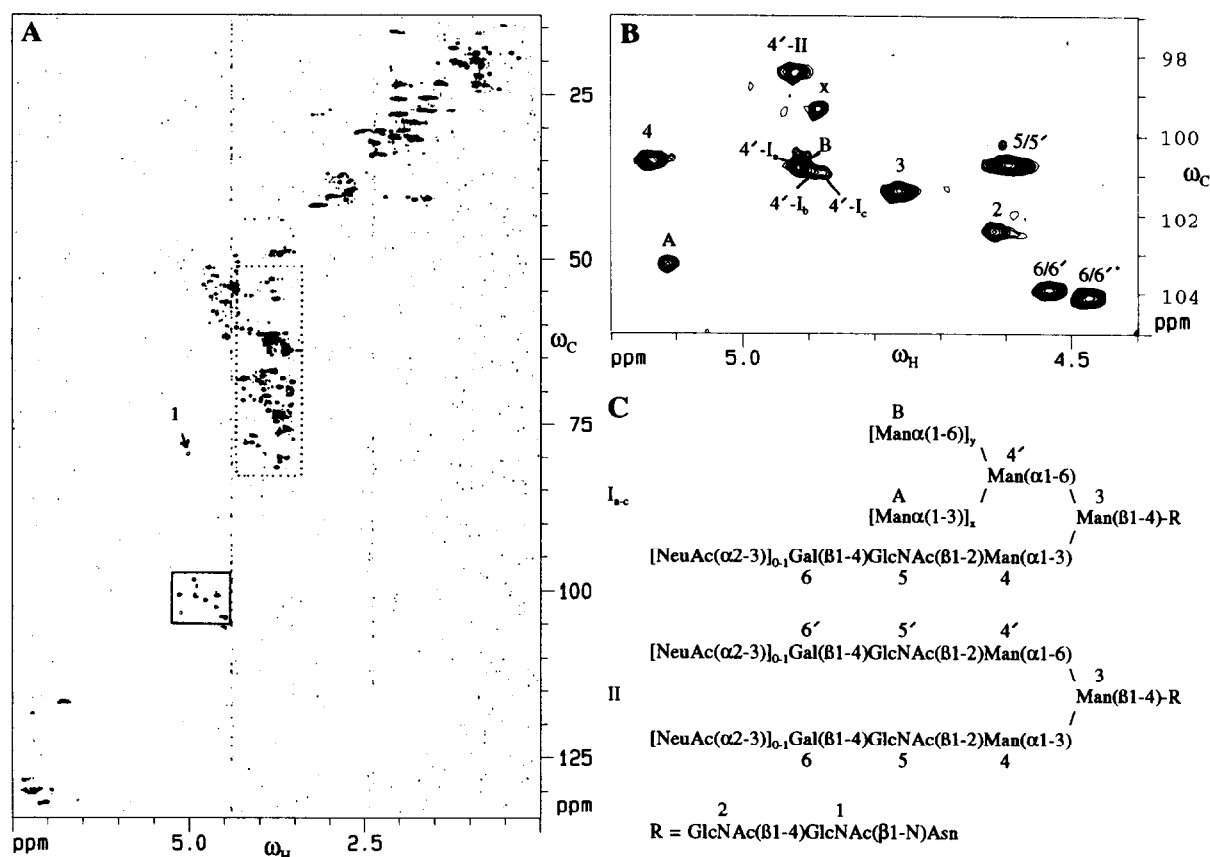


Fig. 2. (A)  $^1\text{H}$ - $^{13}\text{C}$  2D HSQC spectrum of 5 mM  $\alpha$ -hCG in 99.95%  $^2\text{H}_2\text{O}$ , 0.1 M NaCl, 1 mM  $\text{NaN}_3$ . The pH was adjusted to 4.7 by adding dilute  $^2\text{HCl}$  (no correction was made to the pH meter reading for the deuterium isotope effect). The experiment was performed on a Bruker AMX-600 spectrometer equipped with a 5 mm shielded-gradient triple resonance probe at a probe temperature of 328 K. The  $^{13}\text{C}$  carrier frequency was placed at 150.91625 MHz. The HSQC experiment was recorded with 1,024 experiments in the  $t_1$  direction. Per experiment 64 free induction decays of 2048 data points were recorded. The  $^1\text{H}$  carrier frequency was placed at 600.140784 MHz. Quadrature detection in the  $t_1$  dimension was achieved by the States-TPPI method [36]. The spectral width was 7,042 Hz in the  $t_2$  direction, and 21,741 Hz in the  $t_1$  direction. The data set was processed with the Bruker UXNMR software package. In short, time domain data were multiplied with a phase shifted squared sine bell function. The data set of  $2\text{k} \times 2\text{k}$  data points, resulting after Fourier transformation and zero filling, was baseline corrected in both directions with a fifth order polynomial fit. Chemical shifts are given by reference to acetone ( $\delta_{\text{H}} = 2.225$ ) and sodium 3-(trimethylsilyl)-propionate ( $\delta_{\text{C}} = 0$ ). The arrow indicates the anomeric  $^1\text{H}$ - $^{13}\text{C}$  correlation of one of the Asn-linked GlcNAc-1 residues. The area within the dotted box contains most carbohydrate  $^1\text{H}$ - $^{13}\text{C}$  correlations. The anomeric region lies within the solid box, and is enlarged in B. (B) Anomeric region of the  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum (the notation is depicted in Fig. 2C, and taken from Vliegthart et al. [3]). 6/6''\* indicates the  $^1\text{H}$ - $^{13}\text{C}$  correlation of asialo Gal-6/6' residues. The Man  $^1\text{H}$ - $^{13}\text{C}$  correlation indicated by  $\times$  has an upfield  $^{13}\text{C}$  chemical shift relative to Man-4' of structures I<sub>a</sub>-C (C), but could not be assigned as yet based on spectral data [15,19,21,33,34] of the free oligosaccharide structures reported for  $\alpha$ -hCG [1,19-22]. (C) Oligosaccharide chains reported for  $\alpha$ -hCG [1,19-22]. I<sub>a</sub>,  $x = 0$ ,  $y = 0$ ; I<sub>b</sub>,  $x = 1$ ,  $y = 0$ ; II,  $x = 0$ ,  $y = 1$ .

in a region also containing some protein  $\text{H}^\alpha\text{-C}^\alpha$  correlations, and cross-peaks stemming from some amino acid side chains (c.f. Ser, Thr). The typical pattern of two hydrogens attached to a single carbon in the HSQC spectrum, the relayed coherence from the hydroxymethyl protons to carbohydrate ring protons, and the similar connectivity pattern on the  $^{13}\text{C}5$  track in the HSQC-TOCSY spectrum makes assignment of the hydroxymethyl carbons unambiguously. For example, the Man-4 H6 and H6' resonances were assigned based on the relayed correlations (r and s) observed on the  $^{13}\text{C}5$  track at 74.8 ppm. The corresponding C6 atom was identified by the connectivity patterns in the HSQC and the HSQC-TOCSY spectrum, observed at 62.9 ppm.

Apart from the hydroxymethyl  $^1\text{H}$ - $^{13}\text{C}$  correlations, the H2/C2 cross-peaks of GlcNAc residues also appear in a region crowded with cross-peaks stemming from the protein. In this case the relayed connectivity to their corresponding anomeric  $^1\text{H}$ - $^{13}\text{C}$  correlations is used for assignment, as indicated in Fig. 3 for the GlcNAc-5/5' residues (v, at  $\delta_{\text{H}}/\delta_{\text{C}} = 4.59/56.2$ ). Despite the small differences in  $^1\text{H}$  chemical shift of the GlcNAc-5,5' H2, H3 and H4 atoms, as can be seen on the  $^{13}\text{C}1$  track ( $\delta_{\text{C}} = 100.8$ , indicated by an arrow), assignment can be carried out accurately because of the large  $^{13}\text{C}$  chemical shift dispersion as is observed on the H1 track at 4.59 ppm (v(H2),u(H3),t(H4)).

In conclusion, to determine the structure of a

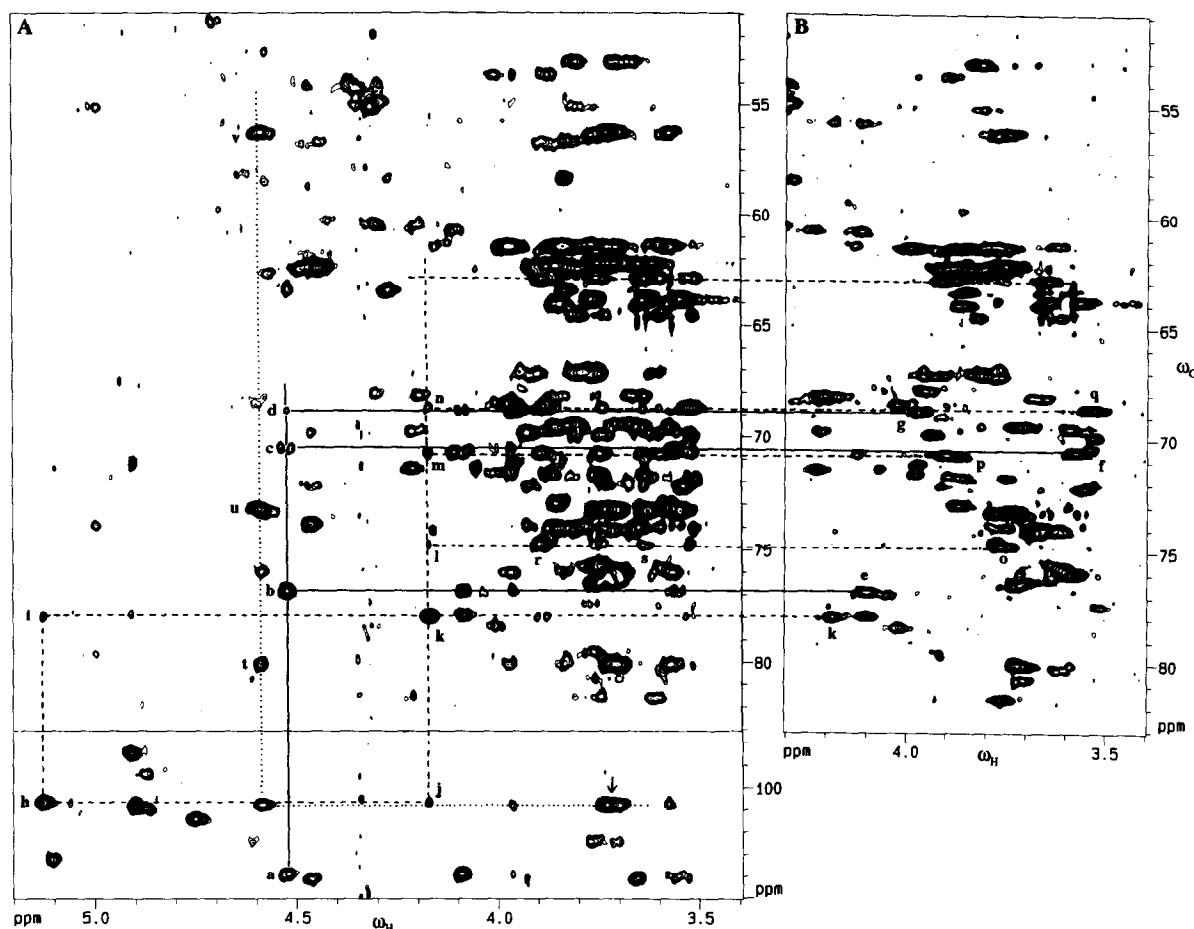


Fig. 3. (A) Part of the  $^1\text{H}$ - $^{13}\text{C}$  2D HSQC-TOCSY spectrum of  $\alpha$ -hCG and (B) a corresponding part of the HSQC spectrum. Experimental conditions were as described in the legend of Fig. 2 with the following differences and additions. In the  $t_1$  direction 679 experiments were recorded, and per experiment 338 free induction decays of 2,048 data points were collected. The  $^1\text{H}$  carrier frequency was placed at 600.140774 MHz. The spectral width was 6,024 Hz in the  $t_2$  direction, and 21,741 Hz in the  $t_1$  direction. The data set was processed as described for the HSQC spectrum. Correlation pathways are shown for Gal-6/6' (solid line) and Man-4 (dashed line). The GlcNAc-5/5' C1 and H1 track are also depicted (dotted lines). The arrow indicates relayed connectivities of GlcNAc-5/5' H1 to H2, H3 and H4.

glycoprotein in solution, or to study the conformation of oligosaccharide chains attached to a glycoprotein, an extensive assignment of the carbohydrate NMR resonances is required. The severe overlap of carbohydrate and protein signals observed in homonuclear spectra makes assignment of these spectra a tedious, and often impossible, task. The application of gradient-enhanced natural abundance heteronuclear experiments greatly simplifies this assignment. Even at a moderate glycoprotein concentration (5 mM) and with moderate acquisition times, high quality spectra are obtained. Most cross-peaks stemming from carbohydrates attached to an intact glycoprotein appear in a region ( $\delta_{\text{H}} = 3.4\text{--}5.2$ ,  $\delta_{\text{C}} = 61\text{--}105$ ) in the spectrum essentially devoid of cross-peaks of the protein backbone. Anomeric  $^1\text{H}$ - $^{13}\text{C}$  correlations are readily assigned, also taking advantage of the reduced overlap among themselves. Furthermore, the heteronuclear relayed experiment powerfully connects these anomeric correlations to their respective skeleton

atoms. This approach is an important step in the implementation of gradient-enhanced heteronuclear experiments, recorded at natural abundance, in the NMR resonance assignment of a glycoprotein.

**Acknowledgements:** The authors thank Dr. G.W.K. van Dedem (Diosynth, Oss) for supplying hCG, and Dr. P. Keiffer (Varian Assoc., Palo Alto) and Dr. P. Szandor (Varian Assoc., Darmstadt) for recording the 750 MHz 1D spectrum of  $\alpha$ -hCG. This investigation was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO), and by the Dutch Cancer Society (project IKMN 92-83).

## References

- [1] Kobata, A. (1992) *Eur. J. Biochem.* 209, 483–501.
- [2] Varki, A. (1993) *Glycobiology* 3, 97–130.
- [3] Vliegthart, J.F.G., Dorland, L. and van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem.* 41, 209–374.
- [4] Kamerling, J.P. and Vliegthart, J.F.G. (1992) *Biol. Magn. Reson.* 10, 1–194.

- [5] Dabrowski, J. (1987) in: *Two-Dimensional NMR Spectroscopy: Applications for Chemists and Biochemists* (Croasmun, W.R. and Carlson, R.M.K., eds.), pp. 349–386, VCH Publishers, New York.
- [6] Bush, C.A. (1989) *Bull. Magn. Reson.* 10, 73–95.
- [7] Homans, S.W. (1990) *Prog. NMR Spectrosc.* 22, 55–81.
- [8] Serianni, A.S. (1992) in: *Glycoconjugates: Composition, Structure, and Function* (Allen, H.J. and Kisailus, E.C., eds.) pp. 71–102, Marcel Dekker, New York.
- [9] Brockbank, R.L. and Vogel, H.J. (1990) *Biochemistry* 29, 5574–5583.
- [10] Wyss, D.F., Withka, J.M., Knoppers, M.H., Sterne, K.A., Recny, M.A. and Wagner, G. (1993) *Biochemistry* 32, 10995–11006.
- [11] Withka, J.M., Wyss, D.F., Wagner, G., Arulanandam, A.R.N., Reinherz, E.L. and Recny, M.A. (1993) *Structure* 1, 69–81.
- [12] Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.* 222, 311–333.
- [13] Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR* 4, 171–180.
- [14] Bock, K. and Pedersen, C. (1983) *Adv. Carbohydr. Chem. Biochem.* 41, 27–65.
- [15] Dill, K., Berman, E. and Pavia, A.A. (1985) *Adv. Carbohydr. Chem. Biochem.* 43, 1–49.
- [16] Medvedeva, S., Simorre, J.-P., Brutscher, B., Guerlesquin, F. and Marion, D. (1993) *FEBS Lett.* 333, 251–256.
- [17] Ryan, R.J., Charlesworth, M.C., McCormick, D.J., Milius, R.P. and Keutmann, H.T. (1988) *FASEB J.* 2, 2661–2669.
- [18] Iles, R.K. and Chard, T. (1993) *J. Mol. Endocrinol.* 10, 217–234.
- [19] Damm, J.B.L., Kamerling, J.P., van Dedem, G.W.K. and Vliegthart, J.F.G. (1988) *Glycoconjugate J.* 4, 129–144.
- [20] Blithe, D.L. (1990) *Endocrinology* 126, 2788–2799.
- [21] Weisshaar, G., Hiyama, J. and Renwick, A.G.C. (1991) *Glycobiology* 1, 393–404.
- [22] Stockell-Hartree, A. and Renwick, A.G.C. (1992) *Biochem. J.* 287, 665–679.
- [23] Hiyama, J., Surus, A. and Renwick, A.G.C. (1990) *J. Endocrinol.* 125, 493–500.
- [24] Bodenhausen, G. and Ruben, D.J. (1980) *Chem. Phys. Lett.* 69, 185–189.
- [25] Morris, G.A. and Freeman, R. (1979) *J. Am. Chem. Soc.* 101, 760–762.
- [26] Bax, A. and Pochapsky, S.S. (1992) *J. Magn. Reson.* 99, 638–643.
- [27] Otting, G. and Wüthrich, K. (1988) *J. Magn. Reson.* 76, 569–574.
- [28] Shaka, A.J., Barker, P.B. and Freeman, R. (1985) *J. Magn. Reson.* 64, 547–552.
- [29] Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Reson.* 53, 521–528.
- [30] Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.* 65, 355–360.
- [31] Griesinger, C., Otting, G., Wüthrich, K. and Ernst, R.R. (1988) *J. Am. Chem. Soc.* 110, 7870–7872.
- [32] John, B.K., Plant, D. and Hurd, R.E. (1993) *J. Magn. Reson.* A101, 113–117.
- [33] Wieruszski, J.-M., Michalski, J.-C., Montreuil, J. and Strecker, G. (1989) *Glycoconjugate J.* 6, 183–194.
- [34] Lerner, L. and Bax, A. (1987) *Carbohydr. Res.* 166, 35–46.
- [35] Hård, K., van Zadelhoff, G., Moonen, P., Kamerling, J.P. and Vliegthart, J.F.G. (1992) *Eur. J. Biochem.* 209, 895–915.
- [36] Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.* 85, 393–399.