

# Complete $^1\text{H}$ and $^{13}\text{C}$ resonance assignments of a 21-amino acid glycopeptide prepared from human serum transferrin<sup>1</sup>

Jianyun Lu<sup>2</sup>, Herman van Halbeek<sup>\*,3</sup>

*Complex Carbohydrate Research Center and Department of Chemistry, The University of Georgia, Athens, GA 30602-4712, USA*

Received 1 May 1996; accepted in revised form 8 September 1996

## Abstract

A 21-amino acid glycopeptide (Gp21) was isolated and purified in multi-milligram yields from commercially available human serum transferrin (HSTF) by a combination of tryptic digestion, Con A affinity chromatography, and reverse phase HPLC. The peptide chain of Gp21 contains a single N-glycosylation site to which a diantennary oligosaccharide is attached. The amino acid sequence and the glycan primary structure of Gp21 have been verified by peptide sequencing, electrospray mass spectrometry, and one-dimensional  $^1\text{H}$  NMR spectroscopy. Different glycoforms were found for the glycan of Gp21 derived from two different batches of commercial HSTF. These glycoforms differ from one another in the number of NeuAc residues (ranging from

Abbreviations: 1D, one-dimensional; 2D, two-dimensional; Con A, concanavalin A; COSY, correlated spectroscopy; DQ, double quantum; DQF, double quantum filtered; DTT, dithiothreitol; ES-MS, electrospray mass spectrometry; FID, free induction decay; HPAEC, high performance anion exchange chromatography; HPLC, high performance liquid chromatography; HSQC, heteronuclear single quantum correlation; HSTF, human serum transferrin; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; RCM, reduced and carboxymethylated; TBS, Tris-buffered saline; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy; TPCK, L-1-tosylamide-2-phenylethyl chloromethyl ketone; TPPI, time-proportional phase incrementation

\* Corresponding author.

<sup>1</sup> A preliminary account of this work was presented at the Annual Meeting of the Society for Glycobiology, Notre Dame, IN, November 1994. For an abstract, see *Glycobiology*, 4 (1994) 753.

<sup>2</sup> Current address: Washington University School of Medicine, 660 South Euclid Campus, Box 8051, St. Louis, MO 63110, USA.

<sup>3</sup> Current address: Rega Institute, Department of Medicinal Chemistry, Katholieke Universiteit Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium.

0 to 2) and/or the number of Gal residues (ranging from 1 to 2). As for the monogalacto species, in the two-dimensional nuclear Overhauser effect (NOE) spectrum of Gp21, interglycosidic NOEs were observed between Man4 in the  $\alpha(1 \rightarrow 3)$  branch and the terminal GlcNAc  $\beta(1 \rightarrow 2)$  residue. No interglycosidic NOE was observed between Man4' in the  $\alpha(1 \rightarrow 6)$  branch and the terminal GlcNAc residue. These observations indicate that the terminal GlcNAc residue in the minor glycoforms of Gp21 is exclusively located in the  $\alpha(1 \rightarrow 3)$  branch of the Gp21 glycan. The occurrence of such a carbohydrate structure in HSTF has not been reported before. The  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of Gp21 have been completely assigned by two-dimensional homonuclear and heteronuclear spectroscopy. The close similarity of the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shift values for the Gp21 glycan with the respective values for the peptide-free diantennary oligosaccharide (Wieruszski et al., *Glycoconjugate J.*, 6 (1989) 183–194) indicates that the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts of the diantennary oligosaccharide are not perturbed by the presence of the Gp21 peptide fragment. The complete  $^1\text{H}$  and  $^{13}\text{C}$  resonance assignments and the full characterization of the primary structure of Gp21 will permit us to study the conformation and dynamics of the N-linked diantennary oligosaccharides while covalently attached to a polypeptide fragment. © 1996 Elsevier Science Ltd.

**Keywords:** Human serum transferrin; N-linked diantennary oligosaccharide; Microheterogeneity; Nuclear magnetic resonance; Resonance assignment; Nuclear Overhauser effect

## 1. Introduction

The N-linked oligosaccharides of glycoproteins have been found to be involved in a variety of biological processes [1,2]. The study of the conformational and dynamic properties of these glycans may aid the understanding of their biological functions. Previous NMR studies have focused on the conformational analysis of peptide-free oligosaccharides [3,4]. However, there is an increasing interest in the conformations of N-linked oligosaccharides covalently attached to polypeptides and proteins [5–9], since there is a distinct possibility that the oligosaccharide conformation may change due to protein–carbohydrate interactions. One interesting example of such a conformational change was discovered in the X-ray crystal structure of the dimeric IgG<sub>1</sub> molecule's F<sub>c</sub>-fragment with a fucosylated N-linked diantennary oligosaccharide attached to each of the monomers [10]. The  $\alpha(1 \rightarrow 6)$  branch of the glycan rested on a hydrophobic section of the protein chain and adopted an orientation that was significantly different from that in the solution conformation of the corresponding free oligosaccharide. Another interesting example was recently presented by Wyss et al. [11,12] for the high-mannose glycan in CD2; the protein–carbohydrate interactions in this intact glycoprotein were characterized by NOE contacts between the two moieties and the glycan conformation was featured with the ManA  $\alpha(1 \rightarrow 3)$  branch folding backwards toward the first three residues of the pentasaccharide core. However, the results from NMR studies on other glycoproteins [6,8] suggest that the presence of the peptide may not have any measurable effect on the solution conformation of the N-linked oligosaccharides. Therefore, more studies are necessary in order to gain further insight into the solution conformation of N-type oligosaccharides covalently attached to polypeptides and proteins.

We are specifically interested in the conformation of the N-linked diantennary oligosaccharides of human serum transferrin (HSTF), an iron-binding glycoprotein



found in human serum. HSTF consists of a single polypeptide chain containing 679 amino acid residues [13,14] and two N-glycosylation sites at Asn413 and Asn611. The glycan chains at both glycosylation sites have been identified as either diantennary or triantennary oligosaccharides [15–17], with the former accounting for about 85–90% of the total carbohydrate content [18–20]. The conformations of the peptide-free diantennary oligosaccharides released from HSTF and human lactotransferrin have been studied by both NMR [21] and molecular modeling [22] methods. However, the effect of the polypeptide on the oligosaccharide conformation has not been studied so far.

It is difficult to study the conformation of the N-linked diantennary oligosaccharide covalently attached to the intact HSTF protein due to the large size of the molecule (about 80 kDa), which exceeds the current MW limit of a protein that can be studied by NMR [23]. Therefore, we sought to obtain suitable glycopeptides derived from HSTF by enzymatic digestion. Based on the HSTF amino acid sequence, we predicted that a complete tryptic digestion could generate two glycopeptides, each of which contains a single glycosylation site. One glycopeptide contains peptide fragment Phe402 to Arg414 (13 amino acid residues), the other contains peptide fragment Gln603 to Arg623 (21 amino acid residues). With Con A affinity chromatography followed by reverse phase HPLC [20], the glycopeptides containing N-linked diantennary oligosaccharides are readily separable from their triantennary counterparts. The glycopeptide family containing 21 amino acid residues and a diantennary oligosaccharide (see Fig. 1) is called Gp21 in this report. Gp21 will serve as a model compound for our future conformational and dynamics studies. Gp21 may be expected to show microheterogeneity in the terminal portions of the two glycan branches [19]. We report here the preparation and complete  $^1\text{H}$  and  $^{13}\text{C}$  resonance assignments of Gp21, and the full characterization of this microheterogeneity.

## 2. Experimental

**Materials.**—HSTF (lot numbers 67F-9454, 50H93161 and 72H9303), TPCK-treated trypsin, and Con A-agarose were purchased from Sigma (St. Louis, MO). Recombinant *N*-glycanase was purchased from Genzyme (Boston, MA). All other reagents were of analytical grade and were commercially available.

**Reduction and carboxymethylation.**—Reactions were carried out according to the procedure described by Fu and van Halbeek [20].

**Tryptic digestion.**—RCM-HSTF (100 mg, 1.22  $\mu\text{mol}$ ) was dissolved in 6 mL of 0.1 M  $\text{NH}_4\text{HCO}_3$  (pH 8.0), and TPCK-treated trypsin (1.0 mg) was added. The mixture was incubated at room temperature for 6 h. Another 1.0 mg of TPCK-treated trypsin was subsequently added, resulting in a total enzyme-substrate ratio of 1:50 (w/w). The digestion was allowed to proceed for another 6 h at room temperature and then quenched by placing the vial in boiling water for 3 min.

**Con A affinity chromatography.**—Con A affinity chromatography was performed at room temperature on a 20 mL Con A-agarose column ( $26.0 \times 1.0 \text{ cm}^2$ ). A solution of the tryptic digest of RCM-HSTF (100 mg) in 1 mL of 0.1 M  $\text{NH}_4\text{HCO}_3$  was loaded onto the column. Tris-buffered saline (TBS; 0.15 M NaCl, 0.01 M Tris, 1 mM  $\text{CaCl}_2$ ,

and 1 mM  $\text{MgCl}_2$ , pH 8.0) was applied to the column and 50 mL of eluant was collected and labeled as Con A-I. Then, TBS containing 10 mM methyl  $\alpha$ -D-glucopyranoside was applied and another 50 mL of eluant was collected and labeled as Con A-II. The Con A-I and II fractions were lyophilized, then desalted on a precalibrated Sephadex G-15 (12 g) column ( $43.0 \times 1.0 \text{ cm}^2$ ) by elution with 7% aqueous 1-propanol. The void volumes (19 mL) were collected and lyophilized.

**Reverse phase HPLC.**—The glycopeptides were purified on a Bio-Rad model 2800 gradient HPLC system. Either an Aquapore RP-300 analytical column ( $2.1 \times 220 \text{ mm}^2$ ) or an Aquapore Octyl (RP-300) semi-preparative column ( $10 \times 100 \text{ mm}^2$ ) from Applied Biosystems (Foster City, CA) was used depending on the amount of sample. Mobile phases A and B were 0.075% TFA in water and 0.060% TFA in 80% aqueous acetonitrile, respectively. The flow rate was 0.2 mL/min for the analytical column and 2.0 mL/min for the semi-preparative column. The desalted Con A-II fraction (10 mg) was dissolved in 1 mL of 0.1 M  $\text{NH}_4\text{HCO}_3$  and the sample was filtered through a microfilter fuge tube (0.45  $\mu\text{m}$  pore size) from Rainin (Woburn, MA) prior to injection. The glycopeptides detected by UV absorbance at 214 nm were collected and immediately dried or neutralized with M  $\text{NH}_4\text{HCO}_3$  to avoid hydrolysis of sialic acid.

**Release of N-linked diantennary oligosaccharides from the glycopeptide by N-glycanase digestion.**—The HPLC purified glycopeptide (10 nmol) was dissolved in 100  $\mu\text{L}$  of 0.1 M  $\text{NH}_4\text{HCO}_3$  (pH 8.0), and incubated at 37 °C for 15 h with recombinant N-glycanase (0.5 units). The digestion was then stopped by placing the sample vial in boiling water for 3 min. Free diantennary oligosaccharides were separated from peptides by reverse phase HPLC as described above. Oligosaccharides did not bind to the column and thus eluted at the void volume. Peptides eluted later in the run.

**Dionex high performance anion exchange chromatography.**—The released N-linked diantennary oligosaccharides were purified using a Dionex Bio-LC system (Sunnyvale, CA) equipped with a Dionex CarboPac PA-1 column ( $4 \times 250 \text{ mm}^2$ ) and a pulsed amperometric detector (Dionex PAD). Eluants A and B were 100 mM NaOH and M NaOAc with 100 mM NaOH, respectively. The mobile phase gradient was 0% B to 25% B over 60 min at a flow rate of 1.0 mL/min. Fractions containing diantennary oligosaccharides were collected, immediately neutralized with 2 M aqueous acetic acid, and then desalted by passing them through a Dowex 50 column ( $\text{H}^+$  form,  $10.0 \times 1.0 \text{ cm}^2$ ).

**Amino acid composition analysis and peptide sequencing.**—Amino acid composition analyses were carried out on an ABI Model 420 Derivatizer equipped with an ABI Model 172 HPLC System. Peptide sequences were determined with an ABI Model 476 Protein Sequencer using the Edman degradation procedure.

**Electrospray mass spectrometry.**—Positive-ion electrospray mass spectra were obtained on an API III System (PE SCIEX, Toronto, Canada). The samples (10–50 pmol/ $\mu\text{L}$ ) collected after reverse phase HPLC were introduced directly into the atmospheric pressure ionization (API) module by injection with a syringe driver, at a flow rate of 5  $\mu\text{L}/\text{min}$ .

**Sample preparation for NMR studies.**—Gp21 derived from HSTF lot number 50H93161 (7.6 mg, sample 1) and derived from HSTF lot number 72H9303 (15.0 mg, sample 2) were used for NMR studies. The Gp21 samples were each subjected twice to

deuterium exchange in D<sub>2</sub>O (99.9%) with 100 mM NH<sub>4</sub>HCO<sub>3</sub> (pD 8.0) and then dissolved in 550  $\mu$ l D<sub>2</sub>O (99.96%). The solutions were titrated with first 2 M DCl ( $\sim$  6  $\mu$ l) and then 0.1 M DCl ( $\sim$  6  $\mu$ l) to reach pD 4.6. Solutions of Gp21 in 90% H<sub>2</sub>O–10% D<sub>2</sub>O (v/v) with pH 4.6 were prepared in a similar fashion.

**NMR spectroscopy.**—All experiments were carried out on a Bruker AMX-600 spectrometer with <sup>1</sup>H and <sup>13</sup>C resonance frequencies of 600.14 and 150.91 MHz, respectively. Various homonuclear 2D <sup>1</sup>H–<sup>1</sup>H NMR experiments were performed to obtain <sup>1</sup>H resonance assignments of the peptide moiety of Gp21 (sample 1). These included DQ spectroscopy [24], DQF-COSY [25], TOCSY [26,27], and NOESY [28–30] in both 9:1 (v/v) H<sub>2</sub>O/D<sub>2</sub>O solution and D<sub>2</sub>O solution. In these experiments the <sup>1</sup>H carrier frequency was set at either the H<sub>2</sub>O or the residual HDO resonance and the spectral width was 6024.10 Hz. In DQ spectroscopy, 512 FIDs with 2048 complex data points and 32 scans each were acquired. The preparation period of the DQ coherences was either 25 or 50 ms and a 90° mixing pulse was used. For the DQF-COSY and TOCSY spectra, 1024 and 800 FIDs, respectively, with 1024 complex data points and 8 scans each were acquired. The MLEV17 sequence [27] flanked by 2.5 ms trim pulses was used for isotropic mixing. For the NOESY spectra, 512 FIDs with 2048 complex data points and 80 scans each were acquired with a 200 ms mixing time.

The HSQC [31] and HSQC-TOCSY [32,33] experiments were carried out on Gp21 (sample 2) in D<sub>2</sub>O. The <sup>1</sup>H and <sup>13</sup>C carrier frequencies were set at 2.75 and 79.07 ppm, respectively, with <sup>1</sup>H and <sup>13</sup>C spectral widths set to 3012.05 and 10500.00 Hz. The refocusing delay  $1/(2^1J_{CH})$  was set at 3.0 ms. The <sup>13</sup>C decoupling during acquisition was achieved using the GARP-1 sequence [34]. For the HSQC spectrum, 340 FIDs with 2048 complex data points and 128 scans each were acquired. For the HSQC-TOCSY spectra, 256 FIDs with 2048 complex data points and 160 or 240 scans each were acquired. The MLEV17 sequence flanked by 1.5 ms trim pulses was used for isotropic mixing with a duration of 20, 30, or 55 ms.

Most of the homonuclear and heteronuclear NMR experiments were carried out at 25 °C. Additional homonuclear experiments were conducted at 15 °C and 5 °C in order to resolve overlapping <sup>1</sup>H resonances. All spectra, except the DQ data, were recorded in a pure phase manner using the time-proportional phase incrementation (TPPI) method [35] in  $t_1$  and quadrature detection in  $t_2$ . The water resonance was suppressed by low-power presaturation during the relaxation delay (1.3–1.7 s) in most experiments.

Data were processed on an IRIS 4D/220GTX workstation (Silicon Graphics) using Felix 2.30 software (Biosym Technologies, San Diego, CA). Typically, the data were zero-filled once in the  $t_1$  dimension, and multiplied with a 60, 80, or 90° shifted squared sinebell function in the  $t_1$  dimension and a Gaussian line-broadening function ( $lb = -1.5$  Hz,  $gb = 0.015$  or  $0.05$ ) in the  $t_2$  dimension prior to Fourier transformation. For the DQF-COSY data, a shifted squared sinebell function was applied in both dimensions. For the heteronuclear 2D NMR data, an exponential line-broadening function ( $lb = 1.0$  Hz) was applied in the  $t_1$  dimension. The baseline correction technique FLATT [36] was applied to the NOESY and TOCSY spectra. The <sup>1</sup>H chemical shifts were referenced to internal acetone (2.225 ppm) and the <sup>13</sup>C chemical shifts were referenced indirectly to DSS with the use of the <sup>1</sup>H frequency of internal acetone [37,38]. As a result, the <sup>13</sup>C chemical shift of internal acetone is 32.85 ppm.

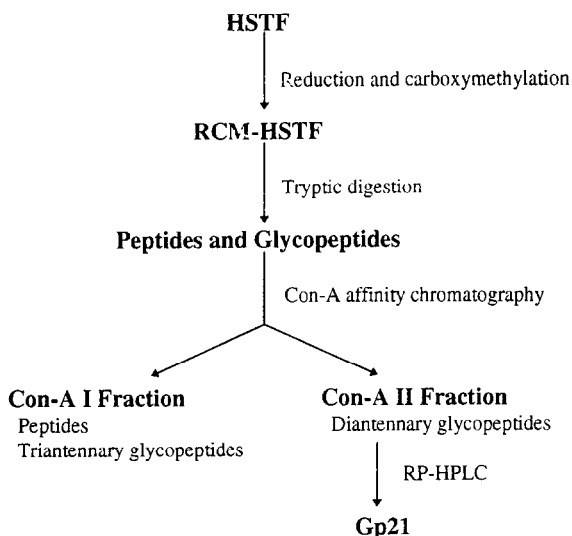


Fig. 2. The procedure used for the isolation and purification of Gp21 from HSTF.

### 3. Results

**Preparation of Gp21.**—The procedure for the isolation and purification of Gp21 is presented schematically in Fig. 2. A typical HPLC chromatogram of a Con-A II fraction containing the diantennary tryptic glycopeptides derived from HSTF is shown in Fig. 3. Each of the five major peaks labeled in Fig. 3 was collected and analyzed by electrospray mass spectrometry. In addition, the sugar-free peptides were obtained from these peaks by *N*-glycanase digestion, and analyzed by ES-MS. The five peptide sequences identified by ES-MS are summarized in Table 1. Gp21 was found in peak 3. A glycopeptide smaller than Gp21 by a mass of  $16.2 \pm 1.0$  was found in peak 4. Apparently, this was a Gp21 derivative in which the N-terminal Gln was converted to pyro-glutamate (pGlu) after release of ammonia (MW = 17.0). The glycopeptide found in peak 2 was the predicted product of a complete tryptic digestion corresponding to glycosylation site Asn413. The glycopeptide found in peak 5 was an elongated analog of peak 2 and is considered the result of incomplete tryptic digestion. The glycopeptide found in peak 1 must be the result of contaminating chymotrypsin activity because its C-terminus is Phe rather than Arg or Lys. The peptide sequencing results of Gp21 were in complete agreement with the reported sequence for residues 603–623 [13,14]. However, the sequencing of peak 4 was blocked in the first round, confirming that it had an unusual N-terminal amino acid residue such as pGlu that lacked a primary amine group.

Three different batches of commercial HSTF have been used to prepare Gp21. One glycoform was obtained for the glycan of Gp21 derived from batch I (HSTF lot number 67F-9454), whereas different glycoforms were obtained for the glycan of Gp21 derived from batches II and III (lot numbers 50H93161 and 72H9303, respectively). The presence of multiple glycoforms is evident in the electrospray mass spectra of Gp21

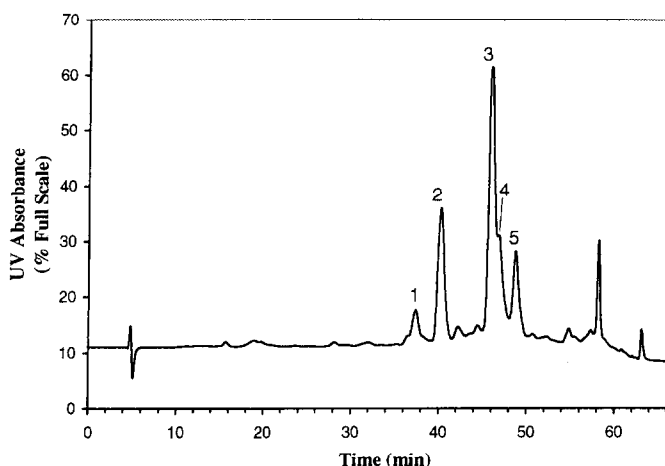


Fig. 3. RP-HPLC chromatogram of the Con A II fraction obtained from the tryptic digest of RCM-HSTF by Con A affinity chromatography.

shown in Fig. 4. The results of the ES-MS analyses of Gp21 are summarized in Table 2, in which the mass of the peptide fragment has been subtracted from the total molecular mass of Gp21 in order to more clearly reveal the presence of different glycoforms. Gp21 from batch I contains only one glycoform, namely the disialylated diantennary oligosaccharide (Glycoform A, see Table 2). However, Gp21 from batch II contains three glycoforms, namely the disialylated (Glycoform A), monosialylated (Glycoform B) and asialylated (Glycoform C) diantennary oligosaccharides. The relative amounts of Glycoforms A, B, and C in this preparation of Gp21 were determined by HPAEC to be 12%, 54% and, 34%, respectively. Gp21 from batch III contained five glycoforms. Besides Glycoforms A, B, and C, Glycoforms D (monosialylated) and E (asialylated) were found, both of which lacked one galactosyl residue.

Since the Gln and pGlu residue at the N-terminus can undergo chemical exchange and reach an equilibrium ratio of 1:1 under the chosen experimental conditions for NMR

Table 1

ES-MS analyses of the sugar-free peptides, released from the glycopeptides collected by RP-HPLC of the Con A-II fraction obtained from the tryptic digest of HSTF (lot number 50H93161)

Peak	Sequence fragment <sup>a</sup>	ES-MS	
		expected mass	measured mass <sup>b</sup>
1	Gln603–Phe619 <sup>c</sup>	1940.0	1939.4 ± 0.8
2	Phe402–Arg414	1477.7	1477.7 ± 0.4
3	Gln603–Arg623	2517.8	2517.4 ± 0.6
4	pGlu603–Arg623	2500.8	2501.2 ± 0.8
5	Phe402–Arg433	3533.9	3533.3 ± 0.6

<sup>a</sup> The sequence number convention is the same as that used by MacGillivray et al. [13].

<sup>b</sup> The measured mass and its standard deviation were calculated from two ions.

<sup>c</sup> This is not a tryptic fragment. See the text for the explanation.



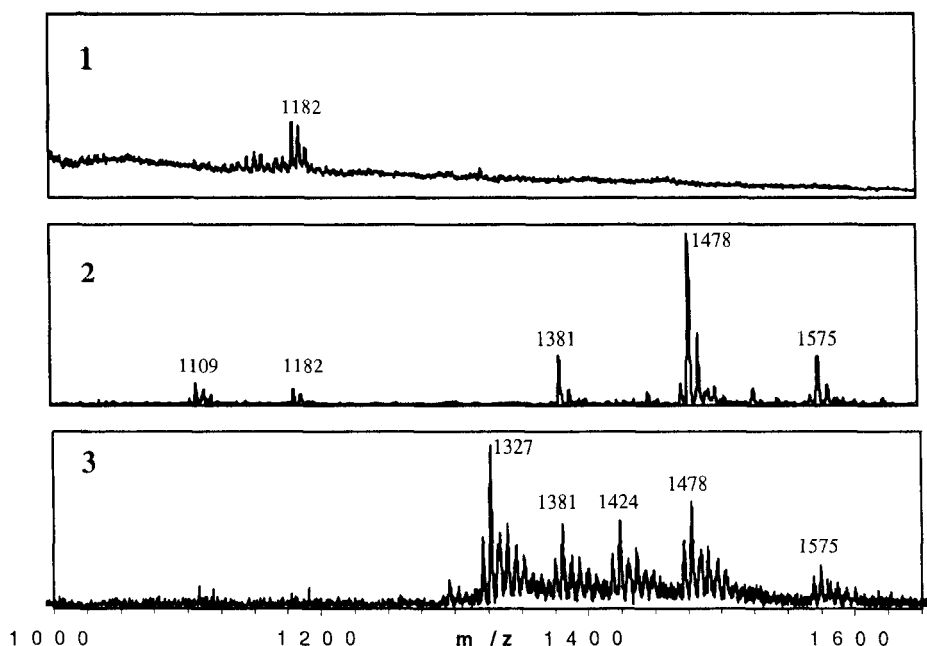


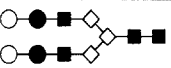
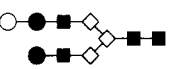
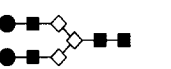
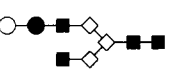
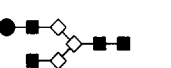
Fig. 4. The ES-MS spectra of Gp21 derived from three batches of commercial HSTF. Panels 1, 2 and 3 show the spectra of Gp21 derived from batches I, II and III, respectively. The  $m/z$  units are indicated for  $[M+4H]^{4+}$  ions (ranging from  $m/z$  1036.0 to 1182.0) and  $[M+3H]^{3+}$  ions (ranging from  $m/z$  1326.0 to 1575.0). The presence of one glycopeptide (MW  $4723.0 \pm 1.4$ ), three glycopeptides (MWs  $4723.7 \pm 0.4$ ,  $4431.5 \pm 0.7$ , and  $4140.0 \pm 1.5$ ), and five glycopeptides (MWs  $4722.4 \pm 1.0$ ,  $4429.6 \pm 1.0$ ,  $4140.1 \pm 1.0$ ,  $4269.4 \pm 1.0$ , and  $3977.8 \pm 1.0$ ), was inferred from the ions in panels 1, 2 and 3, respectively.

studies, peaks 3 and 4 were collected together and labeled as Gp21 in the HPLC preparation. For NMR studies, 7.6 mg (sample 1) and 15.0 mg (sample 2) of Gp21 were obtained from batches II and III, respectively.

**$^1H$  and  $^{13}C$  resonance assignments of Gp21.**—The 1D  $^1H$  NMR spectra of Gp21 recorded in  $H_2O$  and  $D_2O$ , at 25 °C, pH/pD 4.6, are shown in Fig. 5. In the spectrum recorded in  $D_2O$  (Fig. 5a), the anomeric protons of GlcNAc1, Man4 and Man4' and the two imidazole ring protons of His606 were resolved and could be readily recognized using the chemical shift values reported for corresponding glyco-asparagines [39] and a random coil tetrapeptide [40], respectively. Other protons that could be recognized were the sugar N-acetyl methyl protons and the NeuAc H3<sup>ax</sup>. It should be noted that there are two resonances for each anomeric proton of both Man4 and Man4', as is shown in Fig. 5c and d, due to the variability in the presence of the sialic acid at the end of the branch (cf. [39]). The assignment of the two signals for each anomeric proton was aided by the observation that there is a change in the intensity ratios from batch II to batch III. As is shown in Fig. 4, the sialic acid content in Gp21 derived from batch II was higher than in Gp21 from batch III. This difference is reflected by the change in the intensity ratios of the signals from Fig. 5c (sample 1, from batch II) to Fig. 5d (sample 2, from batch III). The intensity of the signal at lower field of each anomeric proton is smaller in batch III

Table 2

ES-MS results of different glycoforms pertaining to the N-linked diantennary oligosaccharide of Gp21 derived from three batches of commercial HSTF

Glycoform <sup>a</sup>	Mass <sup>pred.</sup>	Mass <sup>b</sup>		
		batch I, lot No. 67F-9454	batch II, lot No. 50H93161	batch III, lot No. 72H9303
A 	2224.0	2223.3 ± 1.4	2224.0 ± 0.4	2222.7 ± 1.0
B 	1932.7	N.D. <sup>c</sup>	1931.8 ± 0.7	1929.9 ± 1.0
C 	1641.4	N.D.	1640.3 ± 1.5	1640.4 ± 1.0
D 	1770.5	N.D.	N.D.	1769.7 ± 1.0
E 	1479.2	N.D.	N.D.	1478.1 ± 1.0

<sup>a</sup> The glycoforms have been drawn schematically, consisting of GlcNAc (■), Man (□), Gal (●) and NeuAc (○) residues.

<sup>b</sup> The mass was calculated by subtracting the peptide FW from the measured MW of Gp21.

<sup>c</sup> N.D. stands for not detected.

J. Lu, H. van Halbeek / Carbohydrate Research 296 (1996) 1–21

than in batch II. Therefore, it is concluded that both anomeric protons show a small downfield shift in the presence of a NeuAc residue at the end of the branch. These results are consistent with the data reported [39] for glycopeptides with the same glycoforms as those in sample 1.

The <sup>1</sup>H resonance assignments for the peptide chain were made following the standard homonuclear 2D NMR procedure as outlined by Wüthrich [40] (see Fig. 6). First, amino acid <sup>1</sup>H spin systems were identified in DQF-COSY and TOCSY spectra. For example, the schematic diagrams of a DQF-COSY (Fig. 6b) and a TOCSY (Fig. 6c) spectrum recorded in H<sub>2</sub>O show that the <sup>1</sup>H spin systems for two adjacent residues Val(*i*) and Thr(*i* + 1) are manifested by the cross peaks that correlate the NH with the C<sup>α</sup>H and all the side chain protons. Secondly, sequential connectivities between the identified <sup>1</sup>H spin systems were established based on the observation of sequential NOE cross peaks in a NOESY spectrum. The distances between some of the protons in two adjacent amino acid residues, such as the C<sup>α</sup>H of one residue and the NH of the next residue (denoted as *d*<sub>αN</sub>(*i*, *i* + 1) by Wüthrich [40]), are shorter than 5 Å, and thus give rise to NOEs. A schematic diagram of a NOESY spectrum recorded in H<sub>2</sub>O is shown in Fig. 6d, illustrating the sequential and intraresidue NOE cross peaks that correlate C<sup>α</sup>H of Val(*i*) with NH of Thr(*i* + 1).

The fingerprint region (that is, the backbone NH–C<sup>α</sup>H region) of a DQF-COSY spectrum of Gp21 recorded in H<sub>2</sub>O is shown in Fig. 7a. A total of 33 cross peaks were

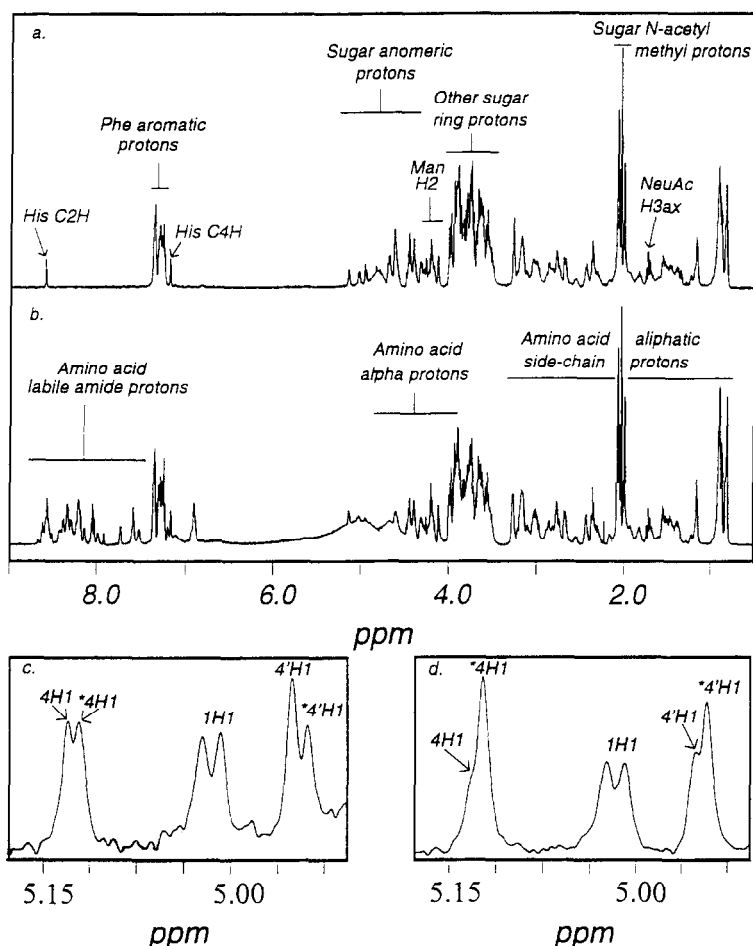


Fig. 5. The 1D  $^1\text{H}$  NMR spectra of Gp21. Panels a and b show the full spectra of Gp21 derived from batch II, recorded in  $\text{D}_2\text{O}$  and  $\text{H}_2\text{O}$ , respectively. Panels c and d show the sugar anomeric regions of Gp21 derived from batches II and III, respectively, recorded in  $\text{D}_2\text{O}$ . The numbers before and after H represent the sugar residue and the proton position in the ring, respectively. An asterisk (\*) indicates that a sugar residue is located in a branch without a terminal sialic acid residue. This convention is used in all figures in this paper.

observed, including 26 backbone  $\text{NH}-\text{C}^\alpha\text{H}$  cross peaks and 7 sugar amide proton to sugar ring proton cross peaks. There are two cross peaks for each of the five consecutive amino acid residues at the N-terminal region, Gln604 to Phe608. This doubling is due to the presence of pGlu, in addition to Gln, at the N-terminus in a portion of the sample. The extended fingerprint region of a TOCSY spectrum of Gp21 recorded in  $\text{H}_2\text{O}$  is shown in Fig. 7b, where the range in the  $F_1$  domain covers the amino acid alpha and aliphatic side chain proton chemical shifts and the sugar proton chemical shifts. In favorable cases, all the proton signals of an amino acid residue could be identified in this region. Furthermore, the  $^1\text{H}$  spin systems for unique amino acid residues, such as Val612, Thr613 and Arg623 were identified.

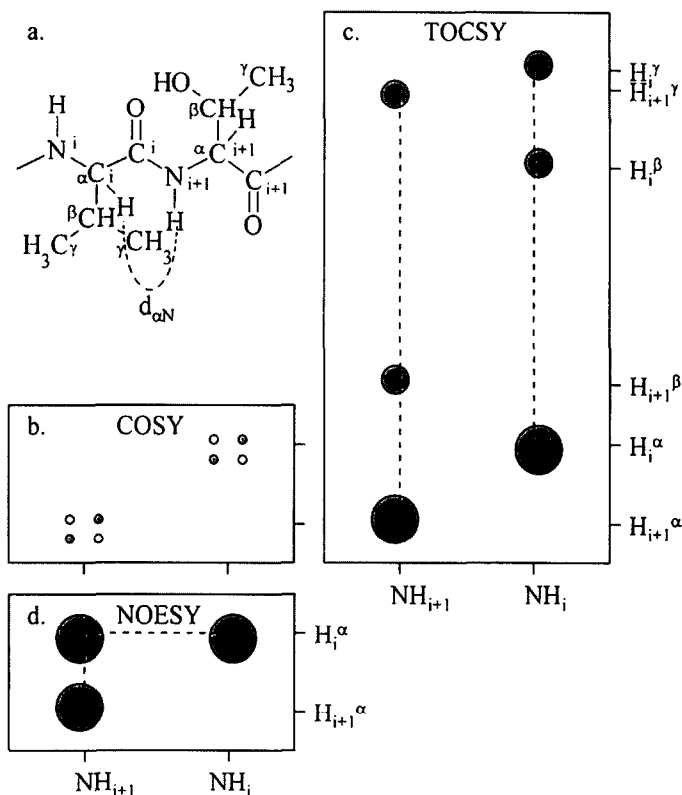


Fig. 6. (a) A dipeptide segment of a polypeptide chain consisting of Val( $i$ ) and Thr( $i+1$ ). The sequential NOE,  $d_{\alpha N}(i, i+1)$ , between Val( $i$ ) C $^{\alpha}$ H and Thr( $i+1$ ) NH, is indicated with a dashed line. (b–d) Schematic diagrams of a DQF-COSY, TOCSY and NOESY spectrum recorded in H<sub>2</sub>O showing through- ( $J$  coupling) and through-space correlations between the protons in Val( $i$ ) and Thr( $i+1$ ), respectively. The cross peaks in the DQF-COSY spectrum have antiphase characteristics, and those in the TOCSY and NOESY spectrum have in-phase characteristics. Here, the filled and open circles represent positive and negative cross peaks, respectively.

Sequential NOE connectivities, namely  $d_{\alpha N}(i, i+1)$ ,  $d_{NN}(i, i+1)$ , and  $d_{\beta N}(i, i+1)$  [in case of Ile, Thr, and Val also  $d_{\gamma N}(i, i+1)$ ], were utilized to establish the sequential connectivities of the amino acid  $^1\text{H}$  spin systems. The fingerprint region of a NOESY spectrum of Gp21 recorded in H<sub>2</sub>O is shown in Fig. 7c. To illustrate the sequential assignment procedure, the connectivity pathway of the peptide fragment from Ser610 to Asn618 is indicated with solid lines. The  $d_{\alpha N}(i, i+1)$  connectivity pathway ran across the entire peptide sequence, whereas the  $d_{NN}(i, i+1)$  connectivity was found in five peptide fragments, Gly609 to Thr613, Asp614 to Cys615, Ser616 to Asn618, Phe619 to Cys620, and Leu621 to Arg623. The complete sequential assignment, especially for Gln603 to His606, was aided by data collected at lower temperatures, i.e., at 15 °C and 5 °C. A complete list of the amino acid  $^1\text{H}$  chemical shifts of Gp21 measured at 25 °C has been compiled in Table 3. The subsequent assignment of the  $^1\text{H}$ -attached  $^{13}\text{C}$

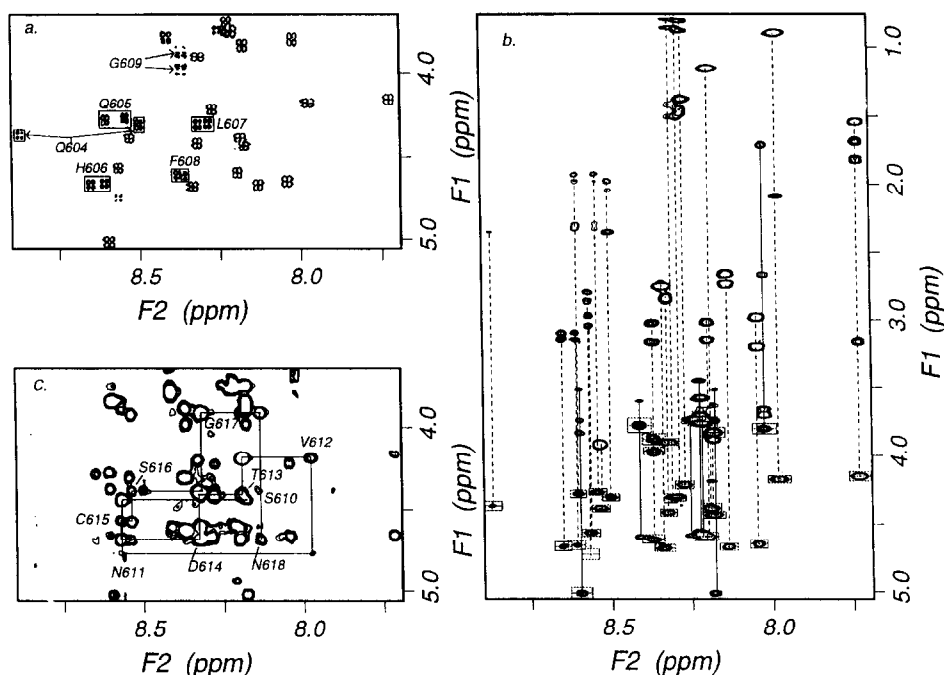


Fig. 7. (a) The fingerprint region (i.e. the backbone  $\text{NH}-\text{C}^\alpha\text{H}$  region) of a DQF-COSY spectrum of Gp21 recorded in  $\text{H}_2\text{O}$  with water presaturation. Two cross peaks for each of the residues Q604, Q605, H606, L607, and F608 are highlighted with boxes. (b) The extended fingerprint region of a TOCSY spectrum recorded in  $\text{H}_2\text{O}$ . The  $\text{NH}-\text{C}^\alpha\text{H}$  cross peaks which also occur in panel a are highlighted with footprint boxes. The amino acid and the monosaccharide  $^1\text{H}$  spin systems are indicated with dashed and solid lines, respectively. (c) Region of the NOESY spectrum corresponding to panel a, recorded in  $\text{H}_2\text{O}$  with water presaturation. The sequential  $d_{\alpha\text{N}}(i, i+1)$  connectivities for the fragment from Ser(S)610 to Asn(N)618 are indicated with solid lines.

resonances of the peptide chain by the HSQC and HSQC-TOCSY spectra was straightforward because all pertinent proton resonances had been assigned. A complete list of the amino acid  $^{13}\text{C}$  chemical shifts of Gp21 measured at  $25^\circ\text{C}$  has been compiled in Table 4.

The complete  $^1\text{H}$  resonance assignment of the carbohydrate moiety has been achieved in a procedure similar to that used for the peptide chain. The anomeric protons (H-1), the H-2 proton of a Man residue, and the two H-3 protons (H-3<sup>ax</sup> and H-3<sup>eq</sup>) of a NeuAc residue as well as the N-acetyl methyl protons of the GlcNAc and NeuAc residues are the so called structural reporter groups [39], and most of their chemical shifts were assigned from 1D spectra. Due to overlap with peptide signals, however, complete analysis required 2D methods.

An expanded region of the carbohydrate portion of a TOCSY spectrum of Gp21 recorded in  $\text{D}_2\text{O}$  is shown in Fig. 8a. Many of the resonances were assigned from traces at the chemical shifts of the structural reporter signals, as shown in Fig. 8a. However, signal overlap in these subspectra led to some ambiguities in the resonance assignments.

Table 3

The peptide  $^1\text{H}$  chemical shifts of Gp21 at 25 °C, pH 4.6 <sup>a</sup>

Residue		NH	C $^\alpha$ H	C $^\beta$ H	Others
Gln	603		4.11	2.16, 2.16	C $^\gamma$ H <sub>2</sub> , 2.44, 2.44
pGlu	603		4.38	2.54, 2.05	C $^\gamma$ H <sub>2</sub> , 2.42, 2.42
Gln	604	8.89	4.37	2.04, 1.97	C $^\gamma$ H <sub>2</sub> , 2.35, 2.35 N $^\gamma$ H <sub>2</sub> , 7.57, 6.91
	604 <sup>b</sup>	8.51	4.31	2.03, 1.96	n.c.
Gln	605	8.62	4.28	1.98, 1.92	C $^\gamma$ H <sub>2</sub> , 2.32, 2.28 N $^\gamma$ H <sub>2</sub> , 7.51, 6.88
	605 <sup>b</sup>	8.56	4.27	1.99, 1.92	C $^\gamma$ H <sub>2</sub> , 2.32, 2.29 N $^\gamma$ H <sub>2</sub> , n.c.
His	606	8.66	4.66	3.15, 3.09	C <sup>2</sup> H, 8.56; C <sup>4</sup> H, 7.15
	606 <sup>b</sup>	8.63	4.66	n.c.	n.c.
Leu	607	8.33	4.31	1.50, 1.42	C $^\gamma$ H, 1.45; C $^\delta$ H <sub>3</sub> , 0.87, 0.80
	607 <sup>b</sup>	8.31	n.c.	n.c.	n.c.
Phe	608	8.38	4.60	3.15, 3.03	C <sup>2,6</sup> H, 7.24; C <sup>3,5</sup> H, 7.33; C <sup>4</sup> H, 7.28
	608 <sup>b</sup>	8.37	n.c.	n.c.	n.c.
Gly	609	8.37	3.98, 3.86		
Ser	610	8.18	4.44	3.89, 3.82	
Asn	611	8.57	4.75	2.86, 2.80	N $^\delta$ H, 8.62
Val	612	7.98	4.18	2.09	CH <sub>3</sub> , 0.90, 0.90
Thr	613	8.20	4.39	4.18	CH <sub>3</sub> , 1.15
Asp	614	8.33	4.68	2.75, 2.75	
Cys	615	8.59	4.57	3.05, 2.97	CM-CH <sub>2</sub> , 3.26 <sup>c</sup>
Ser	616	8.55	4.38	3.93, 3.89	
Gly	617	8.32	3.90, 3.90		
Asn	618	8.13	4.67	2.74, 2.66	N $^\delta$ H <sub>2</sub> , 7.58, 6.90
Phe	619	8.21	4.60	3.15, 3.02	C <sup>2,6</sup> H, 7.23; C <sup>3,5</sup> H, 7.33; C <sup>4</sup> H, 7.28
Cys	620	8.33	4.42	2.86, 2.82	CM-CH <sub>2</sub> , 3.25 <sup>c</sup>
Leu	621	8.29	4.22	1.45, 1.37	CH, 1.51; CH <sub>3</sub> , 0.88, 0.81
Phe	622	8.05	4.65	3.20, 2.98	C <sup>2,6</sup> H, 7.24; C <sup>3,5</sup> H, 7.33; C <sup>4</sup> H, 7.28
Arg	623	7.70	4.16	1.82, 1.68	C $^\gamma$ H <sub>2</sub> , 1.54, 1.54; CH <sub>2</sub> , 3.16 N $^\epsilon$ H, 7.19

n.c.: The shorthand notation of no change due to the presence of pGlu instead of Gln.

<sup>a</sup> Chemical shifts are in parts per million (ppm) referenced to internal acetone (2.225 ppm) and were measured with 0.005 ppm/pt digital resolution.<sup>b</sup> Amino acid residue whose  $^1\text{H}$  chemical shifts may be affected by the presence of pGlu instead of Gln at the N-terminus.<sup>c</sup> The carboxymethyl proton chemical shift of the carboxymethylated Cys.

In addition, some signals are missing in these subspectra, such as the H-5, H-6 and H-6' of Gal6 and Gal6', due to a small coupling constant (< 1.5 Hz) between Gal H-4 and H-5. These problems were solved by performing 2D  $^1\text{H}$ -detected heteronuclear experiments. Fig. 9a shows an expanded region of an HSQC  $^1\text{H}$ - $^{13}\text{C}$  one-bond correlation spectrum of Gp21 recorded in D<sub>2</sub>O. Fig. 9b shows the corresponding expanded region of an HSQC-TOCSY spectrum. Because the HSQC-TOCSY experiment employs the same isotropic mixing technique as the homonuclear TOCSY experiment, the interpretation of the spectrum is very similar. However, sugar  $^1\text{H}$  signals in the HSQC-TOCSY

Table 4  
The peptide  $^{13}\text{C}$  chemical shifts of Gp21 at 25 °C, pH 4.6 <sup>a</sup>

Residue		C $^{\alpha}$	C $^{\beta}$	C $^{\gamma}$	Others
Gln	603	55.34	29.57	33.11	
pGlu	603	59.82	28.21	32.29	
Gln	604	56.22	29.99	33.93	
	604 <sup>b</sup>	56.29	29.78	34.07	
Gln	605	56.15	29.99	34.07	
His	606	55.20	29.24		C $^2$ , 136.58; C $^4$ , 120.33
Leu	607	55.27	42.77	27.20	C $^{\delta}$ , 23.87; C $^{\delta'}$ , 25.16
Phe	608	58.12	40.05		C $^{2,6}$ , 132.30; C $^{3,5}$ , 131.82; C $^4$ , 130.19
Gly	609	45.40			
Ser	610	58.53	64.24		
Asn	611	53.10	39.24		
Val	612	62.61	33.12	20.61, 21.56	
Thr	613	61.93	70.18	21.77	
Asp	614	54.11	40.99		
Cys	615	56.42	36.04		CM-C, 39.77 <sup>c</sup>
Ser	616	59.48	64.04		
Gly	617	45.61			
Asn	618	53.29	39.10		
Phe	619	58.19	39.85		C $^{2,6}$ , 132.30; C $^{3,5}$ , 131.82; C $^4$ , 130.19
Cys	620	55.95	36.24		CM-C, 39.77 <sup>c</sup>
Leu	621	55.67	42.77	27.20	C $^{\delta}$ , 23.66; C $^{\delta'}$ , 25.16
Phe	622	57.71	39.51		C $^{2,6}$ , 132.30; C $^{3,5}$ , 131.82; C $^4$ , 130.19
Arg	623	57.51	31.96	27.34	C $^{\delta}$ , 43.72

<sup>a</sup> Chemical shifts are in ppm referenced to internal acetone (32.85 ppm) and were measured with 0.07 ppm/pt digital resolution.

<sup>b</sup> Amino acid residue whose  $^{13}\text{C}$  chemical shifts may be affected by the presence of pGlu instead of Gln at the N-terminus.

<sup>c</sup> The carboxymethyl carbon chemical shift of the carboxymethylated Cys.

spectrum are projected onto the  $^{13}\text{C}$  chemical shift axis (the  $F_1$  domain in Fig. 9b) rather than the  $^1\text{H}$  chemical shift axis. Because the  $^{13}\text{C}$  chemical shifts of the carbohydrates have a much larger dispersion than the  $^1\text{H}$  chemical shifts, more subspectra are resolved in the HSQC-TOCSY spectrum than in the TOCSY spectrum. A few subspectra are indicated in Fig. 9b with either solid or dashed lines and labels, including those at C-5 and C-6 of Gal6 and 6', and those at C-3, C-4, C-5, and C-6 of Man4. Because of the rather low sensitivity of the HSQC-TOCSY technique, the subspectra are usually incomplete. Nonetheless, the enhanced resolution in the  $F_1$  domain of the HSQC and HSQC-TOCSY spectra enabled us to overcome the obstacle of the  $^1\text{H}$  resonance overlaps of the Gp21 carbohydrate.

Analysis of interresidue NOE cross peaks allowed us to identify doubling of the resonances of the GlcNAc $\beta$ (1  $\rightarrow$  2) and Gal anomeric protons. These resonances, like those of Man4 and Man4' anomeric protons mentioned previously, are affected by the presence of sialic acid in the branches. In addition, we have been able to determine which branch of Glycoforms D and E (see Table 2) lacked a Gal residue by using the sequential NOE cross peaks. Fig. 8b shows an expanded region of a NOESY spectrum

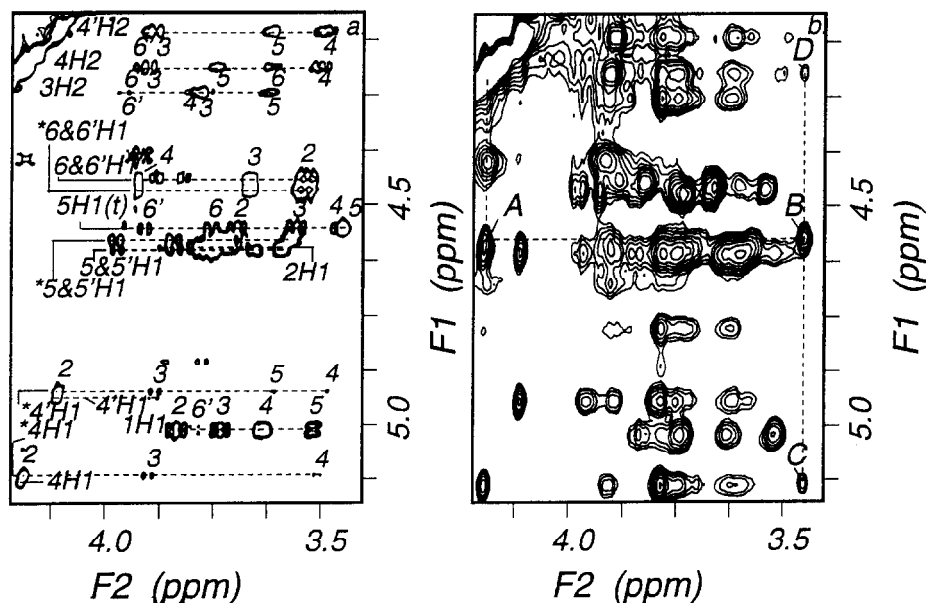


Fig. 8. (a) Part of a TOCSY spectrum of Gp21 showing the sugar residue subspectra (indicated by dashed lines) at their anomeric proton chemical shifts. (b) The same region of a NOESY spectrum of Gp21. See the text for details of the assignment of cross peaks A, B, C, and D. Notice that the digital resolution in panel b is only half of that of panel a in both dimensions.

recorded in D<sub>2</sub>O covering the same region as the TOCSY spectrum shown in Fig. 8a. As indicated with dashed lines and labels, cross peak A shows an NOE contact between H-1 of the terminal GlcNAc residue in the branch that lacks a Gal residue and H-2 of Man<sub>4</sub>; cross peaks C and D show NOE contacts between H-5 of this terminal GlcNAc residue and H-1 and H-2 of Man<sub>4</sub>, respectively. Cross peak B shows an NOE contact between H-1 and H-5 of the GlcNAc residue. However, no interresidue NOE cross peaks were observed between the terminal GlcNAc and any Man<sub>4</sub>' resonances. Therefore, Glycoforms D and E lack the galactosyl residue in their  $\alpha(1 \rightarrow 3)$  branch. The complete lists of the carbohydrate <sup>1</sup>H and <sup>13</sup>C chemical shifts of Gp21 measured at 25 °C have been compiled in Tables 5 and 6, respectively.

#### 4. Discussion

The combination of tryptic digestion, Con A-agarose affinity chromatography, and reverse phase HPLC proved to be an effective way to generate and purify Gp21 in multi-milligram amounts. For example, we were able to isolate a total of 15 mg of Gp21 from the tryptic digests of RCM-HSTF after the multi-step chromatographic purification, in which we started from about 190 mg of the tryptic digest for each of the total four runs. It should be noted that, despite the use of TPCK-treated trypsin to suppress chymotryptic activity, a considerable amount of a chymotryptic glycopeptide (peak 1 in



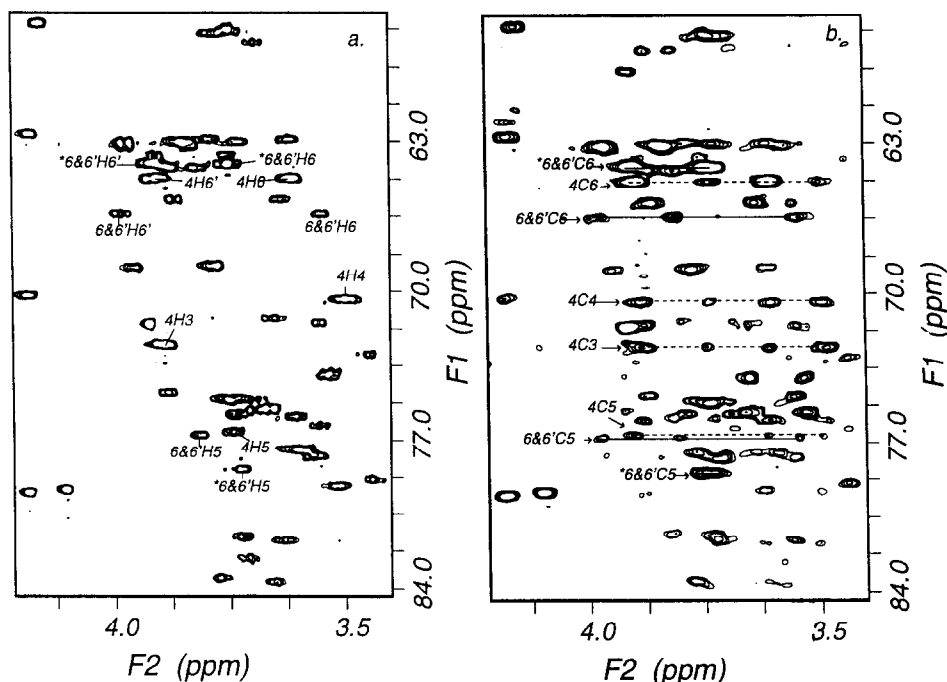


Fig. 9. Corresponding regions of a HSQC spectrum (a) and a HSQC-TOCSY spectrum (b) showing the carbohydrate  $^1\text{H}$  spin systems of Gp21 at different  $^{13}\text{C}$  chemical shifts. The HSQC-TOCSY spectrum was generated by co-addition of the spectra recorded with three different mixing times.

Fig. 3) was generated. We therefore expect that the yield of Gp21 could be further improved by optimization of the experimental conditions, and by purification of the trypsin.

It is known that the glutamine side-chain amino group is much more labile than that of its homologue, asparagine [41]. Thus, glutamine can readily undergo intramolecular cyclization in a weak alkaline or acid solution by releasing  $\text{NH}_3$ . Similarly, the Gln residue at the N-terminal position in a peptide can lose  $\text{NH}_3$  under these conditions. However, Gln residues in the C-terminal position of a peptide and within a polypeptide chain are more stable. Such a conversion of Gln to pGlu has also been noticed by MacGillivray et al. [13] in the peptide fragment T20 (Gln603–Cys615) derived from the CNBr-1 fragment of HSTF, based on amino acid composition and sequencing analyses. Under our tryptic digestion conditions, i.e. pH 8.0, we observed a 20% conversion of Gln603 to pGlu603 after a 12 h digestion. Up to 80% conversion occurred after prolonged incubation at 37 °C with pH 8.0. However, the equilibrium ratio of Gln603 and pGlu603, as determined by  $^1\text{H}$  NMR, was about 1:1 at pH 4.6. The partial conversion (50%) of Gln to pGlu at the N-terminus doubles the number of backbone amide NH resonances of the adjacent five residues. However, there is little or no change in chemical shift of the corresponding alpha and side-chain protons of these amino acid residues, as shown in Table 3. Therefore, it appears that the Gln  $\rightarrow$  pGlu interconversion at the peptide N-terminus should not affect our conformational analysis of Gp21.

Table 5

The carbohydrate  $^1\text{H}$  chemical shifts of Gp21 at 25 °C, pH 4.6 <sup>a</sup>

Residue	H-1	H-2	H-3	H-4	H-5	H-6	H-6'			NH	H(Ac)
GlcNAc1	5.01	3.83	3.73	3.63	3.51	3.62	3.79			8.18	1.99
GlcNAc2	4.59	3.79	3.77	3.71	3.59	3.73	3.87			8.41	2.07
Man3	4.75	4.24	3.77	3.78	3.62	3.79	3.96				
Man4	5.13	4.19	3.90	3.50	3.74	3.61	3.92				
* Man4	5.12	4.19	3.90	3.50	3.74	3.61	3.92				
Man4'	4.94	4.10	3.89	3.49	3.60	3.62	3.90				
* Man4' <sup>b</sup>	4.92	4.10	3.89	3.49	3.60	3.62	3.90				
GlcNAc5	4.60	3.75	3.74	3.65	3.59	3.84	3.98			8.25	2.06
GlcNAc5'	4.60	3.75	3.74	3.65	3.59	3.84	3.98			8.21	2.06
* GlcNAc5	4.59	3.74	3.71	3.72	3.56	3.84	3.98			8.24	2.06
* GlcNAc5'	4.59	3.74	3.71	3.72	3.56	3.84	3.98			8.21	2.06
GlcNAc5 (t) <sup>c</sup>	4.55	3.69	3.55	3.45	3.44	3.75	3.91			8.22	2.06
Gal6	4.44	3.53	3.66	3.92	3.81	3.55	3.98				
Gal6'	4.44	3.53	3.66	3.92	3.81	3.55	3.98				
* Gal6	4.46	3.54	3.66	3.92	3.72	3.75	3.92				
* Gal6'	4.46	3.54	3.66	3.92	3.72	3.75	3.92				
	H-3 <sup>ax</sup>	H-3 <sup>eq</sup>	H-4	H-5	H-6	H-7	H-8	H-9	H-9'	NH	H(Ac)
NeuAcN and N'	1.74	2.68	3.65	3.80	3.69	3.55	3.88	3.63	3.86	8.03	2.03

<sup>a</sup> Chemical shifts are in ppm referenced to internal acetone (2.225 ppm) and measured with 0.005 ppm/pt digital resolution.

<sup>b</sup> The asterisk (\*) is to indicate that a sugar residue is located in the branch that lacks a sialic acid residue.

<sup>c</sup> The terminal GlcNAc5 residue located in the  $\alpha(1 \rightarrow 3)$  branch of the diantennary oligosaccharide.

The disialylated diantennary oligosaccharide (Glycoform A, see Table 2) has been reported to be the most abundant member of the N-linked diantennary oligosaccharide family from HSTF [18–20]. In this study, this was indeed confirmed for Gp21 derived from HSTF batch I, but was not found to be the case for batches II and III. One possible cause of this discrepancy is the hydrolysis of the acid-labile sialic acid under the acidic conditions used for reduction and carboxymethylation and reverse phase HPLC separation (see Experimental). However, all our experiments for batches II and III were carried out with the same procedure and experimental conditions as for batch I, so we do not consider the finding of asialo carbohydrates in batches II and III the result of a systematic error. To prove this, we released the oligosaccharides from intact HSTF by using a procedure that avoided acidic conditions. Thus, *N*-glycanase digestion of the intact HSTF from batch II was carried out, and the released diantennary oligosaccharides were isolated by Con A affinity chromatography and subsequently fractionated by Dionex HPAEC. The ratio of the resulting Glycoforms A, B and C (see Table 2) was similar to that obtained by our standard procedure. Therefore, it was confirmed experimentally that the mildly acidic conditions used in our standard procedure of preparing Gp21 (Fig. 2) did not lead to a significant loss of sialic acid. This is consistent

Table 6

The carbohydrate  $^{13}\text{C}$  chemical shifts of Gp21 at 25 °C, pH 4.6 <sup>a</sup>

Residue	C-1	C-2	C-3	C-4	C-5	C-6	C(Me)
GlcNAc1	81.31	56.69	75.81	81.72	79.21	62.88	25.21
GlcNAc2	104.37	57.92	75.06	82.65	77.51	63.02	25.48
Man3	103.47	73.22	83.49	68.73	77.44	68.87	
Man4	102.67	79.48	72.47	70.36	76.62	64.72	
* Man4 <sup>b</sup>	102.67	79.48	72.47	70.36	76.62	64.72	
Man4'	100.15	79.34	72.47	70.36	75.87	64.72	
* Man4'	100.09	79.34	72.47	70.36	75.87	64.72	
GlcNAc5	102.38	57.78	75.13	83.69	77.71	63.09	25.48
GlcNAc5'	102.38	57.78	75.13	83.69	77.71	63.09	25.48
* GlcNAc5	102.53	57.78	75.13	81.59	77.51	63.09	25.48
* GlcNAc5'	102.53	57.78	75.13	81.59	77.51	63.09	25.48
GlcNAc5 (i) <sup>c</sup>	102.74	58.33	76.35	72.95	78.87	63.63	25.48
Gal6	106.61	73.90	75.53	71.52	76.76	66.35	
Gal6'	106.61	73.90	75.53	71.52	76.76	66.35	
* Gal6	106.07	73.97	75.53	71.52	78.39	64.04	
* Gal6'	106.07	73.97	75.53	71.52	78.39	64.04	
	C-3	C-4	C-5	C-6	C-7	C-8	C-9
NeuAcN and N'	43.04	71.25	54.93	75.60	71.45	74.79	65.67
							C(Me)
							25.48

<sup>a</sup> Chemical shifts are in ppm referenced to internal acetone (32.85 ppm) and measured with 0.07 ppm/pt digital resolution.

<sup>b</sup> The asterisk ( \* ) is to indicate that a sugar residue is located in a branch that lacks a sialic acid residue.

<sup>c</sup> The terminal GlcNAc5 residue located in the  $\alpha(1 \rightarrow 3)$  branch of the diantennary oligosaccharide.

with the results of Rohrer and Townsend [42] who described their study on the stability of NeuAc $\alpha(2 \rightarrow 6)$  and  $\alpha(2 \rightarrow 3)$  linked to Gal and  $\alpha(2 \rightarrow 6)$  linked to GlcNAc during RP-HPLC. They found that there was no loss of sialic acid when glycopeptides were prepared by RP-HPLC using an eluant that contains 0.1% TFA, and were dried immediately.

The lack of the galactose residue in the  $\alpha(1 \rightarrow 3)$  branch of Glycoforms D and E in Gp21 derived from batch III may provide some clues to understand the microheterogeneity of the diantennary oligosaccharide in our samples. It is well known that the glycosidic linkage of the Gal residue in the diantennary oligosaccharide is very stable under mildly acidic and basic conditions at ambient temperature. Therefore, it is likely that the microheterogeneity of the glycan chain as found in the glycopeptide Gp21 was present in the commercial HSTF used as starting material.

Despite the heterogeneity of Gp21 in the N-terminus of the peptide and in the peripheral portion of the carbohydrate, we succeeded in the complete  $^1\text{H}$  and  $^{13}\text{C}$  resonance assignments of each glycoform present in Gp21. We have compared the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts listed in Tables 5 and 6 with those of an asialylated diantennary octasaccharide (Octa) and a disialylated diantennary deca-saccharide (Deca) reported by

Wieruszkeski et al. [43]. Octa and Deca have the same structures as the oligosaccharides of Glycoforms C and A (see Table 2), respectively, except that both lack GlcNAc1 and have a reducing GlcNAc2. The  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts of the GlcNAc5,5' and Gal6,6' residues in the two branches of Gp21 are degenerate, and some of them are affected by the presence of sialic acid. This is in agreement with the observations for Octa and Deca. In addition, the presence of sialic acid does not have a noticeable effect on the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts of the five residues in the glycan core of Gp21 except for the anomeric signals of Man4 and Man4'. This is also similar to the data reported for Octa and Deca. Finally, all of the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts values for the Gp21 glycan are very similar to those of Octa and Deca, except those for GlcNAc2 C-1, C-2 and C-3, because GlcNAc2 is the reducing-end sugar in Octa and Deca. For example, the typical differences in corresponding  $^1\text{H}$  chemical shift values between Gp21 and Octa/Deca are less than 0.02 ppm. These similarities suggest that the Gp21 peptide chain does not have a profound effect on the electronic environment of nuclei in the sugar residues beyond GlcNAc1.

In summary, a combination of enzymatic digestion, lectin affinity chromatography and reverse phase HPLC can be an effective method for the isolation and purification of glycopeptides with different sizes in quantities sufficient for their conformational studies by NMR. The homogeneity of the glycopeptide depends on a number of factors, such as the peptide sequence, the purity and specificity of the enzyme, and the microheterogeneity of the glycan present in the starting glycoprotein. Superficial analysis of the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts of the free and bound diantennary oligosaccharides shows that the Gp21 oligosaccharides are not drastically perturbed by the presence of the peptide fragment. In future studies, we will address the details of the conformation and dynamics of Gp21.

## Acknowledgements

We thank Dr. Lihua Huang for his help in collecting and interpreting the ES-MS spectra, Dr. John Glushka for assistance in NMR spectroscopy, and Dr. John Glushka and Dr. William York for their critical review of the manuscript. We thank Dr. Daotian Fu and Dr. Junko Hihara (Perkin Elmer ABI Division, Foster City, CA) for performing the amino acid composition analysis and the peptide sequencing. This research is supported by the NIH Resource Center Grant for Biomedical Complex Carbohydrates (P41-RR05351).

## References

- [1] A. Kobata, *Eur. J. Biochem.*, 209 (1992) 483–501.
- [2] A. Varki, *Glycobiology*, 3 (1993) 97–130.
- [3] S.W. Homans, R.A. Dwek, and T.W. Rademacher, *Biochemistry*, 26 (1987) 6571–6578.
- [4] J.P. Carver, *Curr. Opin. Struct. Biol.*, 1 (1991) 716–720.
- [5] H. Kessler, H. Matter, G. Gemmecker, A. Kling, and M. Kottenhahn, *J. Am. Chem. Soc.*, 113 (1991) 7550–7563.

- [6] M.R. Wormald, E.W. Wooten, R. Bazzo, C.J. Edge, and A. Feinstein, *Eur. J. Biochem.*, 198 (1991) 131–139.
- [7] J.T. Davis, S. Hirani, C. Bartlett, and B.R. Reid, *J. Biol. Chem.*, 269 (1994) 3331–3338.
- [8] R. Brockbank and H.J. Vogel, *Biochemistry*, 29 (1990) 5574–5583.
- [9] H. van Halbeek, *Curr. Opin. Struct. Biol.*, 4 (1994) 697–709.
- [10] H. Paulsen, *Angew. Chem. Int. Ed. Engl.*, 29 (1990) 823–939.
- [11] D.F. Wyss, J.S. Choi, and G. Wagner, *Biochemistry*, 34 (1995) 1622–1634.
- [12] D.F. Wyss, J.S. Choi, J. Li, M.H. Knoppers, K.J. Willis, A.R.N. Arulanandam, A. Smolyar, D.Z. Reinherz, and G. Wagner, *Science*, 269 (1995) 1273–1278.
- [13] R.T.A. MacGillivray, E. Mendez, J.G. Shewale, S.K. Sinha, J. Lineback-Zins, and K. Brew, *J. Biol. Chem.*, 258 (1983) 3543–3553.
- [14] F. Yang, J.B. Lum, J.R. McGill, C.M. Moor, S.L. Naylor, P.H. van Bragt, W.D. Baldwin, and B.H. Bowman, *Proc. Natl. Acad. Sci. USA*, 81 (1984) 2752–2756.
- [15] G. Spik, B. Bayard, B. Fournet, G. Strecker, S. Bouquelet, and J. Montreuil, *FEBS Lett.*, 50 (1975) 296–299.
- [16] L. März, M.W.C. Hatton, L.R. Berry, and E. Regoeczi, *Can. J. Biochem.*, 60 (1982) 624–630.
- [17] G. Spik, V. Debryne, J. Montreuil, H. van Halbeek, and J.F.G. Vliegthart, *FEBS Lett.*, 183 (1985) 65–69.
- [18] D. Leger, B. Campion, J.P. Decottignies, J. Montreuil, and G. Spik, *Biochem. J.*, 257 (1989) 231–238.
- [19] K. Yamashita, N. Koide, T. Endo, Y. Iwaki, and A. Kobata, *J. Biol. Chem.*, 264 (1989) 2415–2423.
- [20] D. Fu and H. van Halbeek, *Anal. Biochem.*, 206 (1992) 53–63.
- [21] P. de Waard, B.R. Leeftang, J.F.G. Vliegthart, R. Boelens, G.W. Vuister, and R. Kaptein, *J. Biomol. NMR*, 2 (1992) 211–226.
- [22] J. Mazurier, M. Dauchez, G. Vergoten, J. Montreuil, and G. Spik, *Glycoconjugate J.*, 8 (1991) 390–399.
- [23] G. Wagner, *J. Biomol. NMR*, 3 (1993) 375–385.
- [24] T.H. Mareci and R. Freeman, *J. Magn. Reson.*, 51 (1983) 531–538.
- [25] A. Derome and M. Williamson, *J. Magn. Reson.*, 88 (1990) 177–185.
- [26] L. Braunschweiler and R.R. Ernst, *J. Magn. Reson.*, 53 (1983) 521–528.
- [27] A. Bax and D.G. Davis, *J. Magn. Reson.*, 65 (1985) 355–360.
- [28] J. Jeener, B.H. Meier, P. Bachmann, and R.R. Ernst, *J. Chem. Phys.*, 71 (1979) 4546–4553.
- [29] Anil Kumar, R.R. Ernst, and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 95 (1980) 1–6.
- [30] M. Rance, G. Bodenhausen, G. Wagner, K. Wüthrich, and R.R. Ernst, *J. Magn. Reson.*, 62 (1985) 497–510.
- [31] G. Bodenhausen and D.J. Ruben, *Chem. Phys. Lett.*, 69 (1980) 185–189.
- [32] L. Lerner and A. Bax, *J. Magn. Reson.*, 69 (1986) 375–380.
- [33] T.J. Norwood, J. Boyd, J.E. Heritage, N. Soffe, and I.D. Campbell, *J. Magn. Reson.*, 87 (1990) 488–501.
- [34] A.J. Shaka, P.B. Baker, and R. Freeman, *J. Magn. Reson.*, 64 (1985) 547–552.
- [35] D. Marion and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 113 (1983) 967–974.
- [36] P. Güntert and K. Wüthrich, *J. Magn. Reson.*, 96 (1992) 403–407.
- [37] D.H. Live, D.G. Davis, W.C. Agosta, and D. Cowburn, *J. Am. Chem. Soc.*, 106 (1984) 1939–1941.
- [38] A. Bax and S. Subramanian, *J. Magn. Reson.*, 67 (1986) 565–569.
- [39] J.F.G. Vliegthart, L. Dorland, and H. van Halbeek, *Adv. Carbohydr. Chem. Biochem.*, 41 (1983) 209–374.
- [40] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, John Wiley, New York, 1986.
- [41] B. Blombäck, *Methods Enzymol.*, 11 (1967) 398–411.
- [42] J.S. Rohrer and R.R. Townsend, *Glycobiology*, 2 (1992) 457.
- [43] J.-M. Wieruszkeski, J.-C. Michalski, J. Montreuil, and G. Strecker, *Glycoconjugate J.*, 6 (1989) 183–194.