ELSEVIER

Carbohydrate Research 299 (1997) 33–48

CARBOHYDRATE RESEARCH

Synthesis and characterisation of highly glycosylated glycopeptides with T^n-antigenic structures corresponding to human glycophorin A^N

Gunther Klich [a], Hans Paulsen [a,*], Bernd Meyer [a], Morten Meldal [b], Klaus Bock [b]

[a] *Institut für Organische Chemie der Universität Hamburg, Martin-Luther-King Platz 6, D-20146 Hamburg, Germany*

[b] *Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Copenhagen, Denmark*

Received 5 September 1996; accepted 7 December 1996

Abstract

Two highly glycosylated O-glycopeptides corresponding to human glycophorin A^N with T^n-antigenic structures were synthesised. The first glycopeptide has two glycosylated clusters with three and six adjacent 2-acetamido-2-deoxy-D-galactose (GalNAc) glycosylation sites and represents the N-terminal octadecapeptide from Leu-1 to Lys-18. The second glycopeptide, a decapeptide from His-9 to Lys-18, contains as a compact cluster six adjacent GalNAc glycosylation sites. The solid phase synthesis was realised by using the carbohydrate-containing building blocks Fmoc-Ser(Ac_3GalN_3)-Pfp and Fmoc-Thr(Ac_3GalN_3)-Pfp. The synthesised substances were characterised by NMR spectroscopic techniques. The main techniques used were homonuclear TOCSY and NOESY as well as HMQC and HMBC for $^{13}C,^1H$ correlations.

Keywords: Glycopeptides; Solid-phase glycopeptide synthesis; NMR assignment

1. Introduction

The interest in glycoproteins has steadily grown because of their biological significance [1]. The availability of structurally well-defined natural glycopeptides has been limited, because of the many types and the heterogeneity of the oligosaccharides linked to the protein backbone. Techniques for the synthesis of glycopeptides are now available and different glycopeptides have been synthesised for biochemical and structural investigations [2]. We have designed an efficient strategy for the solid-phase peptide synthesis of O-glycopeptides containing GalNAc residues α-linked to L-serine or L-threonine [3]. These kinds of linkages are found in most of the O-glycoproteins isolated from natural sources [4]. The strategy uses the derivatives **1** of Ser and **2** of Thr, glycosylated with 3,4,6-tri-*O*-acetyl-2-azido-2-deoxy-α-D-galactopyranose, as building blocks (Fig. 1). Ser and

* Corresponding author.

0008-6215/97/$17.00
PII S0008-6215(96)00337-0

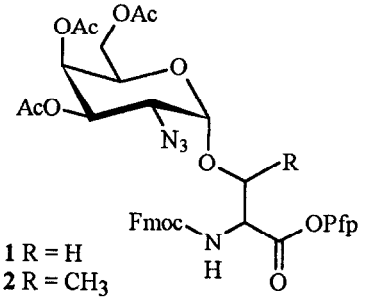


Fig. 1. The galactose containing building blocks **1** and **2** used for glycopeptide synthesis.

Thr in these building blocks are N^{α}-protected with the fluoren-9-ylmethyloxycarbonyl- (Fmoc-) group and C^{o}-protected with the pentafluorophenyl- (Pfp-) ester. Therefore, carbohydrate-containing amino acids can be introduced at any position in the peptide sequence by conventional peptide synthesis techniques. Glycopeptides synthesised in this way contain 2-azido functions in the D-galactose residues. The conversion of azido groups into the desired acetamido derivatives by reduction with thioacetic acid gives the corresponding glycopeptides which are then cleaved from the solid support [5].

The application of this procedure to the synthesis of highly glycosylated glycopeptides with a series of adjacent saccharide side-chains required examination of the progress of the peptide coupling reactions in the presence of bulky saccharide residues. Our target molecule was a sequence of glycophorin A^{N}, an important transmembrane sialoglycoprotein found in erythrocytes [6]. The extracellular part of the 131-amino-acid protein, consisting of 71 residues, is highly glycosylated with 16 oligosaccharides. In Fig. 2 the N-terminus from Leu-1 to Arg-39, called T1-fragment, is presented. This fragment carries 12 *O*- and one *N*-glycosylation sites. The glycosylated hydroxy amino acids carry tetrasaccharides such as α-Neu*p*5Ac-(2 → 3)-β-D-Gal*p*-(1 → 3)-[α-Neu*p*5Ac-(2 → 6)]-α-D-Gal*p*NAc. Asn is glycosylated with a complex type tridecasaccharide [7]. The sequences from Ser-2 to Thr-4 and from Thr-10 to Ser-15 each form a cluster of adjacent glycosylation sites. Former synthetic works were focussed on the terminal five residues of the glycoprotein carrying T- or T^{n}-antigenic structures [8].

Carbohydrate side-chains influence the conformation of a peptide backbone. Theoretical calculations on mucin-analog glycopeptides indicate an extended peptide chain in highly glycosylated glycopeptides as compared to peptides [9]. NMR and theoretical structural studies on the first five residues of the N-terminus carrying T- and T^{n}-antigenic structures have been performed by Pèpe et al. and have shown a relatively fixed conformation [10]. NMR studies of the T1-fragment demonstrated that peptide–carbohydrate interactions induce well-defined peptide backbone structures [11]. Shogren et al. suggest that in mucins the first sugar unit, connected directly to the peptide backbone, is dominant for the influence of the saccharide moiety on the peptide conformation [12]. According to this, the other components of the carbohydrate side-chain attached to GalNAc are relatively flexible and have less influence on the conformation of the peptide backbone. For this reason it is

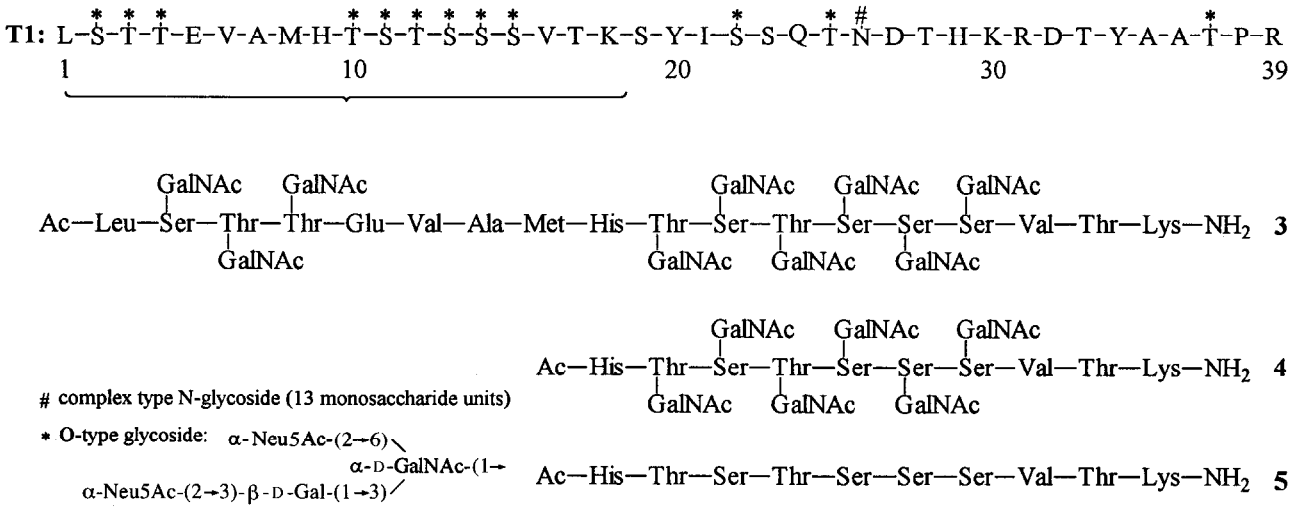


Fig. 2. The T1 glycoprotein fragment (top) and synthesised glycopeptides **3** and **4** and peptide **5**. Left bottom: oligosaccharide structure of *O*-glycosylation sites.

interesting to synthesise the glycopeptides **3** and **4** (Fig. 2), highly glycosylated with GalNAc only.

2. Results and discussion

The synthesis of compound **3** and the smaller fragment **4** was realised by using a continuous-flow peptide synthesiser containing monitoring equipment, i.e. a solid-phase spectrophotometer and a PC-unit [13]. The principle of solid-phase glycopeptide synthesis [4] is illustrated in Fig. 3. The polyethylene glycol poly-*N*,*N*-dimethyl acryl amide copolymer (PEGA) resin **7** [14] was derivatised with a peptide-amide-(Rink-)linker [15]. The amino acids as well as the glycosyl amino acids **1** and **2** were used as N^{α}-Fmoc-protected Pfp-esters. A reagent for activating the coupling reaction is not necessary. The introduction of the amino acids, e.g. N^{α}-Fmoc-Lys(Boc)-Pfp **6**, was realised in the presence of 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine (Dhbt-OH). The presence of Dhbt-OH is necessary in order to follow the completion of the reaction by a colour alteration. In the presence of amino groups, a yellow Dhbt-O^- salt can be detected by UV-absorption at 440 nm [16]. By termination of the coupling reaction the colourless Dhbt-OH is reformed. Monitoring the reaction with a UV detector gives exact information about the reaction state. Coupling of commercially available non-glycosylated amino acid derivatives was performed using 3 equiv of the N^{α}-Fmoc-amino acid-Pfp-esters, whereas for compounds **1** and **2** only an excess of 1.5 equiv was used. After completion of the synthesis the terminal amino acid, for compound **4** His-9, was treated with acetic anhydride (50% in DMF) for *N*-acetylation.

Comparison of the synthesis of the non-glycosylated peptide **5** and the glycosylated compound **4** revealed that the observed reaction times for the glycosylated peptide **4** were faster (1–2 h) than for the non-glycosylated compound **5** (3–6 h). Overall, the couplings were relatively fast reactions compared to those described in ref. [3]. Probably, secondary structure elements were formed during the synthesis of peptide **5** in the resin-bound state, and further coupling reactions were rendered more difficult. During the synthesis of glycopeptide **4** the bulky glycosyl residues are likely to extend the growing peptide, and the free terminal amino functions are in a more favourable position for the next coupling step. This experimental observation indicates a preference of an extended conformation of the peptide backbone in highly glycosylated glycopeptides [17]. However, the peptide synthesis of a glycopeptide with a series of adjacent carbohydrate side-chains seems to be done easily by using building blocks **1** and **2**.

During the formation of compound **4**, at first a polymer-bound glycopeptide arises, terminated at His-9. This compound contains six galactose-2-azido groups which have to be reduced to acetamido functions. Previous studies demonstrated the formation of 2-deoxy-2-thioacetylamino-D-galactose-containing byproducts as a result of the reduction using thioacetic acid [18]. To prevent the generation of such a complex mixture of byproducts by reduction of the six 2-azido-2-deoxy-D-galactose residues, a high purity of thioacetic acid was needed. GC analysis showed that thioacetic acid forms impurities during storage at temperatures over 0 °C. Hence, thioacetic acid was condensed in a nitrogen atmosphere from an ice-cooled container to a liquid-nitrogen-cooled receptacle, using a pressure of 5 kPa to realise a practical rate of condensation. Once this procedure was repeated six times, a purity of 99.8% was measured with no sulfurous odour present. The transformation of the azido group to the acetamido functionality was realised by treating the resin in nitrogen atmosphere with thioacetic acid (99.8%). With a repeated exchange of the reagent within every 7 h the reaction was finished after 48 h. The progress of the reaction was monitored by IR spectroscopy using the azide absorption.

The hexaglycosylated glycopeptide **4** was obtained by cleavage from the polymer with aq 95% trifluoroacetic acid, HPLC purification, and deacetylation with sodium methoxide in methanol at pH 8.5, monitored by NMR spectroscopy (data not shown). During the deacetylation with sodium methoxide the partially deacetylated glycopeptide precipitated and was resolubilised by the addition of water. In this step β-elimination destroyed part of the substance. Separation of compound **4** from resulting salts and β-elimination fragments was achieved by ultrafiltration using a 1000 D ultrafiltration-membrane in water. An overall yield of 26% was obtained. The deacetylation with sodium methoxide in methanol is thus not very effective for glycopeptides which precipitate in methanol solution. Hence, for the larger molecule **3** the deacetylation with hydrazine in methanol in the polymer-bound state was used [19].

The experience from the synthesis of compound **4** was used in the synthesis of the larger glycopeptide **3**. Coupling cycles following at position His-9 were continued and compound **9** was prepared (Fig. 3).

Reduction with highly purified thioacetic acid resulted in the *N*-acetylated product **10**. The *O*-deacetylation of the carbohydrate components took place in a favourable manner as a solid-phase reaction with hydrazine in methanol and gave the glycopeptide **11**. After cleaving compound **11** from the polymer with

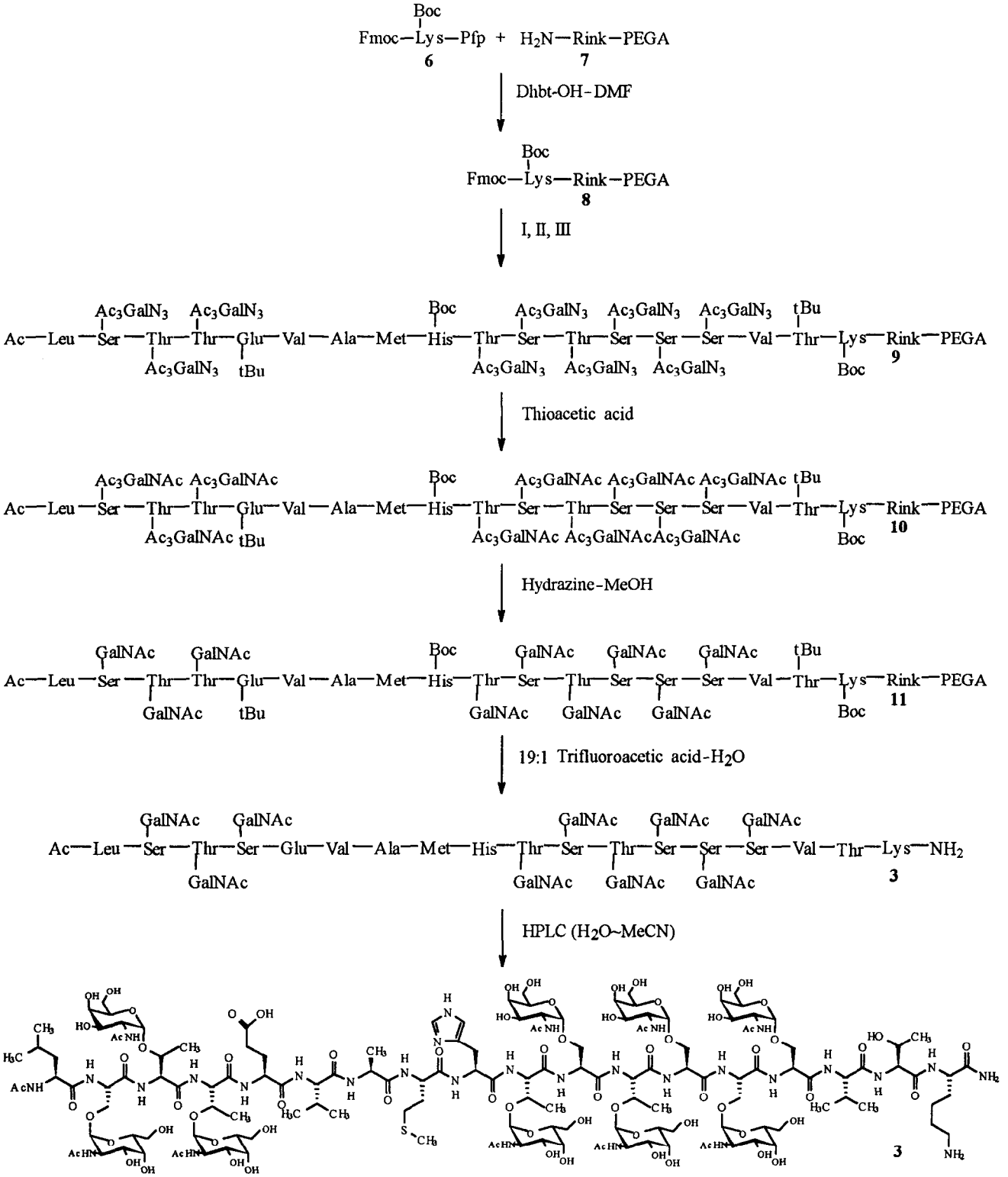
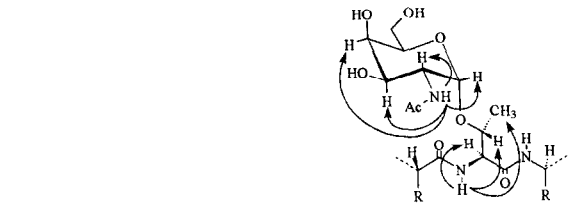


Fig. 3. Principle of solid-phase glycopeptide synthesis: I, 20% piperidine–DMF; II, coupling of N^{α}-Fmoc-amino acid-Pfp-ester, N^{α}-Fmoc-Ser(Ac_3GalN_3)-Pfp-ester, or N^{α}-Fmoc-Thr(Ac_3GalN_3)-Pfp-ester; III, repeat of I and II till the sequence is completed.

trifluoroacetic acid, HPLC analysis of the resulting compound **3** showed some minor byproducts. After repeated separation with HPLC, these byproducts could be identified by ^{1}H NMR spectroscopy and mass spectrometry as a mixture of 2-deoxy-2-thioacetylamino-D-galactose derivatives. The satisfactory overall yield for compound **3** was 51%.

For the determination of the structures of the compounds a detailed NMR analysis was required. The NMR interpretation is not straightforward because of similar chemical shifts and high overlap of GalNAc signals. All described compounds were pure according to analytical HPLC, ^{1}H NMR spectroscopy, and mass spectrometry, and the analytical data were in agreement with the structures. The assignment of the ^{1}H NMR signals of compound **4** is based on TOCSY, DQF-COSY, and NOESY experiments. All signals were assigned starting from NH traces in the TOCSY spectra recorded in H_2O (Fig. 4). NOESY spectra were used to confirm the connectivity between the amino acids. Most information was given by NOEs between Hα or Hβ of Ser and Thr and protons of the GalNAc residues. This was completed by several NOEs between protons of, e.g., Thr γ-Me and NH of GalNAc or between methyl-protons of GalNAc and NH of Ser. To improve the intensity



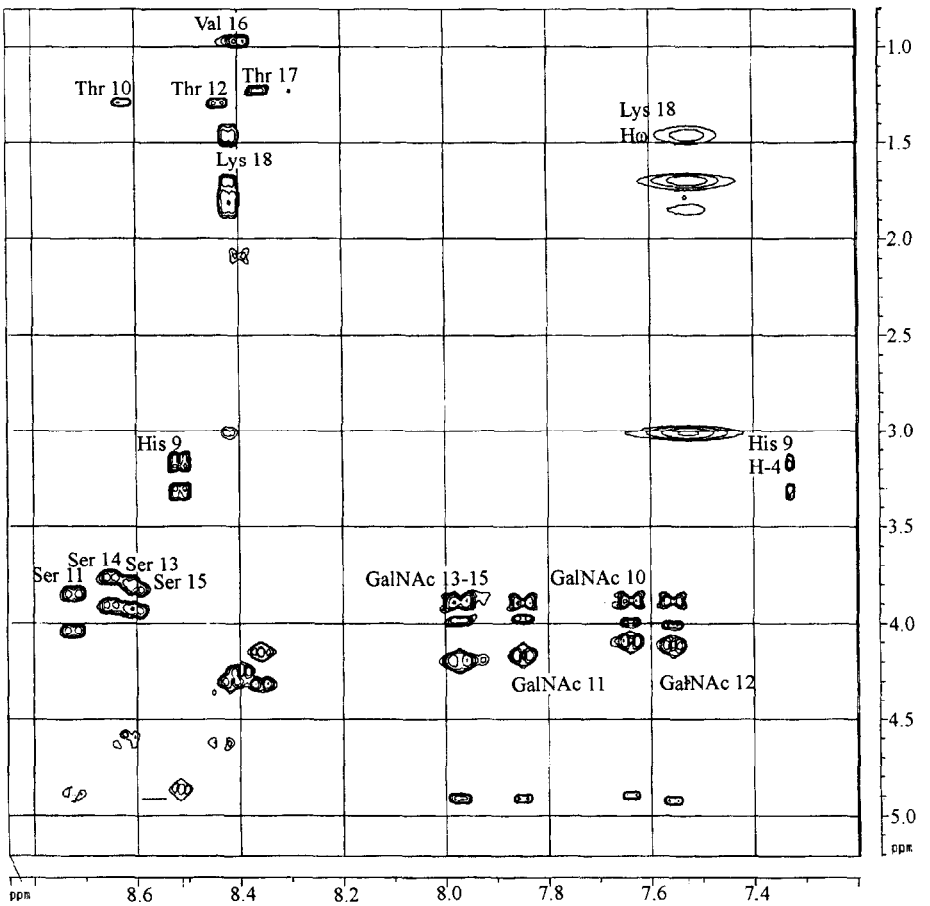


Fig. 4. Part of the TOCSY spectrum of compound **4** recorded in 9:1 H_2O–D_2O. The interactions between NH-protons and protons from carbohydrate residues or side-chains are shown. Some cross-peaks between NH and CH are indicated.

of the NH signals the measurement was performed at pH 3.5, adjusted with trifluoroacetic acid. To analyse the region obscured by the water signal, spectra in D_2O were also recorded. The signals of the peptide backbone amido groups ranged from 8.3 to 8.8 ppm. The NH signals of GalNAc were found in a region from 7.5 to 8.0 ppm and the amino acid side-chain amino protons resonated from 7.1 to 7.7 ppm. The assignment of the resonances descended from the amino acid and the carbohydrate protons could be achieved. These data are presented in Table 1. To confirm the data and to obtain coupling constants, DQF-COSY spectra were evaluated. Coupling constants of all amino-acid protons and those of H-1, H-2, and H-3 from GalNAc were detectable (Table 1). The coupling constants $J_{H\alpha,H\beta}$ of all glycosylated Thr residues have values of 2.4 Hz, which has to be interpreted as being an indication of an almost orthogonal orientation with little or no free rotation around the Cα–Cβ bond [20].

For the analysis of the compounds **3** and **5** the ^{13}C NMR data were used. Furthermore, ^1H NMR spectroscopic techniques were applied, with spectra measured at 500 MHz, and additionally HMQC and

Table 1
^1H NMR chemical shifts (ppm) and coupling constants (Hz, in parentheses) of compound **4** measured at 400 MHz in 99.96% D_2O (pH 6.0) at 300K. Amide protons were detected in 9:1 H_2O–D_2O (pH 3.5) at 300K. COSYDQFTPPI spectra to determine the coupling constants were measured with a sweep width of 12.02 ppm and 8 K acquired data points

	NH	Hα	Hβ$_1$	Hβ$_2$	Hγ$_1$	Hγ$_2$	Hδ	Hε	NH$_2$	NH	CH$_3$	H-1	H-2	H-3	H-4	H-5	H-6
Lys-18	8.44 (8.4) (5.8)	4.30 (8.2) (8.8)	1.86 (10.4)	1.79 (6.8)	1.46 (7.6)	1.46 (7.6)	1.69 (7.6)	3.00 (6.4)	7.53								
Thr-17	8.38 (7.6)	4.32 (6.0)	4.14 (6.8)		1.23												
Val-16	8.42 (8.2)	4.24 (7.8)	2.09 (7.1) (7.1)		0.97	0.94											
Ser-15	8.67 (6.0)	4.68 (6.0) (5.6)	3.92 (10.6)	3.78						7.99 (9.6)	2.06	4.91 (3.6)	4.20 (10.4)	3.87 (3.4)	3.98	3.97	3.75
Ser-14	8.63 (8.0)	4.57 (5.6) (5.2)	3.94 (10.6)	3.80						8.00 (9.6)	2.05	4.91 (3.6)	4.20 (10.4)	3.89 (3.4)	4.01	3.98	3.76
Ser-13	8.62 (8.0)	4.69 (6.0) (5.2)	3.96 (10.4)	3.83						7.99 (9.6)	2.08	4.91 (3.6)	4.20 (10.4)	3.89 (3.4)	3.99	3.98	3.75
Thr-12	8.64 (8.6)	4.64 (2.4)	4.36 (5.8)		1.28					7.57 (9.6)	2.07	4.90 (3.6)	4.12 (10.4)	3.89 (3.4)	3.99	3.99	3.77
Ser-11	8.74 (7.6)	4.90 (6.0) (5.4)	4.03 (10.8)	3.85						7.87 (9.2)	2.03	4.91 (3.6)	4.17 (10.4)	3.90 (3.4)	4.00	3.98	3.76
Thr-10	8.46 (9.2)	4.64 (2.4)	4.36 (5.8)		1.28					7.66 (9.6)	2.08	4.93 (3.6)	4.10 (10.4)	3.88 (3.4)	3.99	3.99	3.76
His-9	8.53 (8.2)	4.87 (5.8) (9.4)	3.31 (15.8)	3.15									8.63 (1.4)		7.33		
Ac/NH$_2$									7.64	7.11	2.01						

HMBC spectra were recorded. Spectra measured in H_2O were also recorded at pH 3.5. The NH signals of the peptide backbone for compound **5** were located between 8.0 and 8.5 ppm; they are shifted by 0.3 ppm to higher field compared to glycopeptide **4**. The amino-acid residues show chemical-shift dispersion for all amide protons as well as for the CHα protons. Differences were observed for the coupling constants between glycosylated and their corresponding non-glycosylated amino acids, particularly for Thr Hα/Hβ, Ser Hβ_1/Hβ_2 ($\Delta J \sim 2$ Hz) and Ser/Thr NH-Hα ($\Delta J \sim 1$ Hz). The various data are presented in Table 2. NOESY spectra were not very useful for the analysis of compound **5**. Only a few neighbouring effects and no interactions between protons of the side-chains were detected. However, HMQC and HMBC experiments were very helpful in the assignment of the carbon resonances. The carbon resonances were determined using HMQC spectra proceeded from proton resonances. In HMBC spectra, magnetic transfers between Hα and carbonyl-carbons of neighbouring amino acids via NH groups were detected in some cases (see Fig. 5). This elegant and simple method led to a complete assignment of the

Table 2
^{1}H NMR chemical shifts (ppm) and coupling constants (Hz, in parentheses) of compound **5** measured at 500 MHz in 99.96% D_2O (pH 5.5) at 300K. Amide protons were detected in 9:1 H_2O–D_2O (pH 3.5) at 300K. COSYDQFTPPI spectra to determine the coupling constants were measured with a sweep width of 9.92 ppm and 8 K acquired data points

	NH	Hα	Hβ_1	Hβ_2	Hγ_1	Hγ_2	Hδ	Hε	NH_2	CH_3	H-2	H-4
Lys-18	8.282 (7.2)	4.332 (9.6) (5.8)	1.880 (14.4) (8.0)	1.799 (4.8)	1.478 (8.0)	1.463 (8.0)	1.712 (8.0)	3.015 (6.0)	7.450			
Thr-17	8.170 (7.6)	4.361 (5.8)	4.200 (6.2)		1.233							
Val-16	8.072 (7.8)	4.230 (7.6)	2.143 (6.8) (6.8)		0.978	0.972						
Ser-15	8.475 (7.0)	4.602 (6.0) (6.0)	3.951 (12.0)	3.909								
Ser-14	8.293 (7.0)	4.547 (6.0) (5.6)	3.927 (12.0)	3.881								
Ser-13	8.343 (7.0)	4.545 (6.0) (5.6)	3.939 (12.0)	3.892								
Thr-12	8.242 (7.8)	4.453 (4.8)	4.326 (6.8)		1.246							
Ser-11	8.293 (7.0)	4.529 (6.0) (6.0)	3.958 (12.0)	3.915								
Thr-10	8.227 (7.6)	4.455 (4.8)	4.267 (6.8)		1.227							
His-9	8.442 (7.8)	4.793 (6.1) (8.5)	3.385 (16.0)	3.187							8.629 (1.4)	7.330
Ac/NH_2	7.066								7.528	2.032		

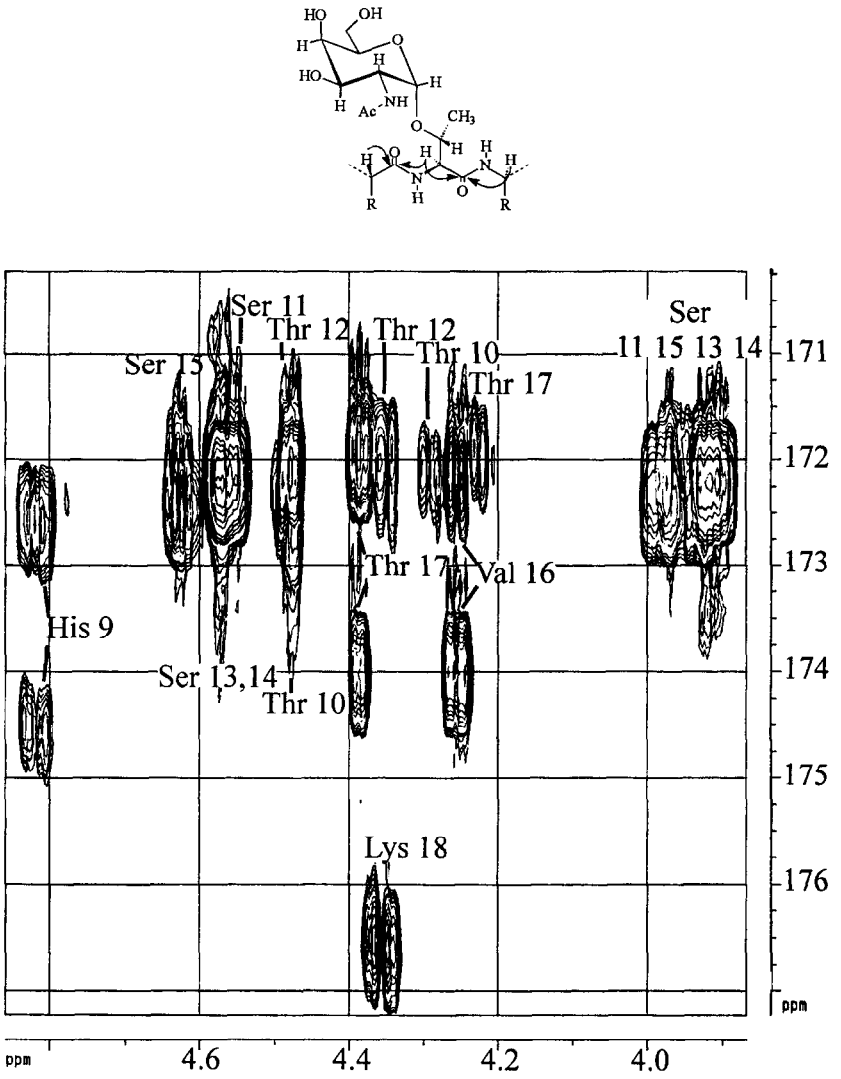


Fig. 5. Part of the HMBC spectrum of compound **5** recorded in D_2O. It shows the resonances from CHα-protons of the peptide backbone and from neighbouring carbonyl carbons. The cross-peaks between CO and Hα are indicated.

Table 3
^{13}C NMR chemical shifts (ppm) of compound **5** measured at 125 MHz in 99.96% D_2O (pH 5.5) at 300K

	CO	Cα	Cβ	Cγ_1	Cγ_2	Cδ	Cε	CH_3	C-2	C-4	C-5
Lys-18	177.19	54.25	31.30	22.99		27.17	40.16				
Thr-17	172.54	60.08	67.93	19.69							
Val-16	174.62	60.69	30.92	19.31	18.65						
Ser-15	173.05	56.59	61.98								
Ser-14	172.79	56.63	61.88								
Ser-13	172.94	56.59	62.01								
Thr-12	172.74	59.99	68.15	19.72							
Ser-11	172.84	56.46	61.98								
Thr-10	172.76	59.83	67.90	19.72							
His-9	173.20	53.55	27.32						134.53	118.15	129.47
Ac	175.15							22.61			

connectivities. The resulting ^{13}C NMR data set is presented in Table 3.

For compound **3** the assignment of the ^{1}H NMR signals was based on TOCSY, ECOSY, and NOESY experiments. The ^{13}C NMR data were analysed using HMQC and HMBC spectra. The chemical shifts of the NH signals of the peptide backbone lie within a range from 8.0 to 8.7 ppm. The GalNAc-NH signals are resonating between 7.4 and 8.0 ppm, and the side-chain amino-group protons are detected in the region from 7.0 to 7.7 ppm. TOCSY spectra led to the determination of chemical shifts of the peptide segments including side-chains. The chemical shifts for the NH signals of Thr-10 and Thr-12 of compound **3** are identical, but for Hβ the chemical shifts for Thr-10 and Thr-12 do not have the same value. The chemical shifts for GalNAc H-1 are scattered over 0.05 ppm in compound **3**, nearly twice as wide as those in compound **4**. The coupling constants for compound **3**, observed in 1D spectra, DQF-COSY and ECOSY, do not differ significantly from the values observed in compound **4**. NOESY, HMQC, and HMBC spectra all gave information about the connectivities. The detected NOE transfers from GalNAc H-1 to Ser/Thr Hα/Hβ, as shown in Fig. 6, showed the interactions in the carbohydrate and

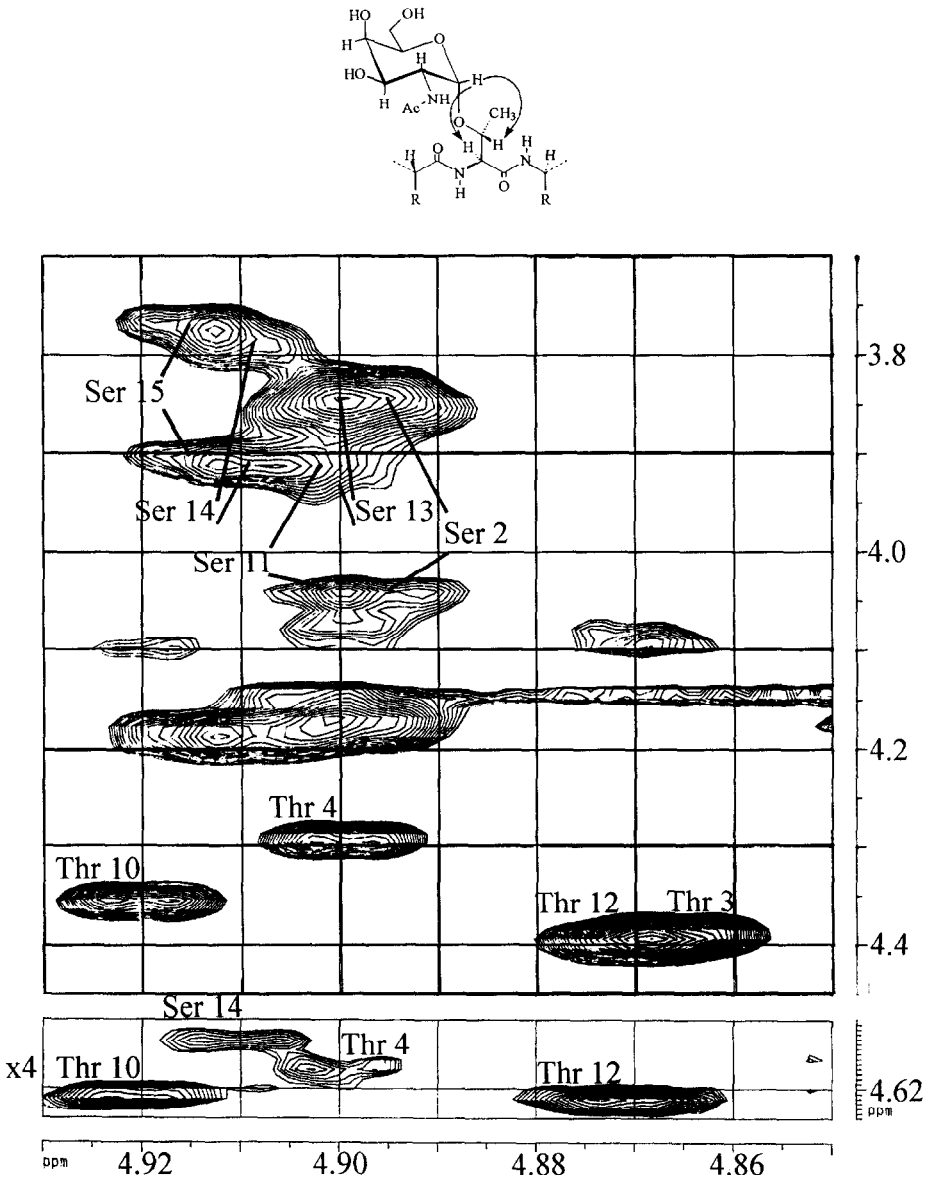


Fig. 6. Part of the NOESY spectrum of compound **3** recorded in D_2O. The interactions between amino acid and carbohydrate parts are shown. The cross-peaks between GalNAc H-1 and Ser/Thr Hα/Hβ are indicated.

Table 4
^{1}H NMR chemical shifts (ppm) and coupling constants (Hz, in parentheses) of compound **3** measured at 500 MHz in 99.96% D_2O (pH 6.0) at 300K. Amide protons were detected in 9:1 H_2O–D_2O (pH 3.5) at 300K. ECOSY spectra to determine the coupling constants were measured with a sweep width of 9.84 ppm and 8 K acquired data points

	NH	Hα	$H\beta_1$	$H\beta_2$	$H\gamma_1$	$H\gamma_2$	$H\delta_1$	$H\delta_2$	Hε	NH_2	NH	Ac	H-1	H-2	H-3	H-4	H-5	H-6a	H-6b
Lys-18	8.443 (7.0)	4.280 (11.3) (5.0)	1.810 (14.6) (7.2) (8.0)	1.760 (8.8) (6.4)	1.473 (13.0) (5.0)	1.431 (6.5)	1.693 (8.0)	1.693 (8.0)	2.993 (6.4)	7.488									
Thr-17	8.363 (8.0)	4.301 (6.3)	4.093 (6.8)		1.212														
Val-16	8.394 (9.8)	4.233 (8.0)	2.086 (7.2) (7.0)		0.968	0.947													
Ser-15	8.646 (7.2)	4.668 (7.2) (5.6)	3.908 (10.4)	3.757							7.892 (9.7)	2.049	4.915 (3.6)	4.178 (11.0)	3.864 (3.4)	3.886	3.967	3.748	3.703
Ser-14	8.609 (9.6)	4.567 (6.8) (5.6)	3.919 (10.4)	3.796							7.986 (9.5)	2.030	4.911 (3.6)	4.185 (11.0)	3.891 (3.2)	3.915	3.987	3.761	3.725
Ser-13	8.600 (9.6)	4.682 (6.0) (5.6)	3.940 (11.2)	3.830							7.964 (9.5)	2.080	4.902 (3.6)	4.200 (10.8)	3.878 (3.0)	3.898	3.975	3.748	3.703
Thr-12	8.390 (9.2)	4.618 (2.3)	4.383 (6.5)		1.280						7.486 (9.5)	2.056	4.874 (3.8)	4.109 (11.0)	3.887 (3.2)	3.910	3.994	3.795	3.743
Ser-11	8.696 (8.8)	4.881 (5.2) (5.6)	4.040 (10.4)	3.848							7.851 (8.8)	2.030	4.907 (3.8)	4.153 (11.2)	3.892 (2.8)	3.914	3.971	3.761	3.725
Thr-10	8.390 (9.2)	4.618 (2.3)	4.343 (6.7)		1.285						7.537 (9.4)	2.065	4.921 (4.0)	4.108 (11.0)	3.876 (3.2)	3.898	3.997	3.761	3.725
His-9	8.597 (6.8)	4.935 (5.6) (9.8)	3.338 (16.4)	3.208										8.652 (1.4)		7.324			

Met-8	8.397 (8.0)	4.410 (5.5) (8.5)	2.080 (14.0) (8.0) (8.0)	1.978 (8.0) (8.0)	2.487				2.053										
Ala-7	8.362 (7.0)	4.279 (7.6)	1.352																
Val-6	8.208 (8.4)	3.973 (8.0)	2.028 (7.2) (7.0)		0.993	0.941													
Glu-5	8.260 (7.0)	4.430 (7.2) (7.2)	1.993 (6.3) (7.8)	1.941 (6.3) (7.8)	2.590 (15.5)	2.508													
Thr-4	8.715 (7.6)	4.588 (2.3)	4.299 (6.4)		1.244						7.689 (9.5)	2.054	4.904 (3.6)	4.096 (11.2)	3.876 (3.0)	3.896	3.988	3.761	3.725
Thr-3	8.568 (7.0)	4.754 (2.3)	4.380 (6.4)		1.297						7.647 (9.5)	2.054	4.875 (3.6)	4.090 (11.2)	3.886 (3.0)	3.898	3.989	3.795	3.743
Ser-2	8.537 (7.6)	4.826 (5.2) (6.8)	4.045 (10.8)		3.825						(9.5)	7.964	2.076 (3.4)	4.898 (11.0)	4.186 (3.2)	3.892	3.915	3.978	3.7613.7
Leu-1	8.198 (6.8)	4.364 (5.9) (10.1)	1.620 (13.0) (6.0)	1.584 (8.0)	1.678 (7.0) (6.8)		0.941	0.898											
Ac/NH_2										7.612	7.086	2.085							

Table 5
^{13}C NMR chemical shifts (ppm) of compound **3** measured at 125 MHz in 99.96% D_2O (pH 5.5) at 300K

	CO	Cα	Cβ	Cγ_1	Cγ_2	Cδ_1	Cδ_2	Cε	CH_3	NCO	C-1	C-2	C-3	C-4	C-5	C-6
Ac	176.02								23.65							
Lys-18	177.63	54.92	31.89	23.55		27.77		40.65								
Thr-17	172.88	60.69	68.49	20.14												
Val-16	174.67	60.63	32.28	19.83	19.40											
Ser-15	172.34	54.46	68.42						23.69	175.81	99.31	51.32	69.29	69.26	72.83	62.78
Ser-14	172.10	54.48	68.28						24.00	175.68	99.58	51.32	69.29	69.94	72.79	62.71
Ser-13	172.38	54.25	69.20						23.68	175.98	99.37	51.00	69.22	69.92	72.83	62.78
Thr-12	172.47	58.87	78.27	19.94					23.76	175.39	100.28	51.49	69.68	69.93	72.82	62.47
Ser-11	172.27	54.56	69.69						23.10	175.63	99.44	51.36	69.34	69.77	72.79	62.71
Thr-10	172.15	58.61	77.47	19.99					23.74	175.29	100.61	51.13	69.46	69.97	72.79	62.71
His-9	173.22	53.49	27.72									135.00		118.68	130.02	
Met-8	174.78	53.68	28.59	31.44				21.49								
Ala-7	176.16	51.19	17.91													
Val-6	174.37	61.08	31.52	19.79	19.76											
Glu-5	174.22	54.08	14.66	30.78		178.25										
Thr-4	172.01	58.20	77.88	20.03					23.90	175.37	100.33	51.27	69.64	69.45	72.79	62.71
Thr-3	172.70	58.63	78.76	19.94					23.90	175.37	100.77	51.27	69.64	69.45	72.82	62.47
Ser-2	172.66	54.86	68.32						23.86	175.86	99.57	51.09	69.14	69.45	72.79	62.71
Leu-1	176.47	53.95	41.39	25.81		23.60	22.32									

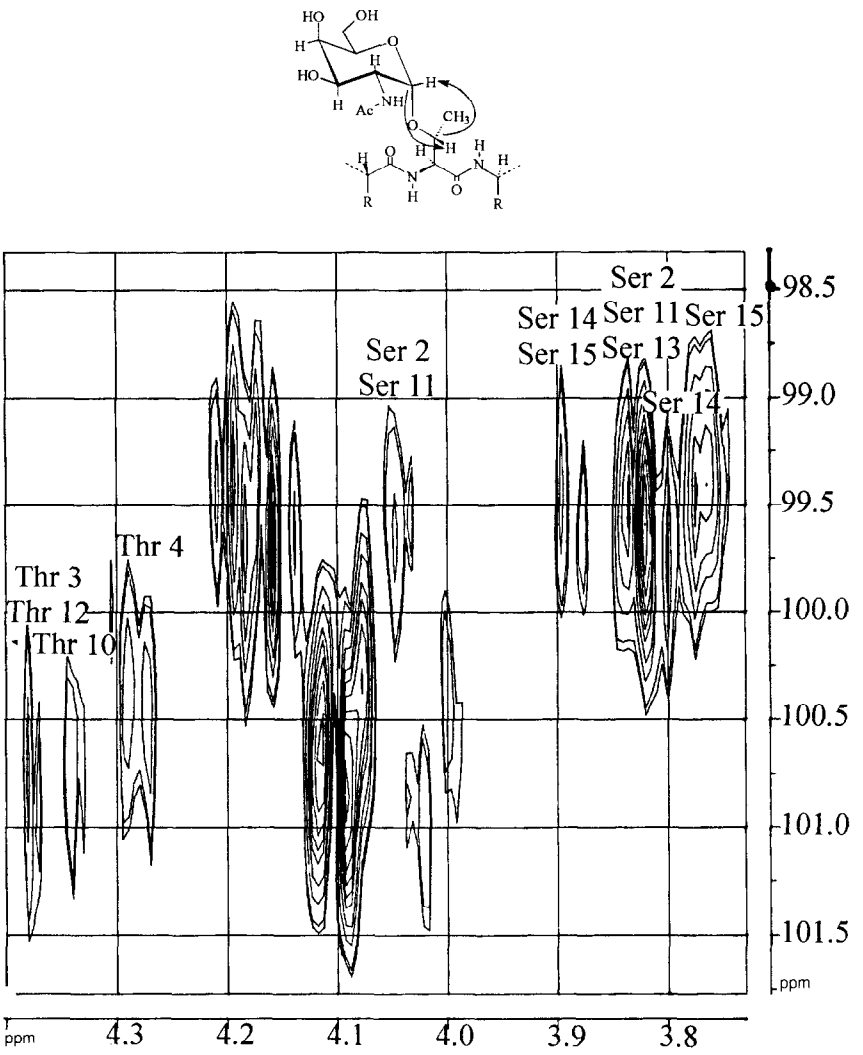


Fig. 7. Part of the HMBC spectrum of compound **3** recorded in D_2O. The through-bond coupling between carbohydrate and peptide residues are shown. The cross-peaks between GalNAc C-1 and Ser/Thr Hβ are indicated.

peptide parts of the molecule. An important piece of information extracted from the HBMC spectra is the through-bond magnetic transfer between GalNAc H-1 and Ser or Thr Cβ and, on the other hand, between GalNAc C-1 and Ser or Thr Hβ. This detected connectivity across the glycosidic linkage from HMBC (Fig. 7) led to complete assigned connectivities for the carbohydrate units. The peptide backbone was examined using NOESY spectra similar to compound **4**, and additional HMBC spectra as described for compound **5**. The resulting NMR data are presented in Tables 4 and 5. Electrospray mass spectrometry confirmed the purity and the structure of the glycopeptides **3** and **4** and peptide **5** (see Experimental).

In conclusion, the highly glycosylated glycopeptides **3** and **4** from glycophorin A^N have been synthesised by automated solid-phase synthesis using the building block approach with azido protection for galactose and the simultaneous conversion of up to 9 azido groups into acetamido functions on the solid phase. The purity and structural identity of the glycopeptides was confirmed and the compounds will be very useful for future enzyme assays and conformational analysis.

3. Experimental

General methods.—TLC was performed on Silica Gel F_{254} (E. Merck). MPLC was performed on Silica Gel Silitech 12-26, 6 nm (ICN) at 300–500 kPa with

distilled solvents. HPLC separations were performed on a Merck/Hitachi HPLC system using a Lichrospher RP-18 column (250 × 25 mm, 10 nm) with a flow rate of 10 mL/min. For analytical HPLC separations a Lichrospher RP-18 column (250 × 4 mm, 10 nm) with a flow rate of 2 mL/min was applied. Peak detection was performed with a photodiode array detector at 215 nm. Solvent system: A, 0.1% CF_3CO_2H in water; B, 0.1% CF_3CO_2H in acetonitrile. NMR spectra were recorded on a Bruker AMX 400 or DRX 500 spectrometer at 300K. Chemical shifts are given in ppm relative to internal acetone (δ 2.22) for solutions in D_2O. Coupling constants are given in Hz. Electrospray mass spectra were recorded on a VG-triple-quadruple electrospray mass spectrometer (positive mode, 50% acetonitrile). IR spectra were recorded on an ATI Mattson Genesis series FTIR 1001 spectrometer. All solvents used for synthesis were distilled at the appropriate pressure. DMF was distilled freshly before use and analysed for free amines by addition of Dhbt-OH prior to use. Suitably protected N^{α}-Fmoc amino acids and Pfp-esters [Dhbt-ester for Fmoc-Thr(tBu)] were purchased from Novabiochem or Bachem. The polyethylene glycol poly-*N*,*N*-dimethyl acryl amide copolymer (PEGA) resin [11] was derivatised with a peptide-amide-(Rink-)linker [12] as previously described [21].

Compounds **3**, **4**, and **5** were synthesised on a continuous-flow peptide synthesiser with real-time monitoring and in-line dissolution of amino acids. The PEGA-resin **7** (257 mg, substitution 175 μmol/g), which had been derivatised with the Rink-linker, was placed in a glass column through which the reagents were pumped or circulated with a flow rate of 1.44 mL/min. Fmoc deprotections were effected with 20% piperidine in DMF for 10 min. The suitably protected amino acids (3 equiv) and the glycosyl amino acid Pfp-esters **1** and **2** (1.5 equiv) were coupled with the equimolar addition of Dhbt-OH. The acylation reactions were monitored by a solid-phase spectrophotometer following the disappearance of the yellow ion pair of Dhbt-OH and unreacted amino groups [12]. After each Fmoc deprotection or coupling reaction, the resin was washed with DMF for 25 min. After completion of the peptide synthesis, the Fmoc group was removed and the terminal amino groups were acetylated with Ac_2O (50% in DMF) for 25 min. Washing with DMF, Et_2O, and CH_2Cl_2 afforded the resin-linked peptide or glycopeptide. Finally, the (glyco)peptide was cleaved from the polymer with concurrent removal of the Boc and *t*Bu side-chain protecting groups by treating the resin with aq 95% CF_3CO_2H (2 mL) for 2 h. The resin was washed with CF_3CO_2H (5 × 2 mL, 2 min each). The combined filtrates were evaporated and the residue was coevaporated several times with toluene and 3:1 toluene–MeOH.

For glycopeptides **3** and **4** the azide reduction was performed after transferring the solid support into a schlenk-funnel before the cleavage from the solid support. An aliquot (100 mg) of the polymer was treated with freshly distilled thioacetic acid (1 mL) (GC > 99.8%) under N_2. Thioacetic acid was renewed at the following times: 1, 2, 4, 7 h, then every 7 h. The reaction was followed by IR spectroscopy through the disappearance of the azide absorption band at ν_{KBr} 2113 cm^{-1}. After 47 h the azide band had disappeared completely. The resin was washed with CH_2Cl_2, MeOH, toluene, and Et_2O (5 × 2 mL of each solvent) and dried.

N^{α}-Acetyl-L-*leucyl*-O-*(2-acetamido-2-deoxy-α*-D-*galactopyranosyl)*-L-*seryl*-O-*(2-acetamido-2-deoxy-α*-D-*galactopyranosyl)*-L-*threonyl*-O-*(2-acetamido-2-deoxy-α*-D-*galactopyranosyl)*-L-*threonyl*-L-*glutamyl*-L-*valyl*-L-*alanyl*-L-*methionyl*-L-*histidyl*-O-*(2-acetamido-2-deoxy-α*-D-*galactopyranosyl)*-L-*threonyl*-O-*(2-acetamido-2-deoxy-α*-D-*galactopyranosyl)*-L-*seryl*-O-*(2-acetamido-2-deoxy-α*-D-*galactopyranosyl)*-L-*threonyl*-O-*(2-acetamido-2-deoxy-α*-D-*galactopyranosyl)*-L-*seryl*-O-*(2-acetamido-2-deoxy-α*-D-*galactopyranosyl)*-L-*seryl*-O-*(2-acetamido-2-deoxy-α*-D-*galactopyranosyl)*-L-*seryl*-L-*valyl*-L-*threonyl*-L-*lysylamide* (**3**). —The assembly of glycopeptide **3** was done according to the peptide sequence starting at the C-terminal end with the following acylation times: Lys(Boc), 3 h; Thr(*t*Bu), 3 h; Val, 5 h; **1**, 2 h; **1**, 2 h; **1**, 1.5 h; **2**, 1.5 h; **1**, 1 h; **2**, 1 h; His(Boc), 3 h; Met, 4 h; Ala, 4 h; Val, 6 h; Glu(*t*Bu), 5 h; **2**, 2 h; **2**, 1.5 h; **1**, 1 h; Leu, 4 h. Washing and drying as described above afforded 516.2 mg of resin. After reducing the azide groups the resin was treated with hydrazine (250 μL) in dry MeOH (3 mL) for 6 h. The resin was washed with MeOH and Et_2O (10 × 2 mL, 2 min each). The cleavage of the final product **3** was followed by purification (3 ×) using preparative HPLC [Buffer A–Buffer B, 100:0 → 90:10 (5 min) → 85:15 (50 min) → 75:25 (5 min)]. Yield: 16.6 mg (51%) (calculated on the substitution of the resin). ES^+MS Calcd for $C_{150}H_{254}N_{31}O_{76}S$: M, 3737.64. Found (> 10%): $[M + 2H + Na]^{3+}/3$, 1254.11 (100%); $[M + 2H + K]^{3+}/3$, 1259.82 (75%); $[M + H + K + Na]^{3+}/3$, 1273.79 (35%); $[M + 3H]^{3+}/3$, 1248.74 (12%). Amino acid analysis: Leu, 1.01; Ser, 5.07; Thr, 5.04; Glu, 0.98; Val, 1.97; Ala, 0.99; Met, 1.02;

His(+GalNAc), 4.72; Lys, 0.96. NMR data are presented in Tables 4 and 5.

N$^{\alpha}$-*Acetyl*-L-*histidyl*-O-(*2-acetamido-2-deoxy-α*-D-*galactopyranosyl*)-L-*threonyl*-O-(*2-acetamido-2-deoxy-α*-D-*galactopyranosyl*)-L-*seryl*-O-(*2-acetamido-2-deoxy-α*-D-*galactopyranosyl*)-L-*threonyl*-O-(*2-acetamido-2-deoxy-α*-D-*galactopyranosyl*)-L-*seryl*-O-(*2-acetamido-2-deoxy-α*-D-*galactopyranosyl*)-L-*seryl*-O-(*2-acetamido-2-deoxy-α*-D-*galactopyranosyl*)-L-*seryl*-L-*valyl*-L-*threonyl*-L-*lysylamide* (**4**).—The assembly of glycopeptide **4** was done according to the peptide sequence starting at the C-terminal end with the following acylation times: Lys(Boc), 3 h; Thr(*t*Bu), 3 h; Val, 5 h; **1**, 2 h; **1**, 2 h; **1**, 1.5 h; **2**, 1.5 h; **1**, 1 h; **2**, 1 h; His(Boc), 3 h. Washing and drying as described above afforded 394.2 mg of resin. 80 mg of the resin were used to reduce the azide groups, after which the glycopeptide was cleaved from the polymer as described above. The cleavage of the protected O-glycopeptide was followed by purification using preparative HPLC [Buffer A–Buffer B, 100:0 → 80:20 (10 min) → 70:30 (20 min) → 50:50 (10 min)]. Yield: 11.1 mg (41%) (calculated on the substitution of the resin). The protected glycopeptide (11.1 mg, 3.64 μmol) was dissolved in MeOH (1 mL) and treated with a 1% soln of NaOMe in MeOH (50 μL). After 2 h the precipitated partial deacetylated glycopeptide was solubilised by the addition of 1 mL water. The deacetylation was followed by NMR ($\delta_{AcOR} \approx 2.0$, $\delta_{AcONa} \approx 1.8$). After 4 h the reaction was stopped by the addition of HOAc (10 μL). The final product **4** was purified by ultrafiltration (1000 D, water). Yield: 5.3 mg (63% in the last reaction, the overall yield for the glycopeptide assembly was 26%). ES$^+$MS Calcd for $C_{91}H_{152}N_{20}O_{48}S$: M, 2294.40. Found (> 10%): $[M + 2H]^{2+}/2$, 1254.11 (100%); $[M + H + 2Na]^{3+}/3$, 778.82 (38%). Amino acid analysis: Ser, 4.03; Thr, 3.06; Val, 0.96; His(+GalNAc), 3.48; Lys, 0.95. NMR data are presented in Table 1.

N$^{\alpha}$-*Acetyl*-L-*histidyl*-L-*threonyl*-L-*seryl*-L-*threonyl*-L-*seryl*-L-*seryl*-L-*seryl*-L-*valyl*-L-*threonyl*-L-*lysylamide* (**5**).—The assembly of peptide **5** was done according to the peptide sequence starting at the C-terminal end with the following acylation times: Lys(Boc), 3 h; Thr(*t*Bu), 3 h; Val, 5 h; Ser(*t*Bu), 4 h; Ser(*t*Bu), 4 h; Ser(*t*Bu), 4 h; Thr(*t*Bu), 4.5 h; Ser(*t*Bu), 5 h; Thr(*t*Bu), 5 h; His(Boc), 6 h. Washing and drying as described above afforded 332.3 mg of resin. After cleavage from the polymer the final product **5** was purified by preparative HPLC [Buffer A–Buffer B, 100:0 → 70:30 (10 min) → 60:40 (20 min) → 50:50 (5 min)]. Yield: 19.9 mg (41%) (calculated on the substitution of the resin). ES$^+$MS Calcd for $C_{43}H_{72}N_{14}O_{18}S$: M, 1075.15. Found (> 10%): $[M + 2H]^{2+}/2$, 538.58 (100%); $[M + H + Na]^{2+}/2$, 549.71 (67%); $[M + H + K]^{2+}/2$, 558.12 (28%). NMR data are presented in Tables 2 and 3.

Acknowledgements

This work was supported by a grant of the science programme of the E.C. SC1*CT92-0765. We are especially grateful to Dr. V. Sinnwell and Dr. J. Pieper for the recording of NMR spectra.

References

[1] A. Varki, *Glycobiology*, 3 (1993) 97–130.
[2] M. Meldal, in Y.C. Lee and R.T. Lee (Eds.), *Neoglycoconjugates*, Academic Press, San Diego, 1994, pp 145–198.
[3] T. Bielfeldt, S. Peters, M. Meldal, K. Bock, and H. Paulsen, *Angew. Chem. Int. Ed. Engl.*, 31 (1992) 857–859.
[4] H. Paulsen, S. Peters, and T. Bielfeldt, in J. Montreuil, J.F.G. Vliegenthart, and H. Schachter (Eds.), *Glycoproteins*, Elsevier, Amsterdam, 1995, pp 87–122.
[5] H. Paulsen, T. Bielfeldt, S. Peters, M. Meldal, and K. Bock, *Liebigs Ann. Chem.*, (1994) 369–379 and 381–387.
[6] G.A. Jamieson and T.J. Greenwalt (Eds.), *Glycophorines of Blood Cells and Plasma*, J.B. Lipincott Company, Philadelphia, Toronto, 1971.
[7] M. Fukuda, M. Lauffenburger, H. Sasaki, M.E. Rogers, and A. Dell, *J. Biol. Chem.*, 262 (1987) 11952–11957.
[8] V. Verez Bencomo, J.-C. Jacquinet, and P. Sinaÿ, *Carbohydr. Res.*, 110 (1982) C9–C11; V. Verez Bencomo and P. Sinaÿ, *Carbohydr. Res.*, 111 (1983) C4–C7; B. Ferrari and A.A. Pavia, *Tetrahedron*, 41 (1985) 1939–1944.
[9] K.J. Butenhof and T.A. Gerken, *Biochemistry*, 32 (1993) 2650–2663.
[10] G. Pèpe, D. Siri, Y. Oddon, A.A. Pavia, and J.-P. Reboul, *Carbohydr. Res.*, 209 (1991) 67–81.
[11] J. Pieper, K.H. Ott, and B. Meyer, *Nat. Struct. Biol.*, 3 (1996) 228–232.
[12] R. Shogren, T.A. Gerken, and N. Jenthoft, *Biochemistry*, 28 (1989) 5525–5536.
[13] A. Dryland and R.C. Sheppard, *Tetrahedron Lett.*, 44 (1988) 859–876; L.R. Cameron, J.L. Holder, M. Meldal, and R. Sheppard, *J. Chem. Soc., Perkin Trans. 1*, (1988) 2895–2901.
[14] M. Meldal, *Tetrahedron Lett.*, 33 (1992) 3077–3080.
[15] H. Rink, *Tetrahedron Lett.*, 28 (1987) 3787–3788.
[16] E. Atherton, J.L. Holder, M. Meldal, R.C. Sheppard,

and R.M. Valerio, *J. Chem. Soc., Perkin Trans. 1*, (1988) 2887–2894.
[17] K. Bock, R. Stuike-Prill, and M. Meldal, in K. Bock and H. Clausen (Eds.), *Complex carbohydrates in drug research. Structural and functional aspects*, Munksgaard, Copenhagen, 1994, pp 229–242.
[18] T. Rosen, I.M. Lico, and D.T.W. Chu, *J. Org. Chem.*, 53 (1988) 1580–1583.
[19] B. Ferrari and A.A. Pavia, *Bioorg. Chem.*, 11 (1982) 85–95.
[20] A.A. Pavia and B. Ferrari, *Int. J. Peptide Protein Res.*, 22 (1983) 539–548.
[21] S. Peters, T. Bielfeldt, M. Meldal, K. Bock, and H. Paulsen, *Tetrahedron Lett.*, 32 (1991) 5067–5070.