Carbohydrate Research, 134 (1984) 141–155
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

STRUCTURAL ANALYSIS OF THE CARBOHYDRATE MOIETIES OF HUMAN TAMM–HORSFALL GLYCOPROTEIN

JANICE WILLIAMS, R. DEREK MARSHALL,
Division of Biochemistry, University of Strathclyde, Glasgow G4 ONR, Strathclyde (Great Britain)

HERMAN VAN HALBEEK, AND JOHANNES F. G. VLIEGENTHART
Department of Bio-Organic Chemistry, University of Utrecht, Croesestraat 79, NL-3522 AD Utrecht (The Netherlands)

(Received January 23rd, 1984; accepted for publication, May 14th, 1984)

ABSTRACT

Glycopeptides present in a pronase digest of human Tamm–Horsfall glycoprotein were fractionated by chromatography on DEAE-Sephadex A25 in 0.1M acetic acid. The separated glycopeptides were characterised by 500-MHz ^{1}H-n.m.r. spectroscopy, in conjunction with sugar and amino acid analysis, and they were shown to be of the *N*-glycosylic, *N*-acetyllactosamine type. Each fraction consisted mainly of a tetra-antennary entity having various degrees of complexity, with lesser amounts of the triantennary structure, and even smaller amounts of the diantennary type. There was extensive heterogeneity in non-reducing terminal groups in each of the glycopeptides, whereas the peptide portions were similar. The extent to which any one of the galactose residues in the *N*-acetyllactosamine units was substituted, and the type of substitution, differed. The substituents were α-NeuAc-(2→6), α-NeuAc-(2→3), and α-NeuAc-(2→3)[β-GalNAc-(1→4)]. The carbohydrate moieties of the glycoprotein were heterogeneous also because of an uneven distribution of the fucose residues, which were attached to GlcNAc residues, both that linked to asparagine and one or more of those present in the *N*-acetyllactosamine units. The α-NeuAc-(2→3)[β-GalNAc-(1→4)]-β-Gal-(1→ sequence forms, at least in part, the Sd^a immunodeterminant. The pK_a of the carboxyl group of the sialic acid residue in this entity is lower than that for molecules lacking GalNAc in this position. Thus, the difference in the number of Sd^a determinants carried by the glycopeptides enabled the latter to be fractionated on DEAE-Sephadex.

INTRODUCTION

Tamm–Horsfall glycoprotein[1,2] is the most abundant protein in normal human urine, and is produced within the kidney by the luminal cells of the thick ascending limb of the loop of Henle and by those morphologically similar cells of

0008-6215/84/$ 03.00 © 1984 Elsevier Science Publishers B.V.

the early distal tubule[3-5]. The glycoprotein, which is associated with the plasma membranes of the cells, is suggested to play a role[3,4] in the mechanism[6] by which urine dilution is effected in the renal tubule and the hypertonicity of the renal medulla is created and maintained. Therefore, it is probably important for the regulation of water balance in the living animal. In spite of the apparent importance of the glycoprotein for renal physiology, as well as in pathology involving the kidney, there is scant information about its molecular structure.

The carbohydrate-enriched fraction obtained from the glycoprotein which had been treated with boiling, aqueous barium hydroxide for several hours[7] was thought[8] to have a molecular weight of 12,300 at a time when methods for investigating glycoproteins were poorly developed. The carbohydrate moieties of the glycoprotein are now known to be considerably smaller. With the use of more modern methods for the isolation of glycopeptides from proteolytic digests of both Tamm–Horsfall glycoprotein and of its asialo-derivative, it has been shown that the protein chain carries at least five carbohydrate moieties with an average molecular weight of 4,300, two of them containing more sialic acid but less galactose than the other three, and that GalNAc is present as part of the structure of at least some of the asparagine-linked carbohydrate units[9].

High-resolution ^{1}H-n.m.r. spectroscopy is a powerful method for the determination of primary structures of carbohydrate chains of glycoproteins[10-13], and we now report on the application of this method to human Tamm–Horsfall glycoprotein.

EXPERIMENTAL

Tamm–Horsfall glycoprotein was isolated from normal, male, human urine by a modification[14] of the method used by Tamm and Horsfall[1,2] and subjected to electrophoresis in polyacrylamide gels, in the presence of sodium dodecyl sulphate, by the modified[15] procedure of Laemmli[16]. Separating and stacking gels contained 15% and 3% of acrylamide, respectively. Solutions (20 μL, 500 μg/mL) of pretreated Tamm–Horsfall glycoprotein were applied to the gels, and Coomassie Blue was used for visualisation[17,18].

Proteolytic digestion of Tamm–Horsfall glycoprotein. — The glycoprotein (20 mg/mL) was digested with pronase (glycoprotein–pronase ratio 30:1) for up to 24 h in 50mM sodium borate buffer (pH 8.3) containing 15mM $CaCl_2$. The digestion was monitored by using a ninhydrin microassay[19], with glycine as standard. The reaction was stopped by addition, to pH 4.0, of M acetic acid.

Isolation of glycopeptides. — The precipitate present in the acidified proteolytic digest was removed by centrifugation (2000*g*, 10 min, 4°), and the supernatant solution was lyophilised. The residue was dissolved in 50mM acetic acid (3 mL) and applied to a column (110 × 2.2 cm) of Sephadex G25 Fine equilibrated with 50mM acetic acid at 4°. The hexose-containing fractions were combined and lyophilised, the residue was dissolved in 0.1M acetic acid (3 mL), and applied to a column

(20 × 2.2 cm) of DEAE-Sephadex A25 equilibrated with 0.1M acetic acid at 4°. Elution was by stepwise increases of NaCl concentration in 0.1M acetic acid. The hexose-containing fractions (GP1-6) were separately lyophilised. A solution of each residue in water was extensively dialysed in an acetylated membrane against water to remove NaCl, and each solution was lyophilised. Visking dialysis membranes (16/32") were acetylated[20] by immersion in 27% acetic anhydride in pyridine for 24 h at 60°. After extensive washing in water, the membranes were stored in aqueous 40% methanol. Trypan Blue was used to test the porosity of the membranes.

Samples of the glycopeptides were dissolved in water (2 mL) for conversion into the *N*-benzyloxycarbonyl derivatives, which were subsequently digested with carboxypeptidase[21,22]. To each sample was added $NaHCO_3$ (400 mg) followed by benzyloxycarbonyl chloride (0.44 mL). Each mixture was shaken for 24 h at 37°, and neutralised first with HCl and then with an equal volume of M Tris–HCl to pH 7.5. Carboxypeptidase A (Type I, 0.85 mg) was added and each preparation was incubated at 37° for 24 h. Carboxypeptidase was removed by isoelectric precipitation at pH 4.6 by the addition of 5M acetic acid, and each supernatant solution was freeze-dried after removal of the precipitate. A solution of each dried product in water was dialysed extensively in acetylated membranes (see above) in order to reduce the loss of the benzyloxycarbonyl-glycopeptides (GPZ1-6).

Carbohydrate analyses. — Fractions from the chromatography columns were analysed for their sugar content by a colorimetric assay based on the method of François *et al.*[23]. To aliquots (0.2 mL) of solutions of glycoprotein, or glycopeptide, or of known concentrations of hexose (mannose–galactose 1:1) containing 3–6 μg of sugar, was added cold (4°) orcinol–sulphuric acid reagent (1.7 mL). The samples were kept at 80° for 14 min and then cooled in ice–water, and the absorbance was read at 505 nm against a water blank. A standard curve was obtained by using mannose–galactose mixtures.

Quantitative analyses of Tamm–Horsfall glycoprotein, glycopeptides, and the benzyloxycarbonyl-glycopeptides were carried out after treatment with methanolic 1.5M hydrogen chloride (24 h, 90°) by g.l.c. of the trimethylsilylated methyl glycosides according to a modification of the procedure of Bhatti *et al.*[24]. The silylating reagent (Tri-Sil; 50 μL) was added and, after 3 min at room temperature, a portion (5 μL) of the supernatant solution was injected into the column, packed with 3% of SE-30 on Diatomite, in a Pye Unicam gas chromatograph. The column temperature was programmed 120°→200° at 1°/min. Mannitol was used as the internal standard.

Amino acid analyses. — Samples (~10 nmol) of the glycopeptides were hydrolysed in 6M HCl (0.2 mL) under nitrogen in sealed ampoules for 24 h at 105°. The acid was then evaporated, the residue was taken up in 0.1M HCl (0.2 mL), and the solution was processed for amino acids and hexosamines using a Locarte amino acid analyser and the following programme: pH 3.25 (0.2M sodium citrate), 25 min; pH 4.25 (0.2M sodium citrate), 55 min; pH 9.35 (M sodium citrate), 75 min.

^{1}H-N.m.r. spectroscopy (500 MHz). — The benzyloxycarbonyl-glycopeptides

(GPZ2–4 and 6) were each repeatedly treated with D_2O at pD 7 with intermediate lyophilisation, and then dissolved in D_2O (0.4 mL, 99.96 atom% D) in a 5-mm glass tube.

^{1}H-N.m.r. spectroscopy was performed[10,11] on a Bruker WM-500 spectrometer (SON hf-NMR facility, Department of Biophysical Chemistry, Nijmegen University), operating in the Fourier-transform mode, and equipped with a Bruker Aspect 2000 computer. Resolution enhancement of the spectra was achieved by Lorentzian to Gaussian transformation. Spectra were recorded at probe temperatures of 27° and 12°, each of which was kept constant within 0.1°. Chemical shifts (δ) are expressed in p.p.m. downfield from the signal for internal sodium 4,4-dimethyl-4-silapentane-1-sulphonate (DSS) at 27° and were measured by reference to internal acetone (δ 2.225 in D_2O at 27°), with an accuracy of 0.002 p.p.m.

RESULTS

The glycoprotein gave a single band in electrophoresis on polyacrylamide gels in the presence of sodium dodecyl sulphate (see also ref. 25).

Preparation of glycopeptides. — Digestion of the Tamm–Horsfall glycoprotein with pronase in borate buffer rapidly and extensively cleaved the protein chain, ~70% of the peptide bonds being split within 6 h and >83% after 24 h (Fig. 1). Gel filtration on Sephadex G25 of the 6-h digest yielded six dominant peaks (I–VI; Fig. 1), only the last two of which absorbed significantly at 280 nm.

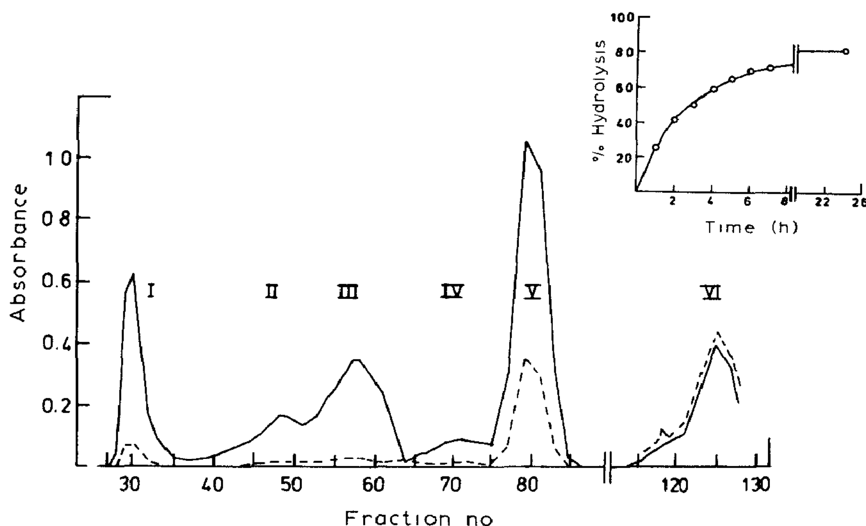

Fig. 1. Gel filtration of a 6-h pronase digest of Tamm–Horsfall glycoprotein (26 mg) on a column (110 × 2.2 cm) of Sephadex G25 (fine) in 0.1M acetic acid (fraction volume, 5.4 mL): glycopeptides (I), other u.v.-absorbing materials (II–VI); absorbances at 280 (----) and 230 nm (——) are shown. The *inset* illustrates the rate of proteolysis, catalysed by pronase, of Tamm–Horsfall glycoprotein (see Experimental).

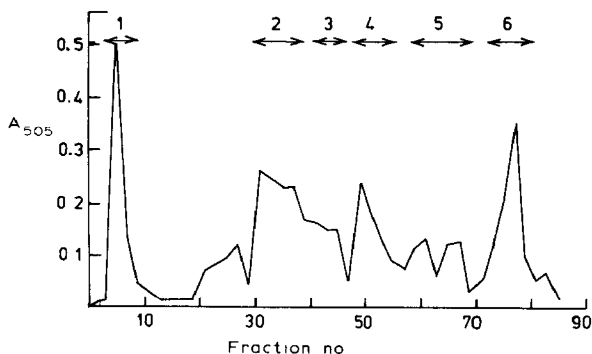

Fig. 2. Ion-exchange chromatography of the partially purified glycopeptides (*cf.* I in Fig. 1) from Tamm–Horsfall glycoprotein (equivalent to 40 mg of neutral sugars) on DEAE-Sephadex A25 in 0.1M acetic acid (3-mL fractions). Elution with NaCl in 0.1M acetic acid was effected at fractions 20 (10mM), 54 (25mM), and 66 (45mM); Fractions 1–6 correspond to glycopeptides GP1–6 (see Experimental).

Peak V had an absorbance spectrum typical of free tyrosine, and Peak VI had one characteristic of tryptophan in the 260–280 nm region, with appropriate absorbance at 230 nm. Thus, all of the aromatic amino acid residues had been released as free amino acids, in amounts which were calculated to be 17.4 mol of tyrosine and 7.8 mol of tryptophan per 80,000 g of the glycoprotein.

Glycopeptides were isolated from a 24-h digest of an amount of the glycoprotein equivalent to 40 mg of neutral sugar. The precipitate derived from the initial digest was devoid of carbohydrate. The 24-h digest gave a pattern of fractionation on Sephadex G25 similar to that shown by the 6-h digest (Fig. 1), with all of the hexose-containing material in the corresponding peak I. Chromatography of fraction I on DEAE-Sephadex A25 gave six fractions (GP1–GP6, Fig. 2).

Carbohydrate and amino acid composition. — The recoveries of neutral sugars in fractions GP1–6, after dialysis and based on colorimetric analysis, were 16.3, 19.0, 7.5, 9.0, 6.8, and 12.0%, respectively (Table I). This corresponds to a total hexose recovery of 71%. The amino acid composition was closely similar for all of the glycopeptides, apart from GP6 which had larger proportions of proline and glycine (Table I). The treatment of the benzyloxycarbonyl-glycopeptides with carboxypeptidase resulted in relatively small changes in the amino acid composition apart from the loss of proline and glycine from GP6.

The molecular weight and compositions of the benzyloxycarbonyl-glycopeptides GPZ1, 3, and 5 were similar. However, glycopeptides GPZ2, 4, and 6 appear to be different from the former and from each other in composition. Therefore, GPZ2, 4, and 6, together with GPZ3 as a representative of the former series, were used for further investigation.

^{1}H-N.m.r. spectroscopy. — The characteristic features of the 500-MHz ^{1}H-n.m.r. spectra of GPZ2–4 and 6, as solutions in D_2O, are summarised in Table II which indicated that each was an *N*-glycosylic glycopeptide of the *N*-acetyllactosamine type. The *N*-type can be concluded from the presence of signals for the

TABLE I

MOLAR COMPOSITION[a] AND RECOVERIES[b] OF GLYCOPEPTIDES, FROM TAMM-HORSFALL GLYCOPROTEIN, AFTER CHROMATOGRAPHY ON DEAE-Sephadex A25; THE SUGAR COMPOSITION OF THE GLYCOPROTEIN (THG) IS ALSO GIVEN

	1		*2*[c]			*3*			*4*			*5*		*6*			*Sum total*	*THG*[d]
	(i)	*(ii)*	*(i)*	*(ii)*	*(iii)*	*(i)*	*(ii)*	*(iii)*	*(i)*	*(ii)*	*(iii)*	*(i)*	*(ii)*	*(i)*	*(ii)*	*(iii)*		
Fuc	1.4	1.6	0 8	0 8	1.1	1.2	1.3	0 9	1.6	1.8	0 9	1.1	1.1	1.8	1.7	1.1	6.0	4.8
Man	4.0	4.0	2.8	2.8	3.0	4.5	4.5	3.0	8.8	8 8	3.0	4.5	4.5	5.3	5.3	3.0	27.3	24.0
Gal	4.6	4.6	3.2	3.3	3.9	5 1	5.5	3 8	7.5	7.7	3.8	5.0	4.9	6.2	6.1	3.9	23.4	23.6
GlcNAc	4.5	4.8	5.5	6.1	5.9	5.1	5.8	5.8	7.2	6.8	5.8	4.9	4.6	6.2	5.5	5.9	27.7	24.8
GalNAc	1.3	1.4	0.8	0 7	0.8	1.2	1 3	1.1	0 7	0.8	0.4	1.4	1 3	2.1	2.0	1.5	6.0	5.4
NeuAc	3.0	3.1	1.4	1.6	1 8	3.0	2 8	2.1	1 0	0.7	0.4	2.7	2 7	2.8	2.6	3.1	11 3	11.9
Lys	0.1	0	0	0		0	0		0.1	0		0 4	0.1	0.2	0			
His	0.2	0	0.3	0		0.2	0		0.1	0.1		0.3	0	0.3	0			
Asp	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1 0	1.0	1.0	1.0	1.0	1.0		
Thr	0.7	0.4	0 4	0.3		0.5	0.3		0.7	0.4		0.5	0.4	0.6	0.4			
Ser	0.9	0.9	0.8	0.7		0.7	0.6		0.8	0.6		1.0	0.7	1.0	0.6			
Glu	0.3	0.4	0.3	0.2		0.4	0.3		0.3	0.3		0.9	0.4	0.6	0.2			
Pro	0 6	0.4	0.7	0.5		0.6	0.4		0.9	0.8		0 6	0.5	1.6	0.6			
Gly	0.6	0.5	0.4	0.3		0.2	0.4		0.3	0.6		0.3	0.4	1.6	0.2			
Ala	0.3	0 3	0.3	0.2		0.2	0.2		0 4	0.3		0	0	0.5	0.3			
Val	0.2	0	0.1	0.1		0	0		0 1	0.1		0	0	0.2	0			
Leu	0.1	0	0 1	0.1		0	0		0	0		0	0	0	0			
Recovery																		
(mg)	6.5		7 6			3.0			3.6			2.7		4.8			28	
(%)	16.3		19.0			7.5			9.0			6.8		12 0			71	
No. of carbohydrate moieties[e]	1 2		2.0			0 49			0.35			0.45		0.66			5 2	

[a]Calculated relative to one mol of Asn [b]An amount of Tamm–Horsfall glycoprotein equivalent to 40 mg of neutral sugar was used. [c]Chromatographic analyses for GP2 (i) and GPZ2 (ii). Those under (iii) are calculated from the n.m.r. data for GPZ2, on the basis of the data in Table II. Only the identified compounds were taken into account for indicating the molar compositions derived from the n.m.r. spectra. [d]Number of residues per 80,000 g of glycoprotein. The ratios between the various sugars are similar to those calculated for the sum total of the glycopeptides (previous column), indicating that cleavage of glycosidic linkages did not occur to any significant degree during proteolytic digestion of the glycoprotein [e]Minimum numbers per 80,000 g of the glycoprotein, based on neutral-sugar recovery.

structural element[10,11] (→4)-β-GlcNAc-(1→N)-Asn. Furthermore, each of the compounds contained the trimannosyl–di-*N*-acetylchitobiose core-unit, which is usually found for *N*-glycosylically-linked carbohydrate chains. Regarding the number of *N*-acetyllactosamine-containing units, the major components of each sample were tetra-antennary in type, as indicated by the chemical shifts of the Man H-1,2 signals[11]. A small proportion of tri-antennae was also present, whereas an additional small proportion of di-antennae was observed for GPZ3 and 4.

Samples GPZ2, 3, 4, and 6 differed with respect to the nature, number, and location of terminal residues in the chain. The terminal sequences found in each sample, together with rough estimations of their relative abundances deduced from the spectral data, are listed in Table II. The marked heterogeneity of the materials did not permit determination of the precise location of each of these elements in any particular branch.

The ^{1}H-n.m.r. parameters that are indicative of the various sequences are included in Table II (*cf.* ref. 11). An Sd^a-immunodeterminant sequence, *i.e.*, β-GalNAc-(1→4)[α-NeuAc-(2→3)]-β-Gal-(1→4)-β-GlcNAc-(1→, can be recognised by the typical set of H-3 chemical shifts for NeuAc [(δ 1.925 (H-3*a*), ~2.66 (H-3*e*)] concomitant with the GalNAc H-1 doublet at δ ~4.72 ($J_{1,2}$ ~8 Hz), whereas the corresponding Gal H-1 doublet is found at δ ~4.54. At a probe temperature of 27°, the GalNAc H-1 doublet was hidden under the signal for the residual HOD in the solvent. However, the doublet could be detected after lowering the temperature to 12°. The combination of the δ value for H-1 and the $J_{1,2}$ value indicated the GalNAc–Gal linkage to be β.

The structure for GPZ2, derived from ^{1}H-n.m.r. and analytical data, is depicted in Fig. 3. Heterogeneity occurred with respect to the terminal residues. Thus, sialic acid was present in the structures α-NeuAc-(2→6)-Gal, α-NeuAc-(2→3)-Gal, and β-GalNAc-(1→4)[α-NeuAc-(2→3)]-β-Gal-(1→4) in the ratios 2:3:4. Also ~2 terminal, unsubstituted galactosyl groups per molecule were present in GPZ2. Most of the fucose present was (1→3)-α-linked to peripheral GlcNAc residues forming part of *N*-acetyllactosamine units[10,11]. The occurrence of α-Fuc-(1→3)-GlcNAc units in the chain comprising the Sd^a-immunodeterminant could not be excluded. In addition, there were indications of a small proportion of α-Fuc-(1→6)-GlcNAc units *N*-glycosylically bonded to asparagine.

The ^{1}H-n.m.r. spectrum of GPZ3 points to a higher degree of heterogeneity than for the former glycopeptide. Again, a mixture of tetra- and tri-antennary, as well as a small proportion of diantennary, structures were present. The same structural elements as described above were found at the non-reducing termini. However, there was less unsubstituted Gal than in GPZ2, and somewhat more β-GalNAc-(1→4)[α-NeuAc-(2→3)]-Gal sequences were present. The amount of fucose was similar to that in GPZ2. Finally, an additional signal was observed at δ 5.10 in the H-1α region of the spectrum, but it could not be interpreted in terms of a known structural element.

From the ^{1}H-n.m.r. spectrum, it can be concluded that GPZ4 also contained

TABLE II

^{1}H-N M R CHARACTERISTICS OF FOUR GLYCOPEPTIDE SAMPLES DERIVED FROM HUMAN TAMM-HORSFALL GLYCOPROTEIN

Structural element	*Reporter group*	*^{1}H Chemical shift*[a] *(δ)*	*Relative abundance (%)*[b] *of element in*			
			GPZ2	*GPZ3*	*GPZ4*	*GPZ6*
Carbohydrate-peptide linkage						
→4)-β-GlcNAc-(1→N)-Asn	GlcNAc-**1** H-1 GlcNAc-**1** NAc	5.06 2.013	100	100	100	100
Branching pattern						
→4) →2)-α-Man-(1→3) [**4**] β-Man-(1→ [**3**] →2)-α-Man-(1→6) [**4′**] →6)	Man-**4** H-1 Man-**4** H-2	5.12–5.13 4.213	100	100	100	100
	Man-**3** H-2	4.213	100	100	100	100
	Man-**4′** H-1 Man-**4′** H-2	4.86 4.097	90	80	80	85
→4) →2)-α-Man-(1→3) [**4**] β-Man-(1→ [**3**] →2)-α-Man-(1→6) [**4′**]	Man-**4′** H-1 Man-**4′** H-2	4.92 4.11	≤10	15[c]	20[c]	15

Terminal sequences[d]						
→4)-β-GlcNAc-(1→4)-β-GlcNAc-(1→N)-Asn 2 / 1 α-Fuc-(1→6)	GlcNAc-**2** NAc	2.091	15	15	15	15
	Fuc H-1	4.86				
	Fuc H-5	4.12				
	Fuc CH_3	1.20				
β-Gal-(1→4)-β-GlcNAc-(1→ / α-Fuc-(1→3)	Fuc H-1	5.12	25	20	20	25
	Fuc CH_3	1.179				
α-NeuAc-(2→3)-β-Gal-(1→4)-β-GlcNAc-(1→	NeuAc H-3*a*	1.80	15	15	—	30
	NeuAc H-3*e*	2.76				
	NeuAc NAc	2.031				
	Gal H-1	4.54				
	Gal H-3	4.11				
α-NeuAc-(2→6)-β-Gal-(1→4)-β-GlcNAc-(1→	NeuAc H-3*a*	1.72	10	10	—	≤10
	NeuAc H-3*e*	2.66				
	NeuAc NAc	2.031				
	Gal H-1	4.44				
β-GalNAc-(1→4)-β-Gal-(1→ / α-NeuAc-(2→3)	GalNAc H-1	4.725	20	30	≤10	40
	Gal H-1	4.54				
	Gal H-3	4.15				
	NeuAc H-3*a*	1.925				
	NeuAc H-3*e*	2.66				
	NeuAc NAc	2.031				

[a]At 27°; values with only two decimals refer to signals which could not be observed separately. [b]Based on integration of signal intensities (accuracy ~10%). The values given for the core elements and extension are expressed relative to one mol of Asn; those of the peripheral element (below the dashed line) are given relative to one β-Gal-(1→4)-GlcNAc unit. The latter can be reconciled with the former by multiplication with the number of Gal residues (see Table I, columns iii). [c]A small proportion of di-antennary (or tri-antennary) glycopeptides might be present (Man-**3** H-2, δ 4.245). [d]Not all signals in the α-anomeric region could be interpreted, and terminal sequences in addition to those listed may occur. Simultaneous occurrence of some combinations of two terminal sequences in the same branch cannot be excluded. Moreover, substantial amounts of non-sialylated, non-fucosylated β-Gal-(1→4)-GlcNAc units are present in GPZ2, 3, and 4 (Gal H-1, δ ~4.47).

α-Fuc-(1→3) 10
α-NeuAc-(2→6) 04
α-NeuAc-(2→3) 06
β-GalNAc-(1→4)
α-NeuAc-(2→3) 08

β-Gal-(1→4)-β-GlcNAc-(1→4)
8 7
β-Gal-(1→4)-β-GlcNAc-(1→2)-α-Man-(1→3)
6 5 4
β-Man-(1→4)-β-GlcNAc-(1→4)-β-GlcNAc-(1→N)-Asn
3 2 1
[α-Fuc-(1→6)] 015
β-Gal-(1→4)-β-GlcNAc-(1→2)-α-Man-(1→6)
6′ 5′ 4′
[β-Gal-(1→4)-β-GlcNAc-(1→6)] 09
8′ 7′

Fig. 3. Structure of glycopeptide GPZ2 The distribution of the 1.0 residue of α-Fuc-(1→3) at GlcNAc residues, indicated as 5, 5′, 7, and 7′, is unknown. The precise β-Gal residues to which the other non-reducing terminal groups are attached cannot be specified as yet.

a mixture of tetra- and tri- (and/or di-)antennary glycopeptides of the *N*-acetyllactosamine-type. The amount of unsubstituted Gal was relatively high and the amount of sialic acid was ≤10%. In particular, the part of the spectrum (δ 5.10–5.15) for H-1α was complex.

The ^{1}H-n.m.r. spectrum of GPZ6 indicated that mostly tetra-antennary, *N*-acetyllactosamine-type glycopeptides were involved. Almost every *N*-acetyllactosamine branch carried a substituent, namely, α-NeuAc-(2→3)-Gal, or α-Fuc-(1→3)-GlcNAc, or β-GalNAc-(1→4)[α-NeuAc-(2→3)]-Gal. The amount of the last structural feature was relatively high. In this fraction, it is possible that the structure β-GalNAc-(1→4)-Gal also occurred. There was also a small proportion of the α-NeuAc-(2→6)-Gal unit in the sample (Table II).

DISCUSSION

Tamm–Horsfall glycoprotein was rapidly and extensively digested by pronase (Fig. 1), in spite of the extensive intramolecular disulphide-bridging, with the release of all the tyrosine and tryptophan residues as free amino acids. The relative proportions of these products confirmed the analytical data reported by Fletcher *et al.*[26]. Relatively little absorbance at 230 nm was exhibited by digests of the glycoprotein, as would be expected if most peptide bonds had been cleaved, even after short periods of digestion. There may be implications in this susceptibility to proteolysis of the glycoprotein as to the fate of that part of it which is associated with the basal plasma membranes of the renal tubular cells in the living animal[3].

The glycopeptide fractions obtained as a result of chromatography on DEAE-Sephadex A25 were all sufficiently similar in amino acid composition (Table I) as to suggest that the peptide content is unlikely to be the cause of the fractionation pattern observed (Fig. 2). Treatment of glycopeptides, after benzyloxycarbonylation, with carboxypeptidase led to little change in the amino acid compositions, apart from GP6. In this glycopeptide, the decrease in the quantity of proline, as well as of other amino acids, suggested that the manipulative procedures in reisolating the acylated glycopeptide, GPZ6, caused removal of a contaminating peptide, at least in part, rather than a shortening of the peptide chain of the glycopeptide.

Differential binding of the glycopeptides to the positively charged chromatographic support, in 0.1M acetic acid (pH 2.8), is probably due, at least in part, to the presence of sialic acid residues and to the type and position of linkages in which they are involved. The pK_a of the carboxyl group of sialic acid residues in glycoconjugates, like those of ionising groups of amino acids in proteins, depends on its precise location within the macromolecule[27]. The pK_a value for this group in the non-reducing terminal trisaccharide β-GalNAc-(1→4)[α-NeuAc-(2→3)]-β-Gal-(1→, which is present in largest amounts in the glycopeptide (GP6) bound most strongly to DEAE-Sephadex A25, is likely to be lower by 1–2 units than that of the sugar residue present in the disaccharide α-NeuAc-(2→3)-β-Gal-(1→, because of

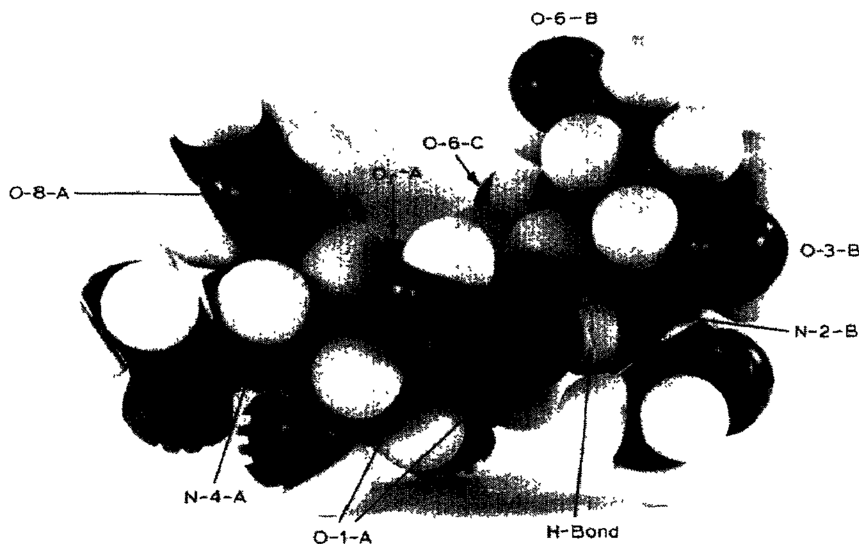

Fig. 4. A molecular model of the trisaccharide terminal unit α-NeuAc-(2→3)[β-GalNAc-(1→4)]-Gal→. Pertinent atoms are identified, with A, B, and C referring to NeuAc, GalNAc, and Gal residues, respectively.

hydrogen bonding between the carboxyl group of NeuAc and the acetamido nitrogen of GalNAc in the trisaccharide. Thus, construction of a molecular model of the trisaccharide on the basis of the ^{13}C-n.m.r. data[28] for gangliosides G_{M1} and G_{D1a}, but with the NAc group of the GalNAc residue oriented such that its carboxyl oxygen is *cis* to H-2 of the sugar moiety, the conformation normally adopted by this group[29,30], yields a structure in which a hydrogen bond is present (Fig. 4). Studies aimed at testing the proposed model, using two-dimensional n.m.r. techniques are in progress.

The overall content of sialic acid residues of the glycopeptides (Table I) and the quantities of the terminal trisaccharide structure (Table II) might explain why GP2, 3, and 6 increase in acidity in that order (Fig. 2). The reason for the relatively strong binding of GP4 to DEAE-Sephadex A25 at pH 2.8 is not obvious, but the complexity of the α-anomeric region of the ^{1}H-n.m.r. spectrum remains to be explained in terms of structural features.

The results of chemical analysis of GPZ2, which is a main product, accord with the n.m.r. data and indicate a tetra-antennary type structure (Table I) having the non-reducing terminal groups shown in Fig. 3. However, there are fairly large differences between the values of sugars found by chemical analysis, especially for GPZ4 and 6, and those calculated on the basis of the n.m.r. data for the identified, mainly tetra-antennary, glycopeptides with respect to mannose and galactose. These glycopeptides are likely to have structures more complex than that of GPZ2. One or more of them may carry an immunodeterminant (closely related to that in

GPZ2), β-GalNAc-(1→4)[NeuAc-(2→3)]-β-Gal-(1→4)-β-GlcNAc-(1→3)-Gal, a pentasaccharide which was isolated in small yield after treatment of Tamm–Horsfall glycoprotein with β-D-galactosidase[31].

These n.m.r. studies provide independent evidence for the presence of β-GalNAc residues, which was proposed previously[31–34] on the basis of chemical and enzymic studies. The ^{1}H-n.m.r. parameters found for this structural element were previously shown for oligosaccharides derived from gangliosides[35], as well as for the Cad-blood-group determinant obtained as an oligosaccharide-alditol from digests of erythrocyte-membrane glycophorin[36].

The presence of β-GalNAc residues in Tamm–Horsfall glycoprotein is consistent with the occurrence of the Sd^{a} immunodeterminant in the molecule[33,34], and with the known specificity of the lectin *Dolichos biflorus* for both α and β anomers of this type of sugar residue[37], but the group necessary for the blood-group specificity contains both sialic acid and GalNAc residues[31]. The present studies confirm that the former sugar is in (2→3)-α-linkage to Gal which is substituted also by a (1→4)-linked β-GalNAc residue, and this forms part of the immunodeterminant group, or groups, associated with Sd^{a}-positive Tamm–Horsfall glycoprotein. The rest of the carbohydrate to which the substituted galactosyl residue is attached differs from that in the Cad immunodeterminant occurring on glycophorin A and B isolated from Cad erythrocytes[36].

If all the GalNAc residues are present in Sd^{a}-immunodeterminant structures, the preparation of pooled Tamm–Horsfall glycoprotein contained at least six residues of the determinant per mol of glycoprotein (Table I), but they were distributed unevenly and sometimes more than one per carbohydrate moiety. GPZ6, representing 0.66 of an overall carbohydrate moiety in the original glycoprotein, carried 1.5–2 immunodeterminant groups per moiety, on average, in the non-reducing terminal positions of the carbohydrate units (Table I), whereas GPZ3 (0.49 overall of a group) carried only one such determinant per tetra-antennary unit. The way in which the distribution of these groups, within a given carbohydrate moiety and between carbohydrate entities in the whole glycoprotein, affects the overall Sd^{a} immunoreactivity of Tamm–Horsfall glycoprotein is unknown. Until we know more about the structure of the protein chain, and, in particular, if there is more than one type in the 80 kDa subunit[38], any speculation concerning distribution of the carbohydrate moieties on the chain(s) is of little value.

Of the total sialic acid present in the glycopeptides ~50% occurs as part of the Sd^{a} immunodeterminant structure, and these may be expected to be resistant to the action of various neuraminidases[39–41]. The occurrence of these entities probably explains, in part at least, earlier observations about the resistance to neuraminidase of a large fraction of the sialic acid residues present in the glycoprotein[42,43].

The carbohydrate moieties are heterogeneous in other ways. Not only is there evidence for differences in the extent and position of fucosylation and for sialylation

at points not involved in the Sd[a] immunodeterminant, but there are some triantennary, and to a lesser extent diantennary, structures also present.

ACKNOWLEDGMENTS

This investigation was supported by The Medical Research Council, the Netherlands Foundation for Chemical Research (SON), the Netherlands Organization for the Advancement of Pure Research (ZWO), and by the Netherlands Foundation for Cancer Research (KWF), grant UUKC 83-13. We thank Mr. James Balmer for help in preparing the Figs.

REFERENCES

1 I. TAMM AND F. J. HORSFALL, *Proc. Soc. Exp. Biol. Med.*, 74 (1950) 108–114.
2 I. TAMM AND F. J. HORSFALL, *J. Exp. Med.*, 95 (1952) 71–97.
3 K. L. SIKRI, C. L. FOSTER, F. J. BLOOMFIELD, AND R. D. MARSHALL, *Biochem. J.*, 181 (1979) 525–532.
4 J. R. HOYER, S. P. SISSON, AND R. L. VERNIER, *Lab. Invest.*, 41 (1979) 168–173.
5 K. L. SIKRI, C. L. FOSTER, N. MCHUGH, AND R. D. MARSHALL, *J. Anat.*, 132 (1981) 597–605.
6 T. MAACK AND E. H. WINDHAGER, in D. BLACK AND W. F. JONES (Eds.), *Renal Disease*, 4th edn., Blackwell, Oxford, 1979, pp. 107–138.
7 A. GOTTSCHALK, *Nature (London)*, 170 (1952) 662–663.
8 A. GOTTSCHALK, in G. E. W. WOLSTENHOLME AND M. O'CONNOR (Eds.), *Chemistry and Biology of Mucopolysaccharides*, Churchill, London, 1958, pp. 287–295.
9 A. M.-M. AFONSO, P. A. CHARLWOOD, AND R. D. MARSHALL, *Carbohydr. Res.*, 89 (1981) 309–319.
10 J. F. G. VLIEGENTHART, H. VAN HALBEEK, AND L. DORLAND, *Pure Appl. Chem.*, 53 (1981) 45–77.
11 J. F. G. VLIEGENTHART, L DORLAND, AND H. VAN HALBEEK, *Adv. Carbohydr. Chem. Biochem.*, 41 (1983) 209–374.
12 H. VAN HALBEEK, L. DORLAND, J. F. G. VLIEGENTHART, W. E. HULL, G. LAMBLIN, M. LHERMITTE, A. BOERSMA, AND P. ROUSSEL, *Eur. J. Biochem.*, 127 (1982) 7–20.
13 H. VAN HALBEEK, L. DORLAND, J. F. G. VLIEGENTHART, N. K. KOCHETKOV, N P ARBATSKY, AND V. A. DEREVITSKAYA, *Eur. J. Biochem.*, 127 (1982) 21–29.
14 A. A. GOODALL AND R. D. MARSHALL, *Biochem. J.*, 189 (1980) 533–539
15 R. PARTON, *J. Gen. Microbiol.*, 89 (1975) 113–123.
16 U. K. LAEMMLI, *Nature (London)*, 227 (1970) 680–685.
17 K. WEBER AND M. OSBORN, *J. Biol. Chem.*, 244 (1969) 4406–4412.
18 R. D. MARSHALL AND P. ZAMECNIC, *Biochim. Biophys. Acta*, 181 (1969) 454–464.
19 S. MOORE AND W. H. STEIN, *J. Biol. Chem.*, 211 (1954) 907–913.
20 L. C. CRAIG AND W. KONIGSBERG, *J. Phys. Chem.*, 65 (1961) 166–172.
21 V. P. BOGDANOV, E. D. KAVERZNEVA, AND TSI DE-FAN, *Abstr. Int. Congr. Biochem., 5th, Moscow*, 1961, p. 25.
22 A. P. FLETCHER, G. S MARKS, R D. MARSHALL, AND A. NEUBERGER, *Biochem. J.*, 87 (1963) 265–273.
23 C. FRANÇOIS, R. D. MARSHALL, AND A. NEUBERGER, *Biochem. J.*, 83 (1962) 335–341.
24 T. BHATTI, R. E. CHAMBERS, AND J. R. CLAMP, *Biochim. Biophys. Acta*, 222 (1970) 339–347.
25 G. M. PHIMISTER AND R. D. MARSHALL, *Biochem. Soc. Trans.*, 11 (1983) 769–770.
26 A. P. FLETCHER, A. NEUBERGER, AND W. A. RATCLIFFE, *Biochem J.*, 120 (1970) 417–424.
27 R. W. LEDEEN AND R. K. YU, in A. ROSENBERG AND C.-L. SCHENGRUND (Eds.), *Biological Role of Sialic Acid*, Plenum Press, New York, 1976, pp. 1–57.
28 P. L. HARRIS AND E. R. THORNTON, *J. Am. Chem. Soc.*, 100 (1978) 6738–6745.
29 R. D. GILARDI AND J. L. FLIPPEN, *Acta Crystallogr., Sect. B*, 30 (1974) 2931–2933
30 A. NEUMAN, H. GILLIER-PANDRAUD, F. LONGCHAMBON, AND D. RABINOVICH, *Acta Crystallogr., Sect. B*, 31 (1975) 474–477.

31 A. S. R. DONALD, A. D. YATES, C. P. C. SOH, W. T. J. MORGAN, AND W. M. WATKINS, *Biochem. Biophys. Res. Commun.*, 115 (1983) 625–631.
32 C. P. C. SOH, W. T. J. MORGAN, W. M. WATKINS, AND A. S. R. DONALD, *Biochem. Biophys. Res. Commun.*, 93 (1980) 1132–1139.
33 A. S. R. DONALD, C. P. C. SOH, W. M. WATKINS, AND W. T. J. MORGAN, *Biochem. Biophys. Res. Commun.*, 104 (1982) 58–65.
34 F. SERAFINI-CESSI AND F. DALL'OLIO, *Biochem. J.*, 215 (1983) 483–489.
35 L. DORLAND, H. VAN HALBEEK, H. WIEGANDT, R. SCHAUER, AND J. F. G. VLIEGENTHART, *Eur. J. Biochem.*, (1984) in press.
36 D. BLANCHARD, J.-P. CARTRON, B. FOURNET, J. MONTREUIL, H. VAN HALBEEK, AND J. F. G. VLIEGENTHART, *J. Biol. Chem.*, 258 (1983) 7691–7695.
37 D. HAMMARSTRÖM, L. A. MURPHY, I. J. GOLDSTEIN, AND M. E. ETZLER, *Biochemistry*, 16 (1977) 2750–2755.
38 A. P. FLETCHER, A. NEUBERGER, AND W. A. RATCLIFFE, *Biochem. J.*, 120 (1970) 425–432.
39 G. DAWSON, in M. I. HOROWITZ AND W. PIGMAN (Eds.), *The Glycoconjugates*, Academic Press, New York, Vol. II, 1978, pp. 287–336.
40 K. SUZUKI, in A. ROSENBERG AND C.-L. SCHENGRUND (Eds.), *Biological Roles of Sialic Acid*, Plenum Press, New York, 1976, pp. 159–181.
41 A. ROSENBERG AND C.-L. SCHENGRUND, in A. ROSENBERG AND C.-L. SCHENGRUND (Eds.), *Biological Roles of Sialic Acid*, Plenum Press, New York, 1976, pp. 295–359.
42 E. KLENK, H. FAILLARD, AND H. LEMPFRID, *Hoppe-Seyler's Z. Physiol. Chem.*, 301 (1955) 235–246.
43 H. FAILLARD, *Hoppe-Seyler's Z. Physiol. Chem.*, 307 (1957) 62–86.