Glycoconjugate Journal (1998) **15:** 275–282

Influence of the amino acid sequence on the MUC5AC motif peptide *O*-Glycosylation by Human gastric UDP-GalNAc : polypeptide *N*-acetylgalactosaminyltransferase(s)

Sylviane Hennebicq[1], Daniel Tetaert[1]*, Benoit Soudan[1,2], Arnold Boersma[1,2], Gilbert Briand[3], Colette Richet[1], Jean Gagnon[4] and Pierre Degand[1]

[1]*Unité INSERM N° 377, Bâtiment G. BISERTE, Place de Verdun, 59045 Lille Cédex (France)*
[2]*Laboratoire de Biochimie et Endocrinologie, USN A, 6 Rue Pr Laguesse, 59045 Lille Cédex (France)*
[3]*Service Commun de Spectrométrie de Masse, Université de Lille II, Faculté de Médecine, Place de Verdun, 59045 Lille Cédex (France)*
[4]*Institut de Biologie Structurale, CEA/CNRS, 41 Avenue des Martyrs, 38027 Grenoble Cédex (France)*

The present work was carried out to study the role of the peptide moiety in the addition of *O*-linked *N*-acetylgalactosamine to human apomucin using human crude microsomal homogenates from gastric mucosa (as enzyme source) and a series of peptide acceptors representative of tandem repeat domains deduced from the MUC5AC mucin gene (expressed in the gastric mucosa). Being rich in threonine and serine placed in clusters, these peptides provided several potential sites for *O*-glycosylation. The glycosylated products were analysed by a combination of electrospray mass spectrometry and capillary electrophoresis in order to isolate the glycopeptides and to determine their sequence by Edman degradation. The *O*-glycosylation of our MUC5AC motif peptides gave information on the specificity and activity of the gastric microsomal UDP-*N*-acetylgalactosamine : polypeptide *N*-acetylgalactosaminyltransferase(s). The proline residues and the induced-conformations are of great importance for the recognition of MUC5AC peptides but they are not the only factors for the choice of the *O*-glycosylation sites. Moreover, for the di-glycosylated peptides, the flanking regions of the proline residues strongly influence the site of the second *O*-glycosylation.

***Keywords*: UDP-GalNAc : polypeptide *N*-acetylgalactosaminyltransferase, *O*-glycosylation, mucin, MUC5AC, gastric mucosa, capillary electrophoresis, Edman degradation**

Introduction

Although threonine or serine residues in proline-rich domains are often suggested as *N*-acetylgalactosamine (GalNAc) *O*-glycosylation sites [1], a number of studies have attempted to identify the peptide motif(s) that predict the site of *O*-glycosylation. These works have either focused on comparing amino acid sequences that flanked the O-linked threonine/serine residues in glycoproteins [2–6] or assessed the *in vitro* glycosylation of model peptides using the UDP-GalNAc : polypeptide *N*-acetylgalactosaminyl-transferase (GalNActransferase), in crude or purified forms [7–18]. From these studies, it has been proposed that the ideal substrate for *in vitro O*-glycosylation reactions could consist of potential serine/threonine residues surrounded on both sides by serine, threonine, proline, glycine and alanine [5, 19, 20] whereas basic and acidic amino acid residues (aspartic acid, glutamic acid, arginine) disturb the GalNAc transferase substrate recognition by their charge distribution and/or charge density modifications [20]. Nevertheless, the recent findings that various species of GalNAc transferases may exist [21], can partly explain the lack of a consensus peptidic sequence for the first step of *O*-glycosylation [3, 7].

Moreover, the *O*-glycosylation may be strongly influenced by the local conformation and tertiary structure of the protein [6, 22]. The *O*-linked glycosylation sites have been predicted to be located within the β-turn either in the second or the third position of the turn [23, 24] and synthetic peptides which are *O*-glycosylated *in vitro* adopt a β-turn as shown by circular dichroism [25]. As a consequence, the proline residues (involved in *cis*/*trans*

* To whom correspondence should be addressed. Fax: (33) 03 20 53 85 62; E-mail: tetaert@biserte.lille.inserm.fr.

0282–0080

conformations) [26] located in the amino acid sequences near *O*-glycosylation sites in most glycoproteins, like mucins, probably cause a conformational change in the polypeptide chain that is essential for the polypeptide GalNActransferase activity [22, 27, 28].

Up to now, few studies have been focused on the *O*-glycosylation sites of mucins. The contribution of primary amino acid sequence to the substrate specifity of human GalNActransferase(s) was recently studied using as acceptor substrates a set of peptides based on the tandem repeat of the epithelial MUC1 and secretory intestinal MUC2 mucins [15, 18]. In our report, the *O*-glycosylation by gastric GalNAc transferase(s) of MUC5AC motif peptides (naturally occurring in the polypeptide backbone of a mucin expressed in the gastric mucosa [29]) have been investigated. An extensive panel of peptides representative of a MUC5AC tandem repeat motif has been selected, with threonine-rich (*TTS*AP*TTS* as well as dimer-form: *TTS*AP*TTS TTS* AP*TTS*) and proline-rich (GTT*PSPVP*) sequences. Capillary electrophoresis at analytical and preparative scales, combined with mass spectrometry and Edman degradation, have proved to be very useful tools to perform this analysis of glycosylated peptides.

Materials and methods

Materials

AG 1-X8 (100–200 mesh, Cl^- form) was purchased from BioRad, (Vitry s/Seine, France). UDP-*N*-[1-^{3}H]acetylgalactosamine (UDP-GalNAc) (8.7 Ci mmol^{-1}) was from New England Nuclear Chemicals (Dupont de Nemours, Les Ulis, France). MES (2-(*N*-morpholino)ethanesulfonate), PMSF (phenyl methane sulfonyl fluoride), AMP (adenosine 5′ monophosphate), E64 (trans-epoxysuccinyl leucylamino (4-guanido) butane), EDTA, Triton X-100, cold UDP-GalNAc were from Sigma (St Louis, MO, USA). Polyvinyl alcohol (MW 15000) was from Fluka (Buchs, Switzerland).

Peptide substrates

The synthesis of the peptide including an aza-proline residue (CαH of proline is replaced by N) in the sequence TTSA-(AzP)-TTS was carried out in liquid phase and the *cis*-prolyl configuration was checked [30]. The different synthetic peptide substrates were from Neosystem (Strasbourg, France): Threonine-rich peptides (TTSAPTTS, TTSASTTS, TTSSPTTS, TTSTTSAP, TTSTTSGP, TTSGPGTT, TTSGAGTT, TTSAPTTSTTSAPTTS) and proline-rich peptides (GTTPSPVP, PSPVPTTS, TTSGPGTTPSPVPTTS). The purity (>95%) of the peptides was assessed by high performance liquid chromatography analysis and capillary electrophoresis, and their masses were verified by electrospray mass spectrometry. Their amino acid sequences were also checked using an Applied Biosystems gas-phase Sequencer 477A.

Enzyme preparation

Macroscopically normal fundic human gastric samples (about 3 g) were obtained from three different patients. Microsome suspensions were prepared by Potter-Elvehjem homogenization of the specimen tissues as previously described [31]. The microsomal pellets were suspended in a 0.2 M NaCl–0.25 M sucrose solution to obtain a final protein concentration of 2.8 to 5.7 mg ml^{-1} and the preparations were stored at −80 °C until use. The protein concentration was determined on a multiparametric automated chemistry Hitachi 717 analyser (Boehringer, Mannheim, Germany) using a Biotrol kit protein assay reagent (Biotrol, Chennevières les Louvres, France).

Analytical methods

In vitro O-*glycosylation reactions*

Assay mixtures in a total volume of 80 µl contained the following components at final concentrations: 3 mM peptide (20 µl); 125 mM MES buffer (pH 7.0) (20 µl) containing 0.2% (v/v) Triton X-100, 12.5 mM $MnCl_2$, 1 mM E64, 1 mM PMSF; 1.25 mM AMP (10 µl); 1 mM UDP-*N*-acetylgalactosamine diluted with 2.5–5 nM UDP-*N*-acetyl[1-^{3}H]galactosamine (0.3×10^6–0.5×10^6 dpm) (10 µl), and 30 µg of microsomal preparation (20 µl). After the addition of the enzyme preparation, samples were incubated for 2 h at 37 °C. The reaction was stopped by the addition of eight volumes (640 µl) of 20 mM sodium borate–1 mM ethylenediaminetetraacetic acid (pH 9.1). The reaction products were passed through AG 1-X8 resin (2 ml), eluted by 3 ml water and the incorporation rates were routinely calculated by scintillation counting of aliquots of the samples with subtraction of values obtained from controls incubated without peptide substrate.

In order to desalt the samples before electrospray mass spectrometry and capillary electrophoresis, Sep-Pak C18 reversed-phase cartridges (Millipore, Waters chromatography, Milford, MA, USA) were used. The cartridges were activated by the passage of 10 ml methanol followed by 10 ml deionized water containing counterion trifluoroacetic acid (TFA, 0.1%). Aqueous eluates from the exchanger AG 1-X8 were applied and polar materials were washed with 10 ml water/TFA whereas the glycopeptidic fractions were obtained by elution with 10 ml acetonitrile 25% in water/TFA.

Mass spectrometry

All mass spectra were obtained on a Perkin-Elmer SCIEX API-I one-quadrupole mass spectrometer (Toronto, Ontario, Canada) equipped with an atmospheric pressure ionization source [32]. The mass spectrometer was operated in the positive mode and polypropylene glycol was used to calibrate the quadrupole. The potential of the spray needle was held at 5.5 kV and the orifice voltage was 90 V. The

samples (approximatively 0.1 μg of peptide and related products in 100 μl), dissolved in 20% acetonitrile/water containing 0.1% formic acid were introduced into the source with a medical infusion pump (model 11, Harvard apparatus, South Natick, USA) at a flow rate of 3 μl min^{-1}. The instrument was scanned in the multichannel analyser/acquisition mode (MCA) over the mass range *m/z* 200–2380 with a 40 s scan time and 5–10 spectra were summed. The Mac Biospec computer program (Macintosh Quadra 700 computer, Cupertino, CA, USA) was used to calculate the molecular masses of the different products in the samples.

Capillary electrophoresis

Capillary electrophoresis was performed on a P/ACE system Model 5000 (Beckman, Fullerton, CA, USA) controlled by the System Gold software V810 (Beckman). UV Absorbance was monitored by a fixed-wavelength detector at 200 nm. The fused silica capillary was 57 cm long × 75 μm internal diameter and fitted in a cartridge with a modified mandrel in order to improve cooling. Migrations were run at 23 °C using a 2 N formic acid buffer and the voltage across the capillary was maintained at 15 kV. The injections were carried out by pressure (duration 2–6 s, approximately 10–20 pmol) and after each separation the capillary was flushed with the appropriate buffer for 1 min [33]. For the separation of the hexadecapeptides, the formic acid 2 N buffer contained 2.5% polyvinyl alcohol MW 15 000 (v/v).

In order to determine the *O*-linkage sites of the peptide substrates, a preparative scale procedure was performed as described in [34] with cumulative collection at their precise migration times of the glycosylated fractions that had been submitted to Edman degradation. The amount of available glycopeptides varied between 1 and 3 nmol and 10 to 15 runs were performed to collect each fraction in sufficient amount.

Identification of O-*glycosylation sites*

The sequencing procedure to determine the positions of glycosylation was the same as that described in [35]. An aliquot of each sample was covalently attached to a Sequalon AA™ membrane (Millipore, USA) according to the manufacturer's instructions. The membrane disc was then placed directly in the protein sequencer (Applied Biosystems gas-phase Sequencer, model 477A) coupled to an HPLC 120A system for the on-line analysis of the phenylhydantoin-derivatives.

The *O*-GalNAc-linked sites were identified as follows: (i) the modification of the repetitive yield percentage that decreased at a *O*-glycosylation position (when compared to the unglycosylated peptide substrate) and was artificially higher at the following cycle; (ii) the appearance of additional PTH products corresponding to the expected PTH-Thr-(*O*-GalNAc), as described by Pisano *et al.* [12], but unfortunately too close to the PTH-Ser under our HPLC conditions for its quantification. The quantity of glycopeptide used for sequence determination was more than 100 pmol.

Results

Incorporation of GalNAc in MUC5AC motif peptides

The effects of varying assay parameters (time, substrate and enzyme concentrations) were first evaluated in order to assess the *O*-glycosylation reaction during its exponential phase of development (data not shown). The parameters were as follows: 2 h of incubation time, 3 mM peptide concentration and 0.35–0.40 g l^{-1} enzyme concentration. Each peptide substrate was tested in duplicate and the results were expressed as mean ± standard error of mean. The relative rates of GalNAc transferase activities (nmol h^{-1} per mg of microsomal proteins) are given in detail in Table 1. The results were comparable whatever the gastric microsomal preparation used (healthy gastric mucosa from three different patients tested in three different experiments).

In our study, the position of the proline residue in the peptide sequence was of importance for the *N*-actylgalactosamine incorporation rate (Table 1). The *O*-glycosylation was enhanced by proline residues placed on the *C*-terminal side of the peptide backbone: for the MUC5AC threonine-rich octapeptide substrates, the incorporation was 211.2 ± 3.7 nmol h^{-1} per mg proteins for TTSTTSAP (*ie* proline in C-terminal) and two-fold lower for TTSAPTTS (104.4 ± 0.4 nmol h^{-1} per mg proteins) when proline is inside the sequence. A similar result was obtained comparing TTSTTSGP (86.9 ± 0.3 nmol h^{-1} per mg proteins) to TTSGPGTT (42.9 ± 0.2 nmol h^{-1} per mg proteins). In the absence of a proline residue in the peptide sequence, we observed a significant decrease in the GalNAc incorporation (TTSASTTS [24.3 ± 3.2 nmol h^{-1} per mg proteins] compared to TTSAPTTS, and TTSGAGTT [19.6 ± 2.0 nmol h^{-1} per mg proteins] compared to TTSGPGTT). The positive effect of proline residues on the affinity of the gastric GalNAc transferase(s) for the MUC5AC motif peptide substrates was partly confirmed by the GalNAc incorporation rate of the proline-rich octapeptide GTTPSPVP (542.7 ± 13.6 nmol h^{-1} per mg proteins). This peptide contained the PSPVP amino acid sequence in its C-terminal side and it appeared that the place of this proline-rich sequence in the peptide backbone was of great importance, since the GalNAc incorporation rate was only of 23.4 ± 0.2 nmol h^{-1} per mg proteins for the peptide PSPVPTTS when the sequence PSPVP was in the N-terminal position. The negative influence of a peptide *cis prolyl* conformation for *O*-glycosylation was also assessed using the TTSA-(AzP)-TTS peptide as substrate. GalNAc incorporation in this peptide was strongly decreased (24.6 ± 0.3 nmol h^{-1} per mg proteins) compared to the rate obtained with TTSAPTTS (104.4 ± 0.4 nmol h^{-1} per mg proteins) and quite comparable to the rate obtained with TTSASTTS (24.3 ± 3.2 nmol h^{-1} per mg proteins) (Figure 1).

Table 1. Capillary electrophoretic features, electrospray mass spectrometry data, *N*-acetylgalactosamine rates, and sites of *O*-glycosylation of the MUC5AC glycopeptides.

Peptides			*Glycopeptides*		
Sequence and m/z *values (mono charged forms)*	*Incorporation rates (nmolGalNAc h^{-1} per mg proteins)*	*Capillary electrophoresis relative migration migration times (min)*	m/z *(mono charged forms)*	*Capillary electrophoresis relative migration migration times (min)*	*O-linked positions (Edman degradation) in the sequence*
Proline-rich MUC5AC motif peptides					
TTSGPGTT **PSPVP** TTS	552.8 ± 3.1	41.1	1690.7	42.84	Position 8
($m/z = 1487.7$)			1893.9	44.74	Positions 8 and 7
PSPVP TTS	23.4 ± 0.2	19.3	990.4	20.01	NI
($m/z = 787.4$)					
GTT **PSPVP**	542.7 ± 13.6	18.6	958.4	20.02	Position 2
($m/z = 755.4$)			1161.5	21.26	Positions 2 and 3
Threonine-rich MUC5AC motif peptides					
TTS **GP** GTT	42.9 ± 0.2	17.9	924.4	19.57	Position 2
($m/z = 721.1$)					
TTS TTS **GP**	86.9 ± 0.3	20.2	955.4	21.66	Position 5
($m/z = 752.4$)			1158.4	22.88	Positions 5 and 2
TTS **GA** GTT	19.6 ± 2.0	17.75	898.4	19.55	Position 2
($m/z = 695.4$)					
TTS **AS** TTS	24.3 ± 3.2	21.0	958.4	23.0	Position 2
($m/z = 755.3$)					
TTS **SP** TTS	80.5 ± 1.9	17.94	984.4	19.34	Position 2
($m/z = 781.3$)					
TTS **A** (**AzP**) TTS	24.6 ± 0.3	21.9	969.4	23.4	Position 2
($m/z = 766.4$)					
TTS **AP** TTS	104.4 ± 0.4	22.4	968.5	24.2	Position 2
($m/z = 765.4$)					
TTSTTS **AP**	211.2 ± 3.7	18.4	968.5	19.62	Position 5
($m/z = 765.4$)			1171.5	20.69	Positions 5 and 4
TTS **AP** TTSTTS **AP** TTS	246.8 ± 0.4	39.9	1714.8	41.7	Position 10
($m/z = 1511.7$)			1917.8	42.3	Positions 10 and 2

NI, non identified.

Moreover, the nature of the amino acid placed before the proline was of great interest. Indeed, replacement of the alanine in TTSTTSAP by a glycine in TTSTTSGP deeply reduced the GalNAc incorporation rate (211.2 ± 3.7 nmol h^{-1} per mg proteins for TTSTTSAP and 86.9 ± 0.3 nmol h^{-1} per mg proteins for TTSTTSGP). This effect might be attributed to the structural disorganization effect of the glycine residue. The replacement of alanine by a serine in TTSAPTTS also decreased the GalNAc incorporation rate (104.4 ± 0.4 nmol h^{-1} per mg proteins for TTSAPTTS and 80.5 ± 1.9 nmol h^{-1} per mg proteins for TTSSPTTS).

Furthermore, the GalNAc incorporation rates for the hexadecapeptides and their most glycosylated corresponding octapeptides were similar: 211.2 ± 3.7 nmol h^{-1} per mg proteins for TTSTTSAP and 246.8 ± 0.4 nmol h^{-1} per mg proteins for TTSAPTTSTTSAPTTS; 542.7 ± 13.6 nmol h^{-1} per mg proteins for GTTPSPVP and 552.8 ± 3.1 nmol h^{-1} per mg proteins for TTSGPGTTPSPVPTTS. Thus, under our conditions, the peptide length was not a parameter that greatly influenced the intensity of the peptide substrate *O*-glycosylation.

Analysis of the *N*-acetylgalactosaminylated peptides

The analysis of the *O*-glycosylated products by capillary electrophoresis complemented by electrospray mass spectrometry (ESMS), allowed us to determine the number of *N*-acetylgalactosamine O-linkage(s) (Table 1). For capillary electrophoresis analysis, the relative migration times of the *O*-GalNAc linked peptides were deduced by comparing the profiles of the reaction products obtained either in the presence or absence of UDP-GalNAc. For example, after

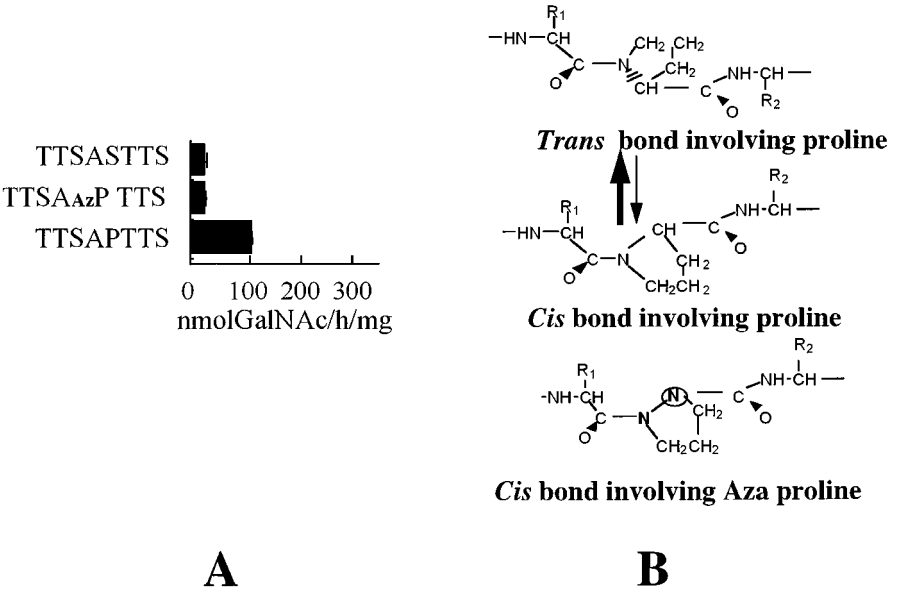

Figure 1. Effect of *cis trans prolyl* conformation of the peptide TTSAPTTS on the *in vitro O*-glycosylation (A) and structure of the different conformations of the peptide (B).

the *O*-glycosylation reaction using the GTTPSPVP peptide as substrate, seven major fractions migrated between 15 and 25 min (Figure 2). On the other hand, the capillary electrophoresis profiles of the products resulting from incubation without UDP-GalNAc showed only five fractions which corresponded to peptidic fragments and to the residual peptide. This allowed us to quantify the peptidic degradation, by integrating the peaks corresponding to the peptidic fragments. This degradation varied between 10 and 20% for all the eight residues peptides, except for the highly glycosylated GTTPSPVP peptide where it was only about 5%. Moreover, the two extra fractions (F2 and F3) at 20.02 min and 21.26 min were attributed to the glycopeptides. The formation of these two glycopeptides was confirmed by ESMS analysis of the resulting *O*-glycosylation products, showing $m/z = 958.4$ and 1161.5 which respectively corresponded to the addition of one and two *N*-acetylgalactosamines on the GTTPSPVP peptide at $m/z = 755.4$ (the *O*-linkage of *N*-acetylgalactosamine induces a 203 addition of mass). After isolation by capillary electrophoresis at the preparative scale, the fractions F2 and F3 were submitted to the Edman degradation in order to determine the sites of *O*-glycosylation (Figure 3). The degradation repetitive yield for each fraction was expressed as a percentage of the degradation yield for the unglycosylated peptide. The fraction F2 corresponded to the monoglycosylated peptide (*O*-linkage on position 2) and the fraction F3 to the di-glycosylated GTTPSPVP (*O*-linkage on positions 2 and 3) indicating that position 2 was the first substituted.

Another example was the hexadecapeptide TTSGPGTTPSPVPTTS described in Figure 4. The separation by capillary electrophoresis of the hexadecapeptide and its related glycopeptides with delayed migration times (Boulis *et al.*, unpublished) was only efficient using 2.5% polyvinyl alcohol in the formic acid buffer. The native peptide TTSGPGTTPSPVPTTS migrated at 41.1 min and the corresponding glycosylated peptides migrated at 42.84 min and 44.74 min (Figure 4A). The electrospray mass spectrometry analysis of the *O*-glycosylation products (Figure 4B) showed the different peptides under their mono $(M+H^+)^+$ and double charged $(M+2H^+)^{2+}$ forms. The TTSGPGTTPSPVPTTS peptide corresponded to m/z 1487.7 and 744.5; the monoglycosylated peptide to m/z 1690.8 and 846.0; and the di-glycosylated peptide to m/z 1893.7 and 947.4. A double charged form $(MH^+ + K^+)^{2+}$ of the unglycosylated hexadecapeptide was also identified at $m/z = 763.5$. It is of interest to notice that there was always more double charged peptide forms than single ones when hexadecapeptide glycosylation products were evaluated by electrospray mass spectrometry unlike the octapeptide analysis.

For each peptide substrate, similar analyses were conducted on the *O*-glycosylated products and, after preparative procedure by capillary electrophoresis, the glycopeptides were submitted to the Edman degradation for the determination of the exact GalNAc *O*-linkage sites (Table 1). Considering the octapeptides TTSTTSAP and TTSTTSGP, we observed that the nature of the amino acid preceding a proline appeared not to influence the first glycosylation site (position 5) but strongly affected the second ones (position 4 for TTSTTSAP and position 2 for TTSTTSGP). This phenomenon was not observed for TTSAPTTS,

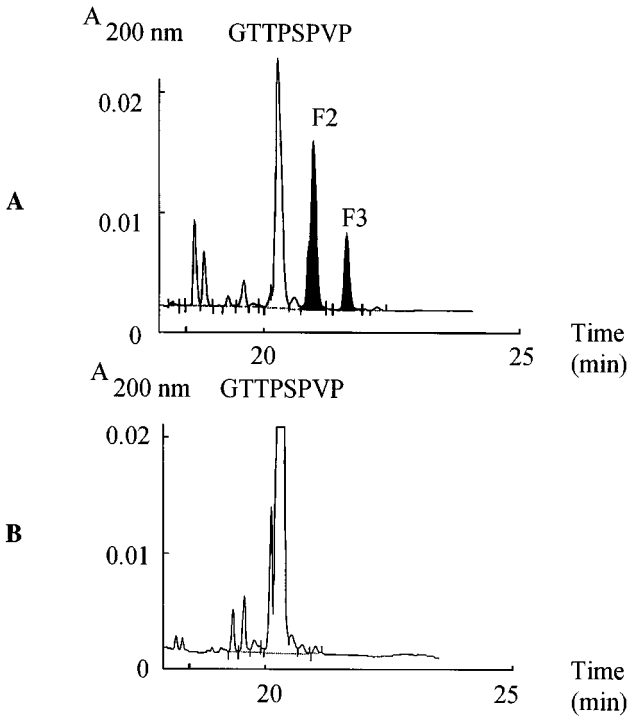

Figure 2. Comparison of capillary electrophoresis profiles of the reaction products obtained by incubation of the peptide GTTPSPVP with a gastric microsomal preparation (A) in presence of UDP-GalNAc (*O*-glycosylation reaction); (B) in absence of UDP-GalNAc. The peaks F2 and F3 correspond respectively to the monoglycosylated and the diglycosylated peptide.

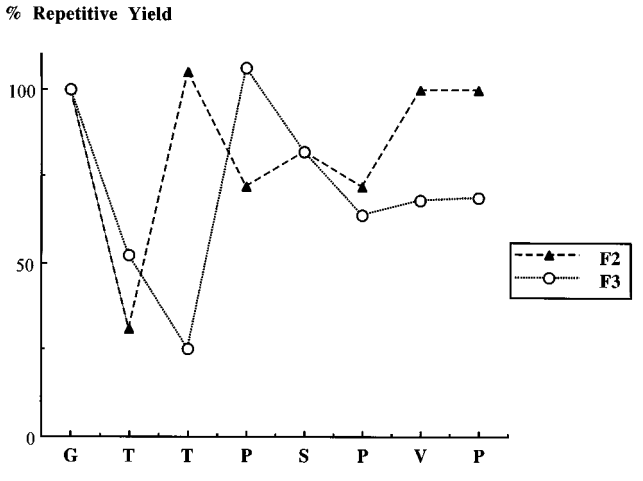

Figure 3. Edman degradation percentage of repetitive yield for fractions F2 (▲) and F3 (○) isolated by preparative capillary electrophoresis of GTTPSPVP *O*-glycosylated products. The peptide fraction was used for 100% determination.

TTSA-(AzP)-TTS, TTSSPTTS, and TTSGPGTT (glycosylation site at position 2) since these octapeptides were only monoglycosylated, probably due to the short length of the peptide backbone on the N-terminal side of the proline.

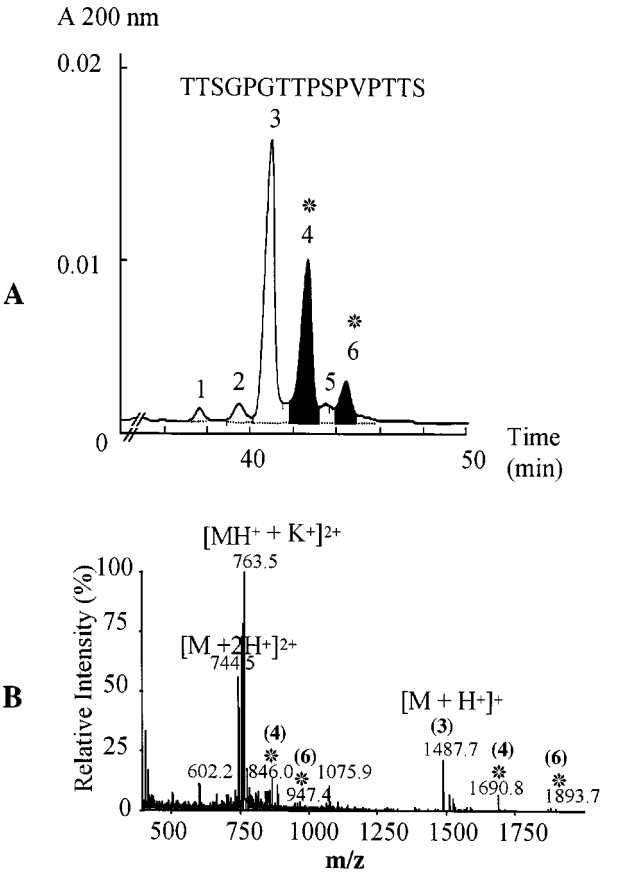

Figure 4. (A) Capillary electrophoresis profile of the reaction products obtained after incubation for 2 h of a human gastric microsomal preparation with the hexadecapeptide TTSGPGTTPSPVPTTS in presence of UDP-GalNAc. The peak 3 corresponds to TTSGPGTTPSPVPTTS. The peaks 1, 2 and 5 correspond to unglycosylated fragments. The *O*-glycosylated fractions (peaks 4 and 6) are marked (*). (B) Electrospray mass spectra of the same products analysed above by capillary electrophoresis. The hexadecapeptide TTSGPGTTPSPVPTTS corresponds to *m/z* 1487.7 (mono charged form) and *m/z* 744.5 and 763.5 (double charged forms). The glycopeptides *m/z* are marked by (*): monoglycosylated peptide (4) and di-glycosylated peptide (6).

Moreover, if the proline greatly influenced the incorporation rate of GalNAc in these octapeptides, it did not modify the first site of glycosylation which was the same for TTSASTTS and TTSAPTTS or TTSGAGTT and TTSGPGTT. Thus, the proline appeared to only influence the GalNAc incorporation rate *in vitro* but not the choice of the first *O*-glycosylation site in MUC5AC threonine rich motif peptides. It is important to note that a prolonged incubation time (up to 6 h) did not modify the results obtained for most of the peptides. The only minor change was the appearance of a very small fraction 3 on the capillary electrophoresis profile of TTSAPTTS and the Edman degradation showed a possible *O*-linkage to the threonine in position 6 (data not shown).

We showed that the GalNAc incorporation into each hexadecapeptide was similar to that of the most glycosylated corresponding octapeptide. Nevertheless, the hexadecapeptide TTSAPTTSTTSAPTTS was glycosylated on positions 10 and 2 when it could have been expected that positions 10 and 9 would be glycosylated (glycosylated sites of the octapeptide TTSTTSAP). This result indicated that the threonine in position −3 relative to a proline was a major site of *N*-acetylgalactosaminylation while position −4 was less favourable. Moreover, it was noteworthy that the total number of *O*-glycosylation sites on an hexadecapeptide was not the sum of the *O*-glycosylation sites described in each constitutive octapeptide (*ie* TTSAPTTS and TTSTTSAP for TTSAPTTSTTSAPTTS). The analysis of the glycosylation sites on the peptide TTSGPGTTPSPVPTTS gave other informations on the gastric GalNAc transferase(s) behaviour towards MUC5AC motif peptides. Indeed, three octapeptide sequences are found in this hexadecapeptide. TTSGPGTT (with a unique glycosylation site in position 2), GTTPSPVP (with two glycosylation sites in positions 2 and 3) and PSPVPTTS (on which the likely monoglycosylated site could not be determined). The motif PSPVP appeared to be of great importance for the recognition of the peptide substrate by GalNAc transferase(s), since the sites of glycosylation were the same for TTSGPGTTPSPVPTTS and GTTPSPVP in regard to this PSPVP motif. Furthermore, the glycosylation was more favoured on positions 8 and 7 of the hexadecapeptide than on the position 2 which was always preferentially glycosylated on the 'threonine rich' octa- and hexadecapeptides.

Discussion

Previous assays using peptides deduced from MUC2 and related glycopeptides as substrates for a purified GalNAc transferase from bovine colostrum suggested that both the primary sequence of the peptide moiety and the first *O*-linked GalNAc peptide affected the recognition of the glycopeptide or peptide as a substrate for the enzyme [18]. As the authors noted, this was in contradiction with the knowledge about clustered *O*-glycans chains attached to the peptide backbone of mucins and these results could be relevant to the use of a purified enzyme as the existence of a family of GalNAc transferases is now currently accepted (for reviews, see [21] and [36]). Consequently, we have chosen to assess the *in vitro O*-glycosylation of MUC5AC motif peptides with crude human gastric microsomal preparations potentially containing all the GalNAc transferases which may be involved in the first step of the reaction. The analysis of enzyme activities on the different MUC5AC motif peptides and of the glycopeptide sequences showed that proline residues increase the recognition of the peptide backbone by UDP-GalNAc : polypeptide *N*-acetylgalactosaminyltransferases. The action of the GalNActransferases appears to be influenced by a substrate with C-terminal structural features on which the enzymes would find good support *in vivo*. In particular, the PSPVP motif has proved to greatly increase the rates of *O*-glycosylation *in vitro*, when it is located in the C-terminal part of the peptide (comparison between GTTPSPVP and PSPVPTTS).

In accordance with Elhammer *et al.* [5], our work also showed that the presence of a proline residue in the +3 position generated an increase of the *in vitro O*-glycosylation reaction (*ie* TTSGAGTT *vs* TTSGPGTT and TTSASTTS *vs* TTSAPTTS). However, when the enzyme has recognized the peptide, the proline is not essential *in vitro* for the choice of the site of *O*-glycosylation by the GalNAc transferase(s), since, *in vitro*, the same position is *O*-glycosylated on these latter peptides (position 2 in all cases). Furthermore, the analysis of the glycopeptide sequences (performed on purified glycopeptides obtained by preparative capillary electrophoresis) also allows us to argue for the presence of several GalNAc transferases in gastric mucosa: *eg* while the glycine of TTSTTSGP is replaced by an alanine behind the same clusters TTS in TTSTTSAP, the GalNActransferase(s) recognize different threonine positions – spaced in TTSTTSGP (positions 2 and 5) and close in TTSTTSAP (positions 4 and 5). The favoured hypothesis would be a multiplicity of enzymes which are required for the *O*-glycosylation of these different peptide sequences [21, 36]. The structural disorganizing effect of the glycine residue could be responsible for the involvement of different GalNAc transferases in recognizing different hydroxy amino acid clusters.

Conformational parameters probably influence the *O*-glycosylation of selected MUC5AC peptides [7] and, in the literature, the glycosylation of longer motifs [16, 37] demonstrated the importance of configurations involving a proline residue. The substrate specificity of the GalNActransferase could therefore be defined as the amino acid sequence which is able to adopt and/or stabilize (*eg trans-prolyl* configuration) an exposed β-turn. Indeed, analysis of the structural properties of the porcine submaxillary gland apomucin revealed the presence of 40% β-turn [38]. The absence of a secondary structure in mature mucins is the result of glycosylation of the peptide core leading to an expanded structure [39]. In our study, these conformational parameters may explain the fact that the GalNAc incorporation rate is strongly decreased when the TTSA-(AzP)-TTS (synthetic peptide with a *cis*-prolyl fixed conformation) is assessed in place of TTSAPTTS (with *trans* preferential conformation). Moreover, the conformational changes induced by the linkage of GalNAc may explain that never more than two glycosylated sites could be detected on the peptides.

No serine residue is found to be glycosylated in this study. The hypothesis may be that, *in vivo*, extensively *O*-glycosylated sequence segments are often devoid of secondary structure feature [39], which *in vitro* can have

a considerable impact on the peptidic acceptor efficiency [20]. Therefore, the higher acceptor efficiency for threonine *vs* serine *in vitro* may be related to the specific orientation of the threonine hydroxyl group which is enhanced by the proline residue, inducing a better configuration for the GalNActransferase approaches. Another hypothesis is that post-translational processes and/or the presence of chaperone molecules should facilitate the *O*-glycosylation of the serine residues *in vivo*.

In conclusion, the influence of the peptidic sequence on human gastric GalNActransferase(s) activity can be attributed to the great importance of proline residue(s) for peptide recognition. Nevertheless, in the absence of proline, the peptide sequence can still be recognized and the site of *O*-glycosylation *in vitro* is the same for peptides with or without a proline residue. Consequently, in addition to proline the peptide sequence may be of importance in the recognition of glycosylation sites in MUC5AC peptides *in vitro*. However, compared to the glycosylation of apomucins *in vivo*, very few hydroxy amino acids are glycosylated in these MUC5AC motif peptides. Thus further studies remain necessary to analyse the mechanisms and rules of the multi-*O*-glycosylation of the apomucins *in vivo*. Moreover, the peptides (especially the multiglycosylated ones) will be good candidates to study the substrate specificity of the different members of the human GalNActransferase family.

References

1 Hounsell EF, Davies MJ, Renouf DV (1996) *Glycoconjugate J* **13**: 19–26
2 Gooley AA, Classon BJ, Marchalek R, Williams KL (1991) *Biochem Biophys Res Commun* **178**: 1194–201.
3 Wilson IBH, Gavel Y, Von Heijne G (1991) *Biochem J* **275**: 529–34.
4 O'Connell BC, Tabak LA, Ramasubbu N (1991) *Biochem Biophys Res Commun* **180**: 1024–30.
5 Elhammer AP, Poorman RA, Brown E, Maggiora LL, Hoogerheide JG, Kozdy F (1993) *J Biol Chem* **268**: 10029–38.
6 Hansen JE, Lund O, Engelbrecht J, Bohr H, Nielsen JO, Hansen JES, Brunak S (1995) *Biochem J* **308**: 801–13.
7 Hansen JE, Lund O, Nielsen JO, Hansen JES, Brunak S (1996) *Nucleic Acids Res* **24**: 248–52.
8 Elhammer A, Kornfeld S (1986) *J Biol Chem* **261**: 5249–55.
9 O'Connell BC, Hagen FK, Tabak LA (1992) *J Biol Chem* **267**: 25010–18.
10 Wang Y, Abernethy JL, Eckhardt AE, Hill RL (1992) *J Biol Chem* **267**: 12709–16.
11. O'Connell BC, Tabak LA (1993) *J Dent Res* **72**: 1554–8.
12. Pisano A, Redmond JW, Williams KL, Gooley AA (1993) *Glycobiology* **3**: 429–35.
13. Wang Y, Agrwal N, Eckhard AE, Stevens RD, Hill RL (1993) *J Biol Chem* **268**: 22979–83.
14 Gooley AA, Williams KL (1994) *Glycobiology* **4**: 413–17.
15. Nishimori I, Perini F, Mountjoy KP, Sanderson SD, Johnson NR, Cerny RL, Gross ML, Fontenot JD, Hollingsworth MA (1994) *Cancer Res* **54**: 3738–44.
16 Nishimori P, Johnson NR, Sanderson SD, Perini F, Mountjoy K, Cerny RL, Gross ML, Hollingsworth MA (1994) *J Biol Chem* **269**: 16123–30.
17 Hagen FK, Van Wuyckhuyse B, Tabak LA (1995) *J Biol Chem* **268**: 18960–5.
18 Brockhausen I, Toki D, Brockhausen J, Peters S, Bielfeld T, Kleen A, Paulsen H, Meldal M, Hagen F, Tabak LA (1996) *Glycoconjugate J* **13**: 849–56.
19 Sørensen T, White T, Wandall HH, Kristensen AK, Roepstorff P, Clausen H (1995) *J Biol Chem* **270**: 24166–73.
20 Nehrke K, Hagen FK, Tabak LA (1996) *J Biol Chem* **271**: 7061–5.
21 Clausen H, Bennett EP (1996) *Glycobiology* **6**: 635–46.
22 Elliott S, Bartley T, Delorme E, Derby P, Hunt R, Lorenzini T, Parker V, Rohdem F, Stoney K (1994) *Biochemistry* **33**: 11237–45.
23 Aubert JP, Biserte G, Loucheux-Lefebvre MH (1976) *Arch Biochem Biophys* **175**: 410–18.
24 Fiat AM, Jolles J, Aubert JP, Loucheux-Lefebvre MH, Jolles P (1980) *Eur J Biochem* **111**: 333–9.
25 Hollosi M, Perclel A, Fasman G (1990) *Biopolymers* **27**: 1549–53.
26 Frommel C, Preissner R (1990) *FEBS Lett* **277**: 159–63.
27 Schachter H, Brockhausen I (1992) *In Glycoconjugates* (Allen HJ, Kisailus EC, eds) pp 263–332. New York: Marcel Dekker.
28 Brandts JF, Halvorson HR, Brennan M (1975) *Biochemistry* **14**: 4953–63.
29 Guyonnet-Duperat V, Audie JP, Debailleul V, Laine A, Buisine MP, Galiegue-Zouitina S, Pigny P, Degand P, Aubert JP, Porchet N (1995) *Biochem J* **305**: 211–19.
30 Bac A, Rivoal K, Thong M, Boussard G, Marraud M, Soudan B, Tetaert D, Degand P (1997) *Lett Pept Sci*, in press.
31 Tetaert D, Soudan B, Lo-Guidice JM, Richet C, Degand P, Boussard G, Mariller C, Spik G (1994) *J Chromatogr* **658**: 31–8.
32 Tetaert D, Briand G, Soudan B, Richet C, Demeyer D, Boersma A, Degand P (1994) *Anal Biochem* **222**: 409–16.
33 Hennebicq S, Soudan B, Tetaert D, Briand G, Richet C, Demeyer D, Degand P (1996) *Carbohydr Res* **280**: 121–6.
34 Biehler R, Schwartz HE (1995) In *Technical Bulletin TIBC* (Beckman Instruments Division, Palo Alto, California) **105**: 1–9.
35 Garin J, Vincon M, Gagnon J, Vignais P (1994) *Biochemistry* **33**: 3772–7.
36 Marth JD (1996) *Glycobiology* **6**: 701–5.
37 Fontenot JD, Tjandra N, Bu D, Ho C, Montelaro RC, Finn OJ (1993) *Cancer Res* **53**: 5386–94.
38 Eckhardt AE, Timpte CS, Abernethy JL, Toumadje A, Curtis Johnson Jr W, Hill RL (1987) *J Biol Chem* **262**: 11339–44.
39 Shogren R, Gerken TA, Jentoft N (1989) *Biochemistry* **28**: 5525–36.

Received 1 May 1997, revised and accepted 11 July 1997