Glycoconjugate Journal (1996) **13:** 849–856

Specificity of O-glycosylation by bovine colostrum UDP-GalNAc: polypeptide α-*N*-acetylgalactosaminyltransferase using synthetic glycopeptide substrates

INKA BROCKHAUSEN[1,2]*, DALE TOKI[1,2], JENNIFER BROCKHAUSEN[1], STEFAN PETERS[3,4], TIM BIELFELDT[3,4], ASTRID KLEEN[3], HANS PAULSEN[3], MORTEN MELDAL[4], FRED HAGEN[5] and LAWRENCE A. TABAK[5]

[1]*Research Institute, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada*
[2]*Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada*
[3]*Institut für Organische Chemie, Universität Hamburg, Hamburg, Germany*
[4]*Department of Chemistry, Carlsberg Laboratory, Copenhagen, Denmark*
[5]*Department of Dental Research and Biochemistry, University of Rochester, Rochester, New York*

Received 18 July 1995, revised 6 December 1995

The factors determining glycosylation of mucin type glycoproteins are not well understood. In the present work, we investigated the role of the peptide moiety and of the presence of O-glycan chains on O-glycosylation by UDP-GalNAc: polypeptide α-*N*-acetylgalactosaminyl-transferase (ppGalNAc-T). We used purified ppGalNAc-T from bovine colostrum and a series of synthetic glycopeptide and peptide substrates most of which contained sequences derived from the tandem repeat region of MUC2 mucin. The rate of incorporation of GalNAc into Thr was significantly greater than toward Ser residues. The presence of one or two GalNAc-Thr moieties in the substrate significantly reduced enzyme activity, and this effect was more pronounced when the disaccharide Galβ1-3GalNAc was present. Thus the sequential attachment of a second GalNAc residue in the vicinity of a pre-existing GalNAc-Thr or Galβ1-3GalNAc-Thr occurs at a slower rate than primary glycosylation of carbohydrate-free peptide. Analysis of products by HPLC showed that the enzyme was selective in glycosylating peptides or glycopeptides with the PTTTPIST sequence in that the preferred primary glycosylation site was the third Thr from the amino-terminal end; secondary glycosylation depended on the site of the primary glycosylation. Negatively but not positively charged amino acids on the carboxy-terminal side of the putative secondary glycosylation site resulted in high activity suggesting charge-charge interactions of substrates with the enzyme. These studies indicate that O-glycosylation by bovine colostrum ppGalNAc-T is a selective process dependent on both the amino acid sequence and prior glycosylation of peptide substrates.

Keywords: polypeptide GalNAc-transferase, substrate specificity, glycopeptides, O-glycosylation, mucin

Abbreviations: Gal, G, D-galactose; GalNAc, *N*-acetyl-D-galactosamine; HPLC, high performance liquid chromatography; ppGalNAc-T, UDP-GalNAc: polypeptide α-GalNAc-transferase EC 2.4.1.41; Ser^{GalNAc}, GalNAcα-Ser; Thr^{GalNAc}, GalNAcα-Thr.

Introduction

UDP-GalNAc: polypeptide α-GalNAc-transferase (ppGalNAc-T) is a ubiquitous enzyme and several species of the enzyme have been purified and cloned [1–8]. The known activities transfer GalNAc primarily to Thr and show much lower activity to Ser although Ser is glycosylated *in vivo*. The various species of ppGalNAc-T may be distinguished by differences in their acceptor substrate specificities.

* To whom correspondence should be addressed.

Mucin tandem repeat regions are thought to be highly O-glycosylated; their exact glycosylation patterns are still unknown. A comparison of O-glycosylation sites of glycoproteins suggests that GalNAc is transferred to selected Thr or Ser residues as the first step in O-glycosylation although no unique peptide sequence is required. The specificity of GalNAc transfer by ppGalNAc-T has been investigated [1, 9, 10] but is still not well understood.

The intracellular localization of ppGalNAc-T (which catalyses the first step in O-glycosylation) has been determined to be in the cis-Golgi [11]. Several reports suggest, however, that the enzyme may be localized more diffusely and extend throughout the Golgi or even to the endoplasmic reticulum [references in 12]. Therefore GalNAc may be added *in vivo* not only to unglycosylated or N-glycosylated peptides but also to glycopeptides that may carry GalNAc or possibly more complex O-glycan chains. The influence of previous glycosylation on peptide substrates on ppGalNAc-T activity is not known and is investigated in the present study. These model studies are aimed at the elucidation of the mechanisms leading to the complex glycosylation patterns of mucins.

We compared the activities of purified bovine colostrum ppGalNAc-T assayed with unglycosylated substrates and substrates containing one or several GalNAc or Galβ1-3GalNAc moieties linked to Ser or Thr residues. Glycopeptides corresponding to the sequences of intestinal mucin MUC2, pig submaxillary mucin tandem repeat and interleukin-2 were chemically synthesized by multiple column glycopeptide synthesis [13, 14] and the purified and completely characterized compounds were used as substrates for ppGalNAc-T. Previous results using unglycosylated substrates were confirmed and extended providing insight into the enzymatic preferences involved in the biosynthesis of highly O-glycosylated peptides. The selectivity of ppGalNAc-T was found to be dependent on previously incorporated sugars as well as the amino acid sequence of substrates.

Materials and methods

MATERIALS

AG1-x8 (100–200 mesh, Cl^- form) was purchased from Bio-Rad. Bovine serum albumin and Triton X-100 were purchased from Sigma. Acetonitrile (190 UV cutoff) was from Fisher Scientific Co. or Caledon Laboratories. UDP-*N*-[1-^{14}C] acetylgalactosamine was purchased from New England Nuclear and diluted with UDP-GalNAc from Sigma.

Glycopeptides and peptides were synthesized by multiple column glycopeptide synthesis by the use of fluoren-9-ylmethoxycarbonyl amino acid pentafluorophenyl ester (Fmoc-AA-OPfp) building blocks and Kieselgur supported polydimethylacrylamide derivatized with an amide linker as the solid support. The glycosylated amino acids were prepared by glycosylation of Fmoc-Ser/Thr-OPfp with per-acetylated glycosyl halides and after purification were used directly for the peptide synthesis [14–18]. The purity and molecular mass of all compounds were determined by HPLC, NMR, FAB-MS and amino acid analysis and these data are published elsewhere [14–17, 19]. All other materials were obtained as previously described [20].

ENZYME

Bovine colostrum ppGalNAc-T was purified as described and had a specific activity of 30.6 μmol min^{-1} mg^{-1} [6]. The concentrated enzyme was stored at −20 °C. For use in the assays, the enzyme was diluted 1:25 or 1:5 with dilution buffer (50 mM Na cacodylate, pH 5.0, 100 mM NaCl, 50% glycerol).

ANALYTICAL METHODS

Polypeptide α-GalNAc-transferase assays

The assay mixture for ppGalNAc-T contained the following components in a total volume of 25 μl: 40 mM Na-cacodylate buffer, pH 6.5, 40 mM 2-mercaptoethanol, 0.1% Triton X-100, 10 mM $MnCl_2$, 0.2 mM UDP-[^{14}C]GalNAc (2250–6070 dpm nmol^{-1}), 5 μl of 1:25 (0.27 μg protein per assay) or 1:5 (1.3 μg protein per assay) diluted bovine colostrum ppGalNAc-T and glycopeptide substrate as indicated in the Tables. Mixtures were incubated at 37 °C for 1 h and the reactions were stopped by 400 μl 20 mM Na-tetraborate per 1 mM EDTA, pH 9. The mixtures were passed through Pasteur pipettes containing 0.5 ml AG 1-x8 resin (Bio-Rad) which were washed with 12.6 ml water. The eluates were collected and counted in 17 ml scintillation fluid. For HPLC analysis, eluates were collected, lyophilized, redissolved in water and injected into the HPLC apparatus. The retention times of standard compounds were used for the identification of the enzyme products. Control assays lacked substrate. Activities were expressed as pmol or nmol of radioactive GalNAc incorporated per assay per hour as presented in the Tables.

To determine kinetic parameters, data were analysed by double reciprocal Lineweaver-Burk transformations at four concentrations of glycopeptide acceptors.

High performance liquid chromatography

HPLC separations were carried out as previously described [20]. Acetonitrile/water mixtures with and without 0.1% trifluoroacetic acid were used as the mobile phase for a reverse phase C18 column as indicated in the Figures. Elutions of compounds were monitored by measuring the absorbance at 195 nm and counting the radioactivity of collected fractions.

Results

General comments on the studies of substrate specificity of polypeptide GalNAc-T

To prevent interference by charge effects, most of the glycopeptide substrates were blocked by an acetyl group at the amino terminal end and by an amide group at the carboxy terminal end. These protecting groups were presumed to resemble natural peptide chains. The substrates were designed to examine the effect of GalNAc and Galβ1-3GalNAc substitutions, the position of the Thr to be glycosylated and the effect of various amino acid substitutions (Tables 1–5). A separate series was synthesized to study the role of amino acids in the −1 (toward the amino terminal) (Table 5) and +1 (toward the carboxy terminal) (Table 4) positions relative to GalNAcα-Thr. Due to the small amounts of substrate available, kinetics were performed only on a limited number of substrates. The formation of product by ppGalNAc-T was proportional to time up to at least 2 h and to enzyme protein concentration up to five times the concentration in the standard assays. All assays, including those for product identification by HPLC, were performed with one hour of incubation time.

Table 1. Specificity of ppGalNAc-T. Effect of the presence of GalNAc or Galβ1-3GalNAc on ppGalNAc-T activity. Substrate sequences are based on the MUC2 sequence PTTTPIST. Most substrate concentrations were about 1 mM; the concentrations of compounds 2, 3 and 1 were 1.4, 0.6 and 1.8 mM respectively. T^{GA} or S^{GA}, GalNAc α-linked to Thr or Ser; T^{GGA}, Galβ1-3GalNAcα-Thr; $T^{GGA\beta}$, Galβ1-3GalNAcβ-Thr. ppGalNAc-T assays were carried out as described in the Methods section using 1:25 diluted enzyme (0.27 μg enzyme per assay). Ac, N-acetyl group at the N-terminus; -NH_2, amide at the carboxyl terminus.

Compound no.	*pmol h^{-1}*	K_m *(nM)*	V_{max} *(nmol h^{-1})*
1 Ac-TPPP	1094	1.3	2.78
2 Ac-PTTTPIST-NH_2	804	1.3	1.20
3 Ac-PTTT^{GA}PIST-NA_2	135		
4 Ac-PTT^{GA}TPIST-NH_2	63	20.5	1.1
5 Ac-PT^{GA}TTPIST-NH_2	25		
6 Ac-PT^{GA}T^{GA}TPIST-NH_2	0		
7 Ac-PTTT^{GGA}PIST-NH_2	18		
8 Ac-PTTT$^{GGA\beta}$PIST-NH_2	4		
9 Ac-PTT^{GGA}TPIST-NH_2	7		
10 Ac-PTT$^{GGA\beta}$TPIST-NH_2	90	12.2	0.78
11 Ac-PT^{GGA}TTPIST-NH_2	0		
12 Ac-PT$^{GGA\beta}$TTPIST-NH_2	3		
13 Ac-PTTS^{GA}PIST-NH_2	75	7.5	0.62
14 Ac-PTS^{GA}TPIST-NH_2	37	12.8	0.75
15 Ac-PS^{GA}TTPIST-NH_2	2		
16 Ac-PTT^{GA}SPIST-NH_2	246	1.3	0.51
17 Ac-PSTT^{GA}PIST-NH_2	29		
18 Ac-PSST^{GA}PIST-NH_2	1		
19 Ac-PTST^{GA}PIST-NH_2	22	2.1	0.11

Table 2. Specificity of ppGalNAc-T. Effect of the presence of GalNAc or Galβ1-3GalNAc and the presence of glycine on ppGalNAc-T activity. Substrate sequences are based on the MUC2 sequence PTTTPIST. Most substrate concentrations were about 1 mM; the concentrations of compounds 2, 11, 20, 25 and 33 were 1.4, 1.8, 5.1, 4.2 and 3.3 mM respectively. T^{GA} or S^{GA}, GalNAc α-linked to Thr or Ser; T^{GGA}, Galβ1-3GalNAcα-Thr; $T^{GGA\beta}$, Galβ1-3GalNAcβ-Thr. ppGalNAc-T assays were carried out as described in the Methods section using 1:25 diluted enzyme (0.27 μg enzyme per assay). Ac, *N*-acetyl group at the N-terminus; -NH_2, amide at the carboxyl terminus.

Compound no.	*pmol h^{-1}*	K_m *(mM)*	V_{max} *(nmol h^{-1})*
2 Ac-PTTTPIST-NH_2	804	1.3	1.20
3 Ac-PTTT^{GA}PIST-NH_2	135		
20 Ac-GTTT^{GA}PIST-NH_2	20		
21 Ac-GTTT^{GA}GIST-NH_2	3		
7 Ac-PTTT^{GGA}PIST-NH_2	18		
22 Ac-GTTT^{GGA}PIST-NH_2	5		
8 Ac-PTTT$^{GAA\beta}$PIST-NH_2	4		
23 Ac-GTTT$^{GGA\beta}$PIST-NH_2	0		
4 Ac-PTT^{GA}TPIST-NH_2	63	20.5	1.1
24 Ac-PTT^{GA}TGIST-NH_2	3		
25 Ac-GTT^{GA}GIST-NH_2	15		
26 Ac-PGT^{GA}TPIST-NH_2	0		
27 Ac-GTT^{GA}TPIST-NH_2	27	6.3	0.18
28 Ac-PTT^{GA}GPIST-NH_2	26	0.9	0.07
9 Ac-PTT^{GGA}TPIST-NH_2	7		
29 Ac-PTT^{GGA}TGIST-NH_2	0		
30 Ac-GTT^{GGA}TPIST-NH_2	3		
10 Ac-PTT$^{GGA\beta}$TPIST-NH_2	90	12.2	0.78
31 Ac-GTT$^{GGA\beta}$TPIST-NH_2	12		
5 Ac-PT^{GA}TTPIST-NH_2	25		
32 Ac-GT^{GA}TTPIST-NH_2	0		
33 Ac-PT^{GA}TTGIST-NH_2	7		
34 Ac-PT^{GA}TGPIST-NH_2	0		
35 Ac-PT^{GA}GTPIST-NH_2	9		
36 Ac-PT^{GA}PTPIST-NH_2	41	2.9	0.18
11 Ac-PT^{GGA}TTPIST-NH_2	0		
37 Ac-GT^{GGA}TTPIST-NH_2	0		
12 Ac-PT$^{GGA\beta}$TTPIST-NH_2	3		
38 Ac-GT$^{GGA\beta}$TTPIST-NH_2	2		

Which Thr residues are glycosylated when PTTTPIST is used as substrate?

Many analogues of the PTTTPIST sequence (derived from the MUC2 repeat region) containing PIST as the last four amino acids, were completely or almost completely inactive as substrates (Tables 1–3) suggesting that the Ser and Thr residues of the PIST sequence are not glycosylation sites. We have therefore concentrated on

Table 3. Specificity of ppGalNAc-T. Sequence analogues from IL-2, porcine submaxillary mucin and MUC2 were used as substrates. All substrate concentrations were 1 mM. T^{GA} or S^{GA}, GalNAc α-linked to Thr or Ser. ppGalNAc-T assays were carried out as described in the Methods section using either 1:25 diluted enzyme (0.27 μg enzyme per assay; section A) or 1:5 diluted enzyme (1.3 μg enzyme per assay, section B). Ac, *N*-acetyl group at the N-terminus; -NH_2, amide at the carboxyl terminus.

Compound no.	*pmol h^{-1}*
Section A:	
IL-2 sequence analogues:	
39 APT^{GA}SS	0
40 APT^{GA}SSS	7
41 APT^{GA}SSA	4
42 APTS^{GA}SS	0
43 APT^{GA}SSSTK	0
44 APT^{GA}SSSTKKT	15
Porcine submaxillary mucin sequence analogues:	
45 Ac-GSSS^{GA}GSPG-NH_2	9
46 Ac-GSS^{GA}SGSPG-NH_2	6
Section B:	
MUC2 sequence analogues:	
47 Ac-PSSSPIST-NH_2	0
48 Ac-PSSS^{GA}PIST-NH_2	0
49 Ac-PSS^{GA}SPIST-NH_2	0
50 Ac-PS^{GA}SSPIST-NH_2	0
51 Ac-PSS^{GA}S^{GA}PIST-NH_2	0
52 Ac-PS^{GA}S^{GA}SPIST-NH_2	0
53 Ac-PS^{GA}SS^{GA}PIST-NH_2	0
54 Ac-PS^{GA}S^{GA}S^{GA}PIST-NH_2	0
55 Ac-PTSSPIST-NH_2	2473
56 Ac-PTS^{GA}SPIST-NH_2	2984
57 Ac-PTS^{GA}S^{GA}PIST-NH_2	221

Table 4. Effect of amino acid substitutions at the +1 (Pro) position of the substrate on ppGalNAc-T activity. Substrate concentrations were about 1 mM. T^{GA}, GalNAc α-linked to Thr. ppGalNAc-T assays were carried out as described in the Methods section using 1:25 diluted enzyme (0.27 μg enzyme per assay).

Compound no.	*pmol h^{-1}*	K_m *(mM)*	V_{max} *(nmol h^{-1})*
3 Ac-PTTT^{GA}PIST-NH_2	135		
58 Ac-PTTT^{GA}GIST-NH_2	39		
59 Ac-PTTT^{GA}SIST-NH_2	45	7.2	0.38
60 Ac-PTTT^{GA}AIST-NH_2	69		
61 Ac-PTTT^{GA}RIST-NH_2	14		
62 Ac-PTTT^{GA}DIST-NH_2	7		
63 Ac-PTTT^{GA}HIST-NH_2	38		
64 Ac-PTTT^{GA}IIST-NH_2	45		
65 Ac-PTTT^{GA}FIST-NH_2	83		
66 Ac-PTTT^{GA}WIST-NH_2	48	37	1.89
67 Ac-PTTT^{GA}YIST-NH_2	50	0.8	0.10

Table 5. Effect of amino acid substitutions at the –1 (Thr) position of the substrate on ppGalNAc-T activity. Substrate concentrations were about 1 mM. GalNAc α-linked to Thr is written as T^{GA}. ppGalNAc-T assays were carried out as described in the Methods section using 1:25 diluted enzyme (0.27 μg enzyme per assay).

Compound no.	*pmol h^{-1}*	K_m *(mM)*	V_{max} *(nmol h^{-1})*
3 Ac-PTTT^{GA}PIST-NH_2	135		
68 Ac-PTPT^{GA}PIST-NH_2	399		
69 Ac-PTST^{GA}PIST-NH_2	22	2.1	0.11
70 Ac-PTAT^{GA}PIST-NH_2	34	9.3	0.46
71 Ac-PTRT^{GA}PIST-NH_2	0		
72 Ac-PTNT^{GA}PIST-NH_2	10	2.0	0.02
73 Ac-PTDT^{GA}PIST-NH_2	100	8.1	0.69
74 Ac-PTQT^{GA}PIST-NH_2	455	0.4	0.51
75 Ac-PTET^{GA}PIST-NH_2	401	3.3	2.03
76 Ac-PTHT^{GA}PIST-NH_2	27		
77 Ac-PTIT^{GA}PIST-NH_2	23		
78 Ac-PTLT^{GA}PIST-NH_2	37		
79 Ac-PTKT^{GA}PIST-NH_2	8	3.9	0.02
80 Ac-PTMT^{GA}PIST-NH_2	49	12.2	0.60
81 Ac-PTFT^{GA}PIST-NH_2	5		
82 Ac-PTWT^{GA}PIST-NH_2	12		
83 Ac-PTYT^{GA}PIST-NH_2	1		
84 Ac-PTVT^{GA}PIST-NH_2	26		

exchanging and variably glycosylating the first four residues of the PTTTPIST sequence.

The smallest active peptide, Ac-TPPP derived from the bovine myelin basic protein sequence (compound **1**), was among the best substrates with an apparent K_m of 1.3 mM and V_{max} of 2.78 nmol h^{-1} per 0.27 μg enzyme (Table 1). This corresponds to 0.17 μmol min^{-1} per mg protein. The octapeptide **2** (Ac-PTTTPIST-NH_2), derived from the MUC2 repeat sequence, was slightly less active with an apparent K_m of 1.3 mM and V_{max} of 1.2 nmol h^{-1} per 0.27 μg enzyme (Table 1).

Activity was observed with glycopeptides **3**, **4** and **5** (Table 1) which carry GalNAc at positions 4, 3 and 2, respectively, of the PTTTPIST sequence. Thus none of these Thr residues are exclusive glycosylation sites. Several PTTTPIST analogues (**68**, **73–75**) were very active (Table 5) although only Thr at position 2 was available for GalNAc transfer indicating that this Thr residue is an acceptor site. In addition, the replacement of Ser by Thr at position 2 (e.g. compare compounds **47** and **55**, Table 3) dramatically increased the activity, again indicating that Thr at position 2 is a glycosylation site. The fact that compound **36** was active (Table 2), suggests that Thr at position 4 is also an acceptor site. Compound **34** (Table 2) containing unsubstituted Thr at position 3 was inactive whereas compound **17** (Table 1) also containing unsubstituted Thr at position 3 showed low activity indicating that this Thr residue may be an acceptor site under certain conditions.

To find the acceptor site in unglycosylated PTTTPIST (compound **2**), the product was analysed by HPLC (Fig. 1). All standards representing all seven possible ppGalNAc-T products were well separated. Only one major radioactive peak eluted after incubation of **2** with ppGalNAc-T and was identified, based on its retention time, as PTTT^{GalNAc}PIST (Fig. 1). Although the product of this reaction (compound **3**) is a substrate (Table 1, Fig. 2) little or no secondary product formation was observed. These results suggest that Thr at position 4 in the PTTTPIST sequence was selectively glycosylated although other potential substrate sites were present.

Thr at position 4 is also a glycosylation site when Thr at position 2 is already glycosylated. The product when compound **5**, PT^{GalNAc}TTPIST (Table 1), is used as a substrate gave a small peak on HPLC at the position of the Ac-PT^{GalNAc}TT^{GalNAc}PIST-NH_2 standard indicating

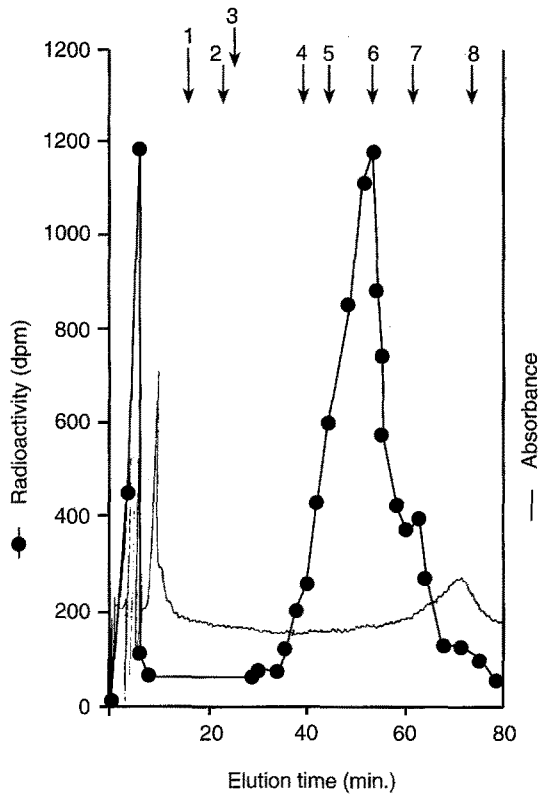

Figure 1. HPLC pattern of ppGalNAc-T enzyme product using Ac-PTTTPIST-NH_2 substrate. Substrate was incubated for 1 h with a 1:5 dilution of ppGalNAc-T as described in the Methods section. Product was then analyzed by HPLC using a C18 column. Compounds were eluted with 8% acetonitrile/92% water in 0.1% trifluoroacetic acid at a flow rate of 1 ml min^{-1}. Standards and their elution times (minutes) were 1, Ac-PT^{GalNAc}T^{GalNAc}T^{GalNAc}PIST-NH_2 (14 min); 2, Ac-PTT^{GalNAc}T^{GalNAc}PIST-NH_2 (25 min); 3, Ac-PT^{GalNAc}T^{GalNAc}TPIST-NH_2 (26 min); 4, Ac-PT^{GalNAc}TT^{GalNAc}PIST-NH_2 (39 min); 5, Ac-PT^{GalNAc}TTPIST-NH_2 (46 min); 6, Ac-PTTT^{GalNAc}PIST-NH_2 (54 min); 7, Ac-PTT^{GalNAc}TPIST-NH_2 (62 min); 8, Ac-PTTTPIST-NH_2 (75 min). Enzyme product eluted as a broad peak with the elution time of standard 6.

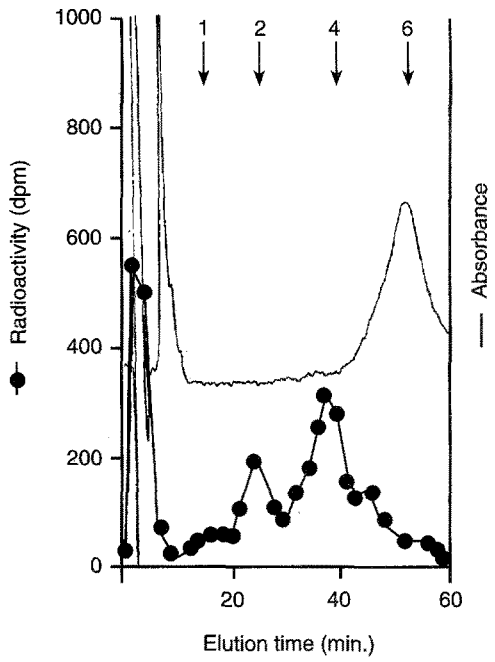

Figure 2. HPLC pattern of ppGalNac-T enzyme product using Ac-PTTT^{GalNAc}PIST-NH_2 substrate. Substrate was incubated and analysed by HPLC with the use of standards as described in Fig. 1. Enzyme product eluted as two peaks with the elution time of standards 2, Ac-PTT^{GalNAc}T^{GalNAc}PIST-NH_2 (25% of radioactivity) and standard 4 Ac-PT^{GalNAc}TT^{GalNAc}PIST-NH_2 (75% of radioactivity).

that Thr at position 4 was glycosylated (data not shown). Glycopeptides lacking an unsubstituted Thr at position 4 may be very active (compounds **3**, **13**, Table 1; compounds **68**, **73–75**, Table 5) indicating that there are also other glycosylation sites.

Product analysis using glycopeptide **3**, PTTT^{GalNAc}PIST (Table 1), as a substrate (Fig. 2) showed that GalNAc was mainly added to Thr at position 2 (75%) and a minor proportion to Thr at position 3 (25%). A small peak was observed with the elution time of the triglycosylated standard. Compound **4**, Ac-PTT^{GalNAc}TPIST-NH_2, was an active substrate (Table 1). However, a good HPLC separation of the possible di-glycosylated products Ac-PTT^{GalNAc}T^{GalNAc}PIST-NH_2 and Ac-PT^{GalNAc}T^{GalNAc}T-PIST-NH_2 could not be achieved.

These results indicate that the choice of Thr to be glycosylated varies with the position of Thr in the amino acid sequence. The presence and position of GalNAc substitution can also affect the site of glycosylation and can induce glycosylation of acceptor sites that are not used for primary glycosylation.

Influence of previous glycosylation

Glycopeptides with the identical PTTTPIST peptide sequence but with one or more GalNAc residues at various positions were compared (Table 1). The presence

of GalNAc reduced the transfer of a second GalNAc significantly (compare compounds **3**, **4** and **5** with **2**). This was less pronounced when GalNAc was attached to Thr at position 4 (compound **3**). When two GalNAc residues were present in the substrate (compound **6**) no activity was observed even though the preferred substrate site at position 4 was available. A slight reduction in activity was found for substrates that contained GalNAc-Ser glycosylation sites (compounds **13–15**, Table 1) instead of GalNAc-Thr.

When the larger disaccharides Galβ1-3GalNAcα or Galβ1-3GalNAcβ were linked to Thr (compounds **7–12**, Table 1) the activity was reduced even further compared to GalNAc alone (compounds **3–5**) with the surprising exception of compound **10** which contained Galβ1-3GalNAcβ attached to Thr in position 3 and which was more active than compound **4** containing GalNAc in the physiological α-linkage at that position.

Effect of glycine in the substrate on enzyme activity

When Pro at position 1 in the PTTT^{GalNAc}PIST substrate sequence (compound **3**) was changed to Gly (compound **20**, Table 2), ppGalNAc-T activity was significantly reduced. A similar reduction by Gly was seen (Table 2) when GalNAc was attached to Thr at position 2 (compounds **5** and **32**) or position 3 (compounds **4** and **27**) or when Galβ1-3GalNAcβ- was attached to Thr at position 3 (compounds **10** and **31**). It is interesting to note that compound **32** is not using its available glycosylation sites due to the effect of Gly. A comparison of compounds **4** and **26** (Table 2) illustrates the unfavourable effect of Gly in the -1 position relative to Thr^{GalNAc}. Relatively inactive substrates resulted when Pro at position 5 was replaced with Gly (compare **5** with **33**, and **4** with **24**) or when both Pro residues were replaced with Gly (compounds **21** and **25**). Gly at other positions in the peptide (compounds **28**, **34**, and **35**) is also unfavourable. The negative effect of Gly can also be observed when comparing compounds **7** and **22** (Table 2). Since the control compounds **8**, **9**, **11** and **12** are inactive as substrates, there was no Gly effect with this series of compounds (**23**, **29**, **30**, **37** and **38**).

Ser as a glycosylation site

To investigate whether GalNAc may be added to Ser, we studied derivatives of the pig submaxillary mucin tandem repeat sequence GSSSGSPG [21], the amino-terminal sequence glycopeptide APT^{GalNAc}SSSTKKT of interleukin-2 [22] and PSSSPIST, which is a modified MUC2 tandem repeat sequence (Table 3). None of the compounds (**39–44**) of the interleukin-2 series were good acceptors; low activity was observed with compound **44**. The compounds of the pig submaxillary mucin series also showed very low activity (compounds **45** and **46**). Likewise, the Ser-rich MUC2 series compounds (**47–54**) were inactive although they all contained the PIST sequence (Table 3); PTTTPIST (compound **2**) was very active (Table 1). It is interesting that the substitution of Thr for Ser in the second position (compare compounds **55** with **47**, **56** with **49**, and **57** with **51**) led to a greatly increased activity (Table 3). Also, comparing active compounds **17** and **19** with inactive compound **18** (Table 1) indicates that Ser is not a favourable acceptor site.

Effect of exchanging specific amino acids at the -1 and $+1$ positions relative to Thr^{GalNAc}

Compound **3**, PTTT^{GalNAc}PIST, was the product of primary glycosylation by ppGalNAc-T (Fig. 1) and a substrate for further GalNAc transfer (Fig. 2). To analyse the role of the peptide sequence in secondary glycosylation, we exchanged the amino acids at the $+1$ and -1 positions relative to Thr^{GalNAc}. We compared the effect of replacing Pro at positon 5 ($+1$ position) of PTTT^{GalNAc}PIST with various amino acids (compounds **58–67**, Table 4). Charged amino acids (Arg in compound **61** and Asp in compound **62**) significantly reduced ppGalNAc-T activity. All other substitutions resulted in smaller losses of activity.

More drastic changes in activity were observed (Table 5) when Thr at position 3 of the PTTT^{GalNAc}PIST sequence was replaced by other amino acids (-1 relative to the position of the first glycosylation). A large decrease in activity occurred when this -1 position was substituted with positively charged amino acids (Arg, Lys, compounds **71**, **79**), or smaller polar amino acids (Asn, compound **72**; Ser, compound **69**). Relatively high activities were seen with Asp, Gln and Glu (compounds **73–75**). Thus a positive charge in the -1 position interfered with ppGalNAc-T activity but a negative charge was more favourable suggesting a charge-charge interaction between the enzyme and the substrate in this subsite. These effects may be mediated by changes in secondary structure although short peptides are not expected to have any significant secondary structure in solution. These conclusions have to be supported by additional data. Bulky hydrophobic amino acids (Phe, Trp, Tyr, compounds **81–83**) in the -1 position almost abolished activity; however, Pro (compound **68**) was favourable. Compound **69** containing Ser was significantly less active than compound **3** (containing Thr), possibly due to the higher flexibility of Ser-containing peptides. However, Ser replacing Thr at position 4 (compare compounds **4** and **16**, Table 1) resulted in a marked increase in activity.

Discussion

The present results suggest that both the primary sequence of the peptide moiety and peptide glycosylation of the substrate affect the recognition of the glycopeptide or

peptide as a substrate for bovine colostrum ppGalNAc-T. The presence of glycosylation in the substrate inhibited ppGalNAc-T activity in a manner directly relating to the number and size of the O-glycan chains. Secondary glycosylation was more difficult. Carbohydrate chains may sterically hinder the recognition of substrate by ppGalNAc-T. Alternatively, the more rigid conformation of the peptide induced by the carbohydrate [23] may not be favourable for substrate binding. In contrast to our observations, highly O-glycosylated proteins such as mucins or glycophorins often contain clustered O-glycan chains attached to adjacent amino acids suggesting more efficient *in vivo* synthesis or the possible presence of another enzyme with a specificity distinct from that of the bovine colostrum enzyme. The specific glycosylation sites of heavily glycosylated mucin tandem repeat regions and the O-glycan structures attached to these sites are as yet unknown. Based on our studies, it can be predicted that spacing between sites would be preferred as opposed to clusters of sugars at adjacent sites. Since secondary glycosylation at an adjacent site appears to be more difficult, the biosynthesis of clustered glycosylation sites may be accomplished by the simultaneous addition of several GalNAc residues rather than by sequential events. No conserved consensus amino acid sequences for O-glycan glycosylation have been identified in mucins.

We found that Thr was consistently preferred over Ser as the site of GalNAc attachment. This is in agreement with our previous findings and those of several laboratories that UDP-GalNAc: polypeptide α-GalNAc transferase from various tissues preferentially glycosylates Thr residues rather than Ser residues [6, 9, 10, 24]. This difference in reactivity reflects the inherent difference in conformational properties and hydrophobic character of Ser and Thr. The interpretations are of course limited by the design of these peptides and the *in vivo* glycosylation of intact mucin molecules may differ.

The specific O-glycosylation sites in a number of non-mucin glycoproteins have been determined and Pro residues [25] are especially abundant at the -1 and $+3$ positions relative to the glycosylation site. There appears to be a high selectivity of ppGalNAc-T for substrates containing Pro. This is consistent with our findings that PTXXP sequences form excellent substrates. The role of Pro may be mediated by its ability to force a rigid and favourable conformation of the peptide that promotes the steric exposure of the hydroxyl of Thr. In contrast, Gly consistently reduced the activity, possibly by allowing greater and unfavourable conformational flexibility.

Pig submaxillary gland ppGalNAc-T also transferred GalNAc to two Thr residues of Pro rich sequences of bovine myelin basic protein [26]. The preferred site was, however, not Thr in the TXXP sequence but the first Thr in the TPRTPPP sequence [27]. Wang *et al.* [1] did not find a consensus sequence for the pig submaxillary ppGalNAc-T. Thr but not Ser residues were glycosylated at various levels. A preferred glycosylation site was the third Thr in the VKTEATTFI sequence. In another study [9], the pig submaxillary gland ppGalNAc-T glycosylated Ser residues, but with much lower efficiency than Thr. The pig submaxillary gland mucin sequence, VLGTTAV, formed reasonably good substrates which is consistent with our finding that the second Pro in PTXXP sequences can be substituted by other hydrophobic amino acids.

ppGalNAc-T extracted from pig trachea [24] was also very active with the TPPP substrate but not SPPP, and the enzyme glycosylated both Thr residues in the myelin basic protein VTPRTPPP sequence. BHK cells were found to lack an enzyme that could glycosylate SPPP but these cells were active with TPPP substrate [28]. The lack of Ser glycosylation in this study is consistent with previous findings that bovine colostrum ppGalNAc-T did not glycosylate the erythropoietin sequence PPDAASAAPLR but could act when Ser was replaced by Thr [29].

We observed a strong influence on enzyme activity of substitutions at the -1 and $+1$ positions relative to the primary glycosylation site of MUC2 peptide PTTTPIST. The data suggests that the enzyme interacts with negatively charged residues and this interaction promotes catalysis. Nishimori [30] also reported that TRPP or TRAP and TRP sequences did not form substrates for breast cancer cell ppGalNAc-T. Previous studies on glycopeptide substrates have indicated that charges on the amino termini, and especially on the carboxy termini of short peptides, interfered with the activity of core 1 β3-Gal-T [22]. The present results suggest that charge-charge interactions are also important determinants for ppGalNAc-T activity.

By a comparison of known glycosylation sites, O'Connell *et al.* [31] suggested that positions -6, -1 and $+3$ were important for ppGalNAc-T activity. Using peptides with von Willebrand factor sequences it was found that the -3 and -2 positions were also important and charged residues (Glu and Arg) at -1 markedly decreased ppGalNAc-T activity [29]. The enzymes from pig and bovine submaxillary glands were found to glycosylate peptides with the TPPP sequence when the amino acid in the -1 position was Arg and RTPPP was a good substrate [32, 33].

The present results call for thorough investigations of the three dimensional factors that cause a peptide or glycopeptide to be a good substrate for ppGalNAc-T. The presentation of Thr or Ser may depend on distant factors in the protein and may be very difficult to predict. Small peptides are usually much more flexible and lack the conformational influence of distant structures. Differences in reactivity observed for Thr at positions 2 and 4 in the substrate sequence PTTTPIST may be due to the position

of the former in the vicinity of the N-terminus. The effect of Pro may be more important in small peptides compared to glycoproteins. Glycopeptides such as those used in this study show more conformational rigidity than carbohydrate-free peptides due to the relatively large sugar residues that can form hydrogen bonds. The ppGalNAc-T may bind to the hydroxyl of Thr in a specific three dimensional configuration requiring the methyl group which is lacking in the much less active Ser-containing substrates. There may not be a specific interaction between the enzyme and the peptide in the position adjacent to the glycosylation site and consequently a number of substrate sequences are tolerated.

We previously showed that the processing of O-glycans to core 1, Galβ1-3GalNAc, and other core structures depends on peptide structure and glycosylation [22, 34]. Thus, not only the addition of GalNAc to peptide, but also the synthesis of various types of mucin oligosaccharide core structures may be controlled by the peptide structure and glycosylation of substrates.

Acknowledgements

The authors thank Jimmy Yang for carrying out HPLC analyses. This research was supported by a grant from the Medical Research Council of Canada to I. Brockhausen, by a European Community Science Project grant SCI-CT92-0765 to M. Meldal and H. Paulsen, and by NIH grant DE 08108 to L. Tabak.

References

1. Wang Y, Abernethy JL, Eckhardt AE, Hill RL (1992) *J Biol Chem* **267**: 12709–16.
2. Sugiura M, Kawasaki T, Yamashina I (1982) *J Biol Chem* **257**: 9501–7.
3. Takeuchi M, Yoshikawa M, Sasaki R, Chiba H (1985) *Agric Biol Chem* **49**: 1059–69.
4. Elhammer A, Kornfeld S (1986) *J Biol Chem* **261**: 5249–55.
5. Homa FL, Hollander T, Lehman DJ, Thomsen DR, Elhammer AP (1993) *J Biol Chem* **268**: 12609–16.
6. Hagen FK, van Wuyckhuyse B, Tabak LA (1993) *J Biol Chem* **268**: 18960–65.
7. Sørensen T, White T, Wandal HH, Kristensen AK, Roepstorff P, Clausen H (1995) *J Biol Chem* **270**: 24166–73.
8. White T, Bennett EP, Takio K, Sørensen T, Bonding N, Clausen H (1995) *J Biol Chem* **270**: 24156–65.
9. Wang Y, Agrawal N, Eckhardt AE, Stevens RD, Hill RL (1993) *J Biol Chem* **268**: 22979–83.
10. Elhammer AP, Poorman RA, Brown E, Maggiora LL, Hoogerheide JG, Kézdy FJJ (1993) *J Biol Chem* **268**, 10029–38.
11. Roth J (1987) *Biochim Biophys Acta* **906**: 405–36.
12. Brockhausen I (1995) In *Glycoproteins, New Comprehensive Biochemistry, Vol 29a* (J Montreuil, JFG Vliegenthart, H Schachter eds) pp. 201–59, Amsterdam: Elsevier Pub.
13. Meldal M, Holm CB, Bojesen G, Jacobsen MH, Holm A (1993) *Int J Peptide Protein Res* **41**: 250–60.
14. Peters S, Bielfeldt T, Meldal M, Bock K, Paulsen H (1992) *J Chem Soc Perkin Trans* **1**: 1163–71.
15. Peters S, Bielfeldt T, Meldal M, Bock K, Paulsen H (1991) *Tetrahedron Lett* **32**: 5067–70.
16. Paulsen H, Bielfeldt T, Peters S, Meldal M, Bock K (1994) *Liebigs Ann Chemie* 369–87.
17. Paulsen H, Peters S, Bielfeldt T, Meldal M, Bock K (1995) *Carboh Res* **268**: 17–34.
18. Kleen A, Paulsen H (1995) unpublished.
19. Paulsen H, Bielfeldt T, Peters S, Meldal M, Bock K (1994) *Liebigs Ann Chemie* 381–87.
20. Brockhausen I, Romero P, Herscovics A (1991) *Cancer Res* **51**: 3136–42.
21. Timpte CS, Eckhardt, AE, Abernethy JL, Hill RL (1988) *J Biol Chem* **263**: 1081–88.
22. Brockhausen I, Möller G, Merz G, Adermann K and Paulsen H (1990) Biochemistry **29**: 10206–12.
23. Bock K, Stuike-Prill R, Meldal M (1994) In *Complex carbohydrates in drug research, Structural and functional aspects* (Bock K, Clausen H, eds) pp. 229–42, Munksgaard: Copenhagen.
24. Cottrell JM, Hall RL, Sturton RG, Kent PW (1992) *Biochem J* **283**: 299–305.
25. Wilson IBH, Gavel Y, von Heijne G (1991) *Biochem J* **275**: 529–34.
26. Cruz TF, Moscarello MA (1983) *Biochim Biophys Acta* **760**: 403–10.
27. Persaud R, Fraser P, Wood DD, Moscarello MA (1988) *Biochim Biophys Acta* **966**: 357–61.
28. Hughes RC, Bradbury AF, Smyth DG (1988) *Carbohyd Res* **178**: 259–69.
29. O'Connell BC, Hagen FK, Tabak LA (1992) *J Biol Chem* **267**: 25010–18.
30. Nishimori I, Perini F, Mountjoy KP, Sanderson SD, Johnson N, Cerny RL, Gross ML, Fontenot JD, Hollingsworth MA (1994) *Cancer Res* **54**: 3738–44.
31. O'Connell B, Tabak LA, Ramasubbu N (1991) *Biochem Biophys Res Comm* **180**: 1024–30.
32. Young JD, Tsuchiya D, Sandlin DE, Holroyde MJ (1979) *Biochemistry* **18**: 4444–48.
33. Briand JP, Andrews SP, Cahill E, Conway NA, Young JD (1981) *J Biol Chem* **256**: 12205–7.
34. Granovsky M, Bielfeldt T, Peters S, Paulsen H, Meldal M, Brockhausen J, Brockhausen I (1994) *Eur J Biochem* **221**: 1039–46.