Glycobiology vol. 7 no. 8 pp. 1053–1060, 1997

Charge distribution of flanking amino acids inhibits O-glycosylation of several single-site acceptors *in vivo*

Keith Nehrke, Kelly G.Ten Hagen, Fred K.Hagen and Lawrence A.Tabak[1]

Departments of Dental Research and Biochemistry, School of Medicine and Dentistry, University of Rochester, 601 Elmwood Avenue, Box 611, Rochester, NY 14642, USA

[1]To whom correspondence should be addressed

From surveys of known O-glycosylation sites and *in vitro* glycosylation assays with synthetic peptide acceptors, it appears that the presence of charged amino acids near serine/threonine residues reduces the likelihood of O-glycosylation by UDP-GalNAc polypeptide:N-acetylgalactosaminyltransferases (ppGaNTases). Previously, we demonstrated that the *in vivo* O-glycosylation of a sequence derived from a known glycosylation site of human von Willebrand factor (PHMAQVTVGPGL) was markedly reduced when charged residues were substituted at position −1 and +3 relative to the single threonine. In contrast, acidic residues at positions −2, +1, and +2 had no effect (Nehrke *et al.*, 1996), suggesting that charge distribution but not charge density was important. To determine whether the charge distribution effect on O-glycosylation is limited to a specific sequence context or restricted to unique isoforms of ppGaNTase, we have analyzed the *in vivo* O-glycosylation of six secreted recombinant reporter proteins in three different cell backgrounds. The differential presence of known ppGaNTase transcripts was determined in each cell type by Northern blot analysis. Each reporter, which contains a single site of O-glycosylation, was O-glycosylated in a cell-background-specific manner; digestion with O-glycanase and α-N-acetylgalactosaminidase following mild acid hydrolysis suggested that simple type II core structures were acquired. However, in COS7 cells, one reporter peptide acquired glycosaminoglycans in preference to mucin-type O-glycans. Regardless of cell background or the reporter examined, the substitution of glutamic acid residues at positions −1 and +3 markedly diminished the level of mucin-type O-glycosylation. Charge distribution would appear, therefore, to play a more general role in determining the extent to which solitary O-glycosylation sites are modified.

Key words: glycosyltransferases/mucins/O-glycosylation

Introduction

O-glycosylation imparts unique physicochemical properties to molecules such as mucin-glycoproteins (Jentoft, 1990) which play an important role in the protection of mucosal surfaces (Tabak, 1995). The first step in O-glycosylation is catalyzed by a family of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases (ppGaNTases); N-acetylgalactosamine (GalNAc) is transferred from the sugar donor UDP-GalNAc to Ser or Thr of the peptide acceptor. Thus, the addition of the core GalNAc represents both the initial regulatory step in mucin glycosylation as well as a key structural determinant of mucin function.

Since not all potential acceptor sites become O-glycosylated, signals must exist to specify which hydroxyamino acids acquire O-glycans. The relative contribution of conformational accessibility versus the role played by the nature of the amino acids flanking a potential acceptor site in determining the initiation of O-glycosylation has been debated at length. Aubert and Loucheux-Lefebvre (1976) first suggested that the hydroxyamino acids involved in O-glycosidic linkages belonged to a β-turn as predicted by the algorithm of Chou and Fasman (1978). While conceptually appealing, few studies have experimentally demonstrated that conformation influences O-glycosylation. Elliot *et al.* (1994) determined that disruption of the Cys 7-Cys 161 disulfide bond which alters proper folding of erythropoietin (as ascertained by reactivity with conformational dependent monoclonal antibody) reduces O-glycosylation of Ser 126 to 5–10% of wild-type levels. However, such a relationship does not preclude an effect by the nature of the amino acids which flank potential O-linked sites as well. From surveys of known O-glycosylation sites, it is apparent that charged residues are rarely found at positions which flank glycosylated Ser or Thr (O'Connell *et al.*, 1991; Wilson *et al.*, 1991; Carlsson *et al.*, 1993; Elhammer *et al.*, 1993; Chou *et al.*, 1995; Hansen *et al.*, 1995; Strub *et al.*, 1997). The presence of charged residues at positions −1 and +3 relative to an O-glycosylation site inhibits *in vitro* incorporation of GalNAc into synthetic peptide substrates (O'Connell *et al.*, 1991, 1992; Nishimori *et al.*, 1994; Stadie *et al.*, 1995), although a peptide which is derived from the V3 loop of the HIV gp120 (CIRIQRGPGRAFVTIGKJGNM) contains a lysine residue at position +3, and is glycosylated by recombinant ppGaNTase-T3, *in vitro* (Bennett *et al.*, 1996).

Previously, we analyzed the *in vivo* glycosylation of a known O-glycosylation site (PHMAQVTVGPGL) derived from human von Willebrand factor in COS7 and MCF-7 cells. Substitution of Glu at both the −1 and +3 positions decreased the *in vivo* O-glycosylation of the site to about 10–40% (occupancy) of wild-type levels. However, overall charge density did not compromise glycosylation; reporter with Glu substitutions at positions −2, +1, and +2 was glycosylated to nearly wild-type levels. Collectively, these observations suggest that the specific charge distribution within the primary sequence of a protein can influence O-glycosylation. To further test this hypothesis we have studied the glycosylation of additional sequeons derived from known O-glycosylation sites of five different proteins. Since there are multiple ppGaNTase isoforms and varied repertoires are expressed in different cell types (Hagen *et al.*, 1995, 1997; Sørenson *et al.*, 1995; White *et al.*, 1995; Bennett *et al.*, 1996; Zara *et al.*, 1996), we have also examined the glycosylation in three diverse cell backgrounds;

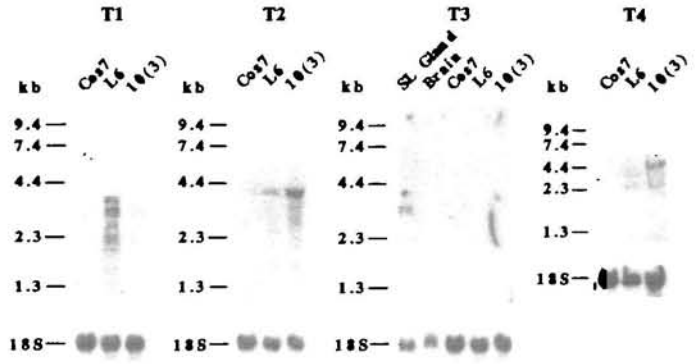

Fig. 1. Northern analysis of ppGaNTase-T1, -T2, -T3, and -T4 in COS7, L6 and 10(3) cells. Total RNA from each cell line was extracted, electrophoresed on 1% formaldehyde-agarose gels and transferred to Hybond-N membranes. Northern blots were hybridized with the probes listed at the top of each panel (-T1-T4) as well as an 18S rDNA control probe (bottom panels). Each lane contains 7.5 μg of total RNA. Cell lines are listed at the top of each lane and size markers are indicated on the left.

i.e., COS7 cells, a monkey kidney cell line; 10(3) cells, a murine fibroblast p53-null cell line; and, L6 cells, a rat skeletal muscle cell line. We show, first, that the repertoire and abundance of RNAs coding for the various known ppGaNTase isoforms differs among these cell lines and, second, that the presence of charged residues at positions −1 and +3 decreases O-glycosylation of peptides expressed in each of these cell types. Although we do not know which isoform(s) glycosylated which reporter *in vivo*, this result, taken together with the overlapping specificities of ppGaNTases observed *in vitro* (Sørenson *et al.*, 1995; White *et al.*, 1995; Bennett *et al.*, 1996; Zara *et al.*, 1996; Hagen *et al.*, 1997), implies that the amino acid charge distribution of a potential substrate may play a significant role in determining the activity of multiple isoforms *in vivo*.

Results

Analysis of ppGaNTase transcripts in COS7, L6 and 10(3) cells by Northern blot analysis

Northern blot analysis with isoform specific primers was performed to determine the relative abundance of ppGaNTase-T1, -T2, -T3, and -T4 transcripts in COS7, L6, and 10(3) cells. We took advantage of the high sequence conservation among isoforms derived from different species (Hagen *et al.*, 1995; Bennet *et al.*, 1996) to create probes that could be used with RNA derived from monkey, rat and mouse cell lines. Our results indicate that COS7 cells contain barely detectable levels of ppGaNTase-T1 and -T2 transcripts with no detectable -T3 or -T4 message observed. L6 cells express ppGaNTase-T1, -T2, and -T4 message, but -T3 transcript was not detected. ppGaNTase-T2 and -T4 transcripts were readily detectable in 10(3) cells whereas low levels of ppGaNTase-T1 message was observed; there was no evidence of -T3 transcript (Figure 1). Evidence for additional isoforms is growing (e.g., Hagen *et al.*, 1995, 1997; Hennet *et al.*, 1995), and thus it is likely that additional transferases may be expressed in each of these cell lines.

O-glycosylation in COS7 cells of six single sites derived from naturally occurring amino acid sequences

Five known glycosylation sites derived from native proteins (human erythropoietin, rEP; colony stimulatory factor, rCStF; bovine high molecular weight kininogen (rHMWK; and human glycoporin C, rGph-C) and one putative site from ovine submandibular gland mucin (rSMM) were selected (Elhammer *et al.*, 1993; Oh-eda *et al.*, 1990). Each of the O-glycosylation sites contained at least four amino acids derived from the native protein flanking either side of the potentially glycosylated threonine or serine (Table I). The sequences were expressed as secreted chimeric reporter proteins using an SV40 promoter-driven mammalian expression vehicle (Figure 2) as described previously (Nehrke *et al.*, 1996). Briefly, the reporter proteins were recovered from the culture media of transiently transfected COS7 cells by immunoprecipitation with an anti-FLAG antibody. To facilitate detection, bound materials were then radiolabeled with γ[^{32}P]ATP using heart muscle kinase. Serine residues within the heart muscle kinase site (RRASV) which acquire O-glycan would not be detectable in this system since the glycosylation would preclude subsequent phosphorylation.

Radiolabeled materials were desorbed and then resolved by denaturing Tricine SDS–PAGE (Figure 3). In a high percentage gel and crosslinking format, glycosylation of the small molecular weight reporters leads to a notably reduced mobility. The unglycosylated reporters are expected to migrate at ap-

Table I. Reporter proteins

Vector	Peptide sequence	Derived from[a]
pKN33	^{149}PDAASAAPL	human erythropoietin (rEP)
oligonucleotide: 5′-TA(TGCCCGACGCCGCTTCAGCTGCCCCGCTC)-3′		
pKN35	^{162}ALQPTQGAM	granulocyte colony stimulatory factor (rCStF)
oligonucleotide: 5′-TA(TGGGAGCTCTTCAGCCTACCCAGGGAGCTATGCTC)-3′		
pKN37	^{142}GPVVTAQYD	bovine high molecular weight kininogen (rHMWK)
oligonucleotide: 5′-TA(TGGGGCCCGTTGTCACAGCTCAGTATGAC)-3′		
pKN39	^{8}AFAFTAGVD	ovine submandibular gland mucin (rSMM)
oligonucleotide: 5′TA(TGGCTGGCGCCGGTACAGCTGGCGTAGAC)-3′		
pKN41	^{48}DPGMSGWPD	human glycophorin C (rGph-C)
oligonucleotide: 5′-TA(TGGACCCCGGGATGTCAGGATGGCCTGAC)-3′		

Numbers in suprascript refer to position of amino acid in sequence of native protein.
[a]The complete sequence of the ovine submandibular gland mucin is not known (see Elhammer et al., 1993 for sequence used)
The respective −1, +3E mutant reporters (pKN34, 36, 38, 40 and 42) were cloned using oligonucleotides that were nearly identical to those above, but had the underlined codons replaced with the sequence GAA, encoding glutamic acid.

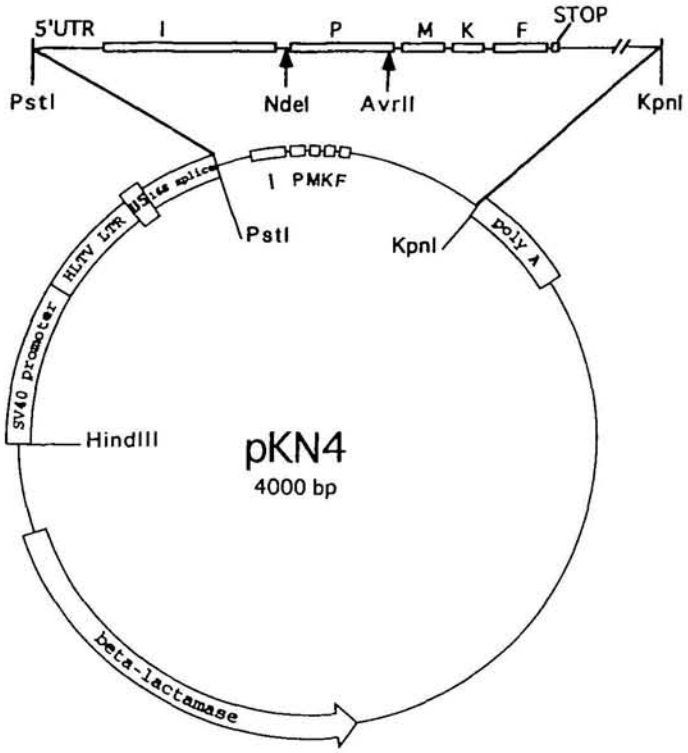

Fig. 2. Map of the eukaryotic expression vector pKN4. Reporter secretion was driven by the insulin secretory signal (I), which is fused in order to a peptide (P) sequence containing a single glycosylation site and its flanking sequence, a 6× His tag (M), a heart muscle kinase site (K), and the FLAG octapeptide antibody recognition epitope (F). The NdeI and AvrII cloning sites are not unique in pKN4. O-glycosylation site reporter-peptides included, rHVF (human von Willebrand factor), rEP (erythropoietin), rCStF (colony stimulatory factor), rHMWK (high molecular weight kininogen I), rSMM (submandibular gland mucin), and rGph-C (glycophorin C).

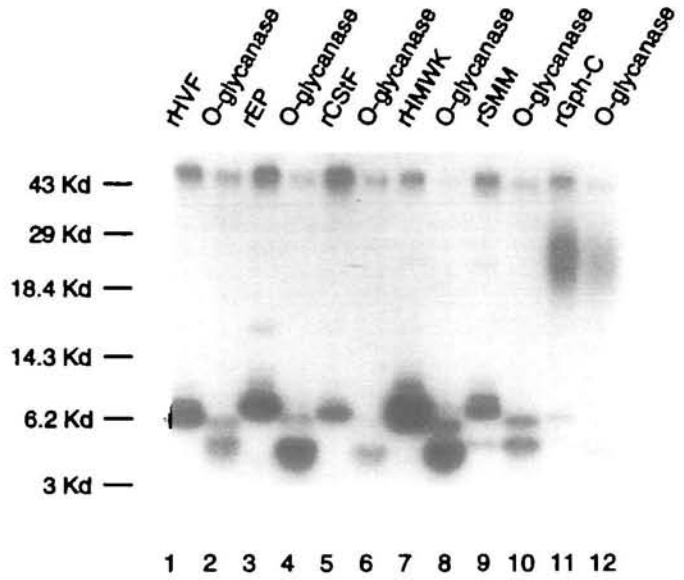

Fig. 3. O-glycosylation in COS7 cells of six single-site chimeric fusion proteins. COS7 cells were transiently transfected with expression constructs designed to secrete reporter proteins containing a single glycosylation site containing flanking sequence from human von Willebrand factor (lane 1), erythropoietin (lane 3), granulocyte colony stimulatory factor (lane 5), high molecular weight kininogen (lane 7), submandibular gland mucin (lane 9), or glycophorin C (lane 11). Immunoprecipitation via an antibody recognition tag and labeling of a bovine heart muscle kinase site in the reporter allowed visualization of the expressed protein. To assess O-glycosylation of the reporters, acid hydrolysis was used to remove sialic acid, then O-glycanase to cleave Gal β 1,3 GalNAc from the protein cores (lanes 2, 4, 6, 8, 10, and 12). The digestion products were analyzed by Tricine SDS–PAGE in a 16.5%T, 6%C resolving phase along with undigested reporters, as shown above. Size markers are shown on the left. The lower apparent molecular weight of the glycosidase-digested material was due to removal of a carbohydrate side chain. Differential sialic acid addition led to a slight heterogeneity of the glycosylated reporter in some samples; nonspecific and autophosphorylation resulted in the signal shown above 43 kDa (data not shown).

proximately 4.5 K, while the glycosylated species should migrate at around 6 K. The charge of the reporter will affect to some extent its mobility due to the small size of the reporter proteins, leading to differences in migration between various reporters containing the same total number of amino acids. The percent occupancy at the single O-glycosylation site was then determined by densitometric measurement of the separated glycosylated and unglycosylated species using a Molecular Dynamics PhosphorImager (Nehrke *et al.*, 1996). In this manner, sites in the rHVF, rEP, rCStF, and rHMWK and rSMM reporters were found to be glycosylated at over 95% occupancy when secreted from COS7 cells (Table II; Figure 3). In contrast to the other reporter proteins, rGph-C migrated predominantly as a high molecular weight form of approximately 25 K, as well as two minor forms of approximately 4.5 and 6 K.

To verify that each reporter protein was O-glycosylated, the purified proteins were subjected to mild acid hydrolysis and endoglycosidase digestion using O-glycanase. In all cases, acid hydrolysis and O-glycanase digestion led to a decrease in the apparent molecular weight of the labeled protein present at approximately 6 kDa, consistent with the removal of the O-glycan (i.e., from 6 K to 4.5 K; Figure 3). Some residual undigested material can be seen in the digested samples (lanes 2, 4, 6, 8, 10, and 12); this corresponds to reporter protein that has been acid hydrolyzed but still includes the core sugar residues, and digestion with a large excess of enzyme results in the disappearance of these bands (data not shown). The undigested material following mild acid hydrolysis migrates as a single band, suggesting that the multiple bands observed in many of the glycosylated reporter proteins are due to heterogeneous sialic acid content. Lectin precipitation of the reporters with jacalin agarose, which recognizes the same Gal β 1,3 GalNAc core structure as O-glycanase, corroborated the results of the glycosidase digestions (data not shown). However, there was little effect of O-glycanase on the 25 K species from the rGph-C reporter suggesting that it represented an alternatively glycosylated form of the reporter.

Glutamic acid at the −1 and +3 positions abolishes O-glycosylation of single site-acceptors in COS7 cells independently of flanking sequence

The O-glycosylation of rEP, rCStF, and rHMWK, rSMM and rGph-C reporters was markedly reduced by substitution of the residues at the −1 and +3 positions with glutamic acid (Table II; Figure 4). To verify that O-glycans were not acquired by the mutant −1,+3E reporters, acid hydrolysis and exo- and endoglycosidase digestions were performed (Figure 5). The mutated reporter peptides (with the exception of the rHVF reporter) were minimally affected (10–20%) by these treatments confirming that the mutated reporters fail to acquire significant levels of O-glycan. Previously, we have shown that substitution of glutamic acid at positions −1/+3 in rHVF reduces O-

Table II. *In vivo* glycosylation of reporter peptides

	Reporter designation	Peptide sequence[a]	Cell type and % glycosylation			NetOglyc score[b]	Chou-Fastman[c] prediction	Garnier[c] prediction
			COS7	10(3)	L6			
1	rHVF	Met-Ala-Gln-Val-Thr-Val-Gly-Pro-Leu	>95	>95	>95	0.212	Sheet	Sheet
2	−1+3E	. . . Glu . . . Glu .	10–40	<10	<10	0.036	Helix	Helix
3	rEP	Pro-Asp-Ala-Ala-Ser-Ala-Ala-Pro-Leu	>95	>95	>95	0.639	Helix	Helix
4	−1+3E	. . . Glu . . . Glu .	<10	<10	<10	0.326	Helix	Helix
5	rCStF	Ala-Leu-Gln-Pro-Thr-Gln-Gly-Ala-Met	>95	>95	>95	0.203	Turn	Helix
6	−1+3E	. . . Glu . . . Glu .	<10	<10	<10	0.056	Helix	Helix
7	rHMWK	Gly-Pro-Val-Val-Thr-Ala-Gln-Tyr-Asp	>95	>95	>95	0.207	Sheet	Sheet
8	−1+3E	. . . Glu . . . Glu .	<10	<10	<10	0.107	Helix	Helix
9	rSMM	Ala-Gly-Ala-Gly-Thr-Ala-Gly-Val-Asp	>95	61	70	0.065	Turn	Helix
10	−1+3E	. . . Glu . . . Glu .	<10	<10	<10	0.054	Helix	Helix
11	rGph-C	Asp-Pro-Gly-Met-Ser-Gly-Trp-Pro-Asp	P.G.[d]	21	30	0.796	Turn	Turn
12	−1+3E	. . . Glu . . . Glu .	P G.	<10	<10	0.442	Turn	Turn

[a]The amino acids immediately flanking the glycosylation site, which is shown in bold, are indicated.
[b]The score of the NetOglyc server for each sequence (out to eight amino acids on either side of the hydroxyamino acid) is shown, with a score of over 0.5 indicating predicted glycosylation.
[c]The Chou-Fastman and Garnier predictions of secondary structure at the glycosylation site are indicated.
[d]P.G. indicates that proteoglycan rather than mucin type O-glycosylation is occurring.

glycosylation (Nehrke *et al.*, 1996); in the present study this ranged from 10–40%. The slight differences in mobility between the glycosidase-treated wild-type sequences and the −1+3E mutants is due most likely to the effect of charged amino acids on the migration of these small reporter proteins through the SDS–PAGE gel, as noted previously (Nehrke *et al.*, 1996).

The −1,+3E change reduced the modification that led to the formation of a 25 K species of rGph-C (Figure 5, compare lanes 11 and 12). Since proteoglycans are known to be extended to great lengths in some cell backgrounds, we treated rGph-C with chondroitinase ABC (Figure 5, lane 13). This resulted in the disappearance of the 25 K form of the reporter and a concomitant increase in the signal of the 6 K form (the result of digestion to the initial xylose residue from which glycosaminoglycan chains are extended). We conclude that most of the rGph-C is modified with glycosaminoglycans when expressed in COS7 cells, although a small percentage becomes O-glycosylated with mucin-type oligosaccharide (Figure 4, lane 11 vs. 12).

O-Glycosylation of reporter proteins in murine fibroblast and rat skeletal muscle cell lines

We have previously demonstrated that the −1, +3 to glutamic acid change in the HVF reporter could severely reduce O-

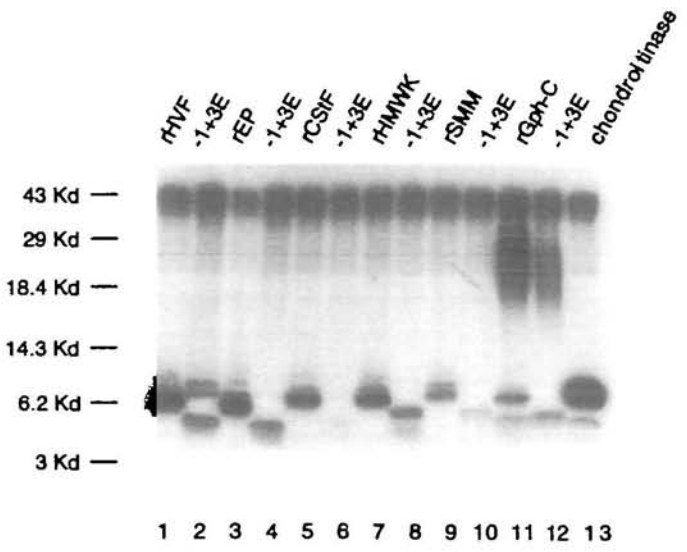

Fig. 4. Ablation of O-glycosylation in six flanking sequence contexts by replacement with glutamic acid at the −1 and +3 positions. rHVF (lane 1), rEP (lane 3), rCStF (lane 5), rHMWK (lane 7), rSMM (lane 9), and rGph-C (lanes 11 and 13) along with their respective mutant counterparts containing glutamic acid at the −1 and +3 positions relative to the glycosylation site (lanes 2, 4, 6, 8, 10, and 12) were immunoprecipitated from transiently transfected COS7 cell media, labeled and analyzed by Tricine gel electrophoresis. The rGph-C reporter was, in addition, treated with chondroitinase ABC prior to gel analysis (lane 13).

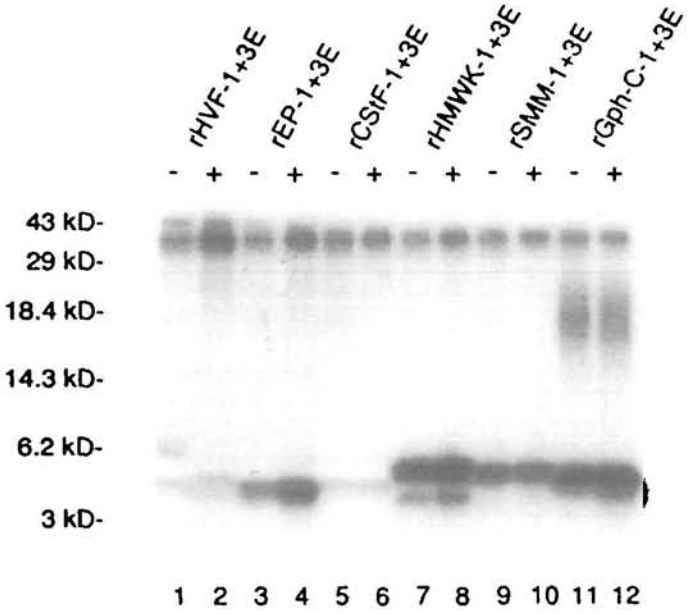

Fig. 5. Endoglycosidase analysis of −1,+3E reporter proteins secreted from transfected COS7 cells. The reporter proteins, listed at the top of each set of lanes, were first immunoprecipitated and labeled by kinase treatment. Samples either received no treatment (−) (lanes 1, 3, 5, 7, 9, and 11) or were treated following mild acid hydrolysis with a combination of O-glycanase and α-N-Acetylgalactosaminidase (+) (lanes 2, 4, 6, 8, 10, and 12). All material was analyzed by Tricine SDS–PAGE.

glycosylation in MCF-7, a mammary carcinoma cell line, as well as COS7 cells (Nehrke *et al.*, 1996). To determine if this effect could be observed in additional cell backgrounds expressing different repertoires of ppGaNTases, we examined two other cell lines representing divergent cell types: L6 is a rat skeletal muscle cell line, and 10(3) is a murine fibroblast p53-null cell line. Each was transiently transfected with the reporter constructs and the reporter proteins were analyzed as described above.

In L6 cells, the wild-type sequences were O-glycosylated, with four of the six at over 95% occupancy (rHVF, rEpo, rCStF (Figure 6A, lanes 1, 3, and 5), and rHMWK (Figure 6B, lane 1)), while the rSMM and rGph-C (Figure 6B, lanes 3 and 5) sequences were glycosylated at about 70% and 30% occupancy, respectively (Table II). In all of the cases examined, glycosylation was reduced to negligible (<15%) levels in the respective −1,+3E mutants (Table II; Figure 6A,B).

Similar findings were observed for reporter proteins secreted from transfected 10(3) cells; the occupancy of the sites in rHVF, rEpo, rCStF, and rHMW was >95% (Table II; Figure 7, lanes 1, 3, 5, and 7). In contrast, rSMM and rGph-C (Figure 7, lanes 9 and 11) were occupied to 61% and 21% levels, respectively (Table II). None of the −1,+3E mutants acquire appreciable levels of O-glycan (Figure 7, lanes 2, 4, 6, and 8). As in 10(3) cells, L6 cells expressing rGph-C do not modify the reporter with glycosaminoglycan.

Sequential treatment of the fusion proteins with a combination of mild acid hydrolysis, O-glycanase and, in the case of the −1,+3E mutants, α-N-acetylgalactosaminidase (to ensure that no unextended GalNAc has been added to the mutant reporters) verified the presence or absence of mucin type O-glycosylation in L6 (Figure 8A,B) and 10(3) (Figure 8C,D) cells. Two bands are observed with the rHMWK mutated reporter which we attribute to proteolysis since the mobility of neither band is affected by exo- or endoglycosidase treatment (Figure 8B).

Although we have used the same promoter to express each reporter peptide and all transfections were performed in tandem with equivalent amounts of plasmid, the recoveries of some of the reporters, particularly the mutants, were reduced (e.g., the rGph-C wild-type reporter is not well glycosylated in the L6 cell background and little signal is recovered; Figure 6B, lane 6). In part, this may reflect variation in transfection efficiency. Alternatively, the underglycosylated proteins may not be secreted as efficiently or they may be degraded. We

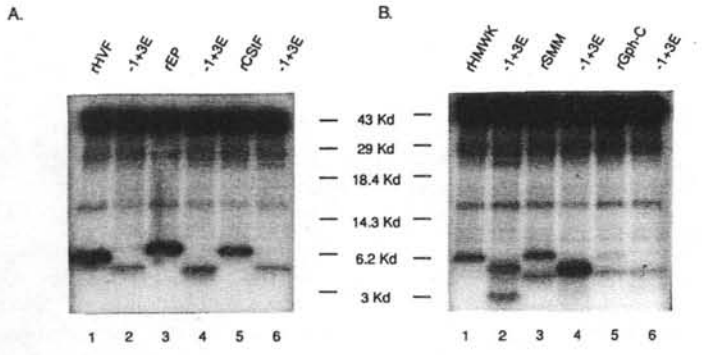

Fig. 6. Analysis of six glycosylated reporter proteins and their −1, +3E derivatives in L6 cells, a rat skeletal muscle cell line. Chimeric reporter proteins (A and B, lanes 1, 3, and 5) and their mutant derivatives (A and B, lanes 2, 4, and 6) were immunoprecipitated, labeled, and analyzed by Tricine SDS–PAGE, as shown.

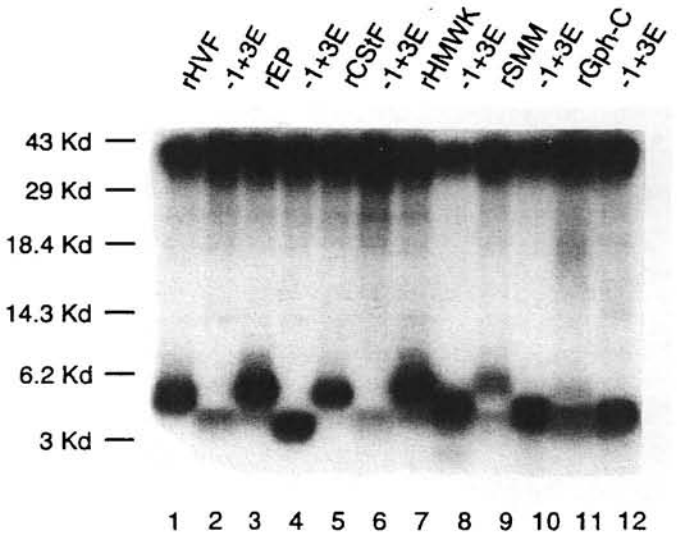

Fig. 7. Analysis of six glycosylated reporter proteins and their −1, +3E derivatives in 10(3) cells, a murine fibroblast p53 null cell line. Chimeric reporter proteins (lanes 1, 3, 5, 7, 9, and 11) and their mutant derivatives (lane 2, 4, 6, 8, 10, and 12) were immunoprecipitated, labeled, and analyzed by Tricine SDS–PAGE, as shown.

have observed this phenomenon with other substrates as well (Nehrke *et al.*, 1996).

Comparison of in vivo *glycosylation results with a neural net prediction method for O-glycosylation*

Four of the reporter proteins (rHVF, rCStF, rHMWK, and rSMM) were predicted to not be O-glycosylated by the NetOglyc server (Table II). However, when tested *in vivo,* these substrates were found to be nearly fully occupied in the three different cell backgrounds. In the other two cases (rGph-C and rEP), the server predicted correctly that glycosylation would occur. The rGph-C reporter is modified as a proteoglycan in COS7 cells, whereas, in 10(3) and L6 cells, mucin-type O-glycans were acquired, although the occupancy of the site in each cell background was less than 30%. The predictive Net O-glyc scores of all the −1, +3E mutant reporter proteins were less than their wild-type counterparts, suggesting that the NetOglyc server predictions are in general agreement with the experimental effect of charged amino acids which flank potential O-glycosylation sites.

Comparison of predicted secondary structure of parent and Glu-mutated reporters

We subjected each reporter peptide to secondary structure prediction using either the Chou-Fasman (1978) or Garnier (1978) method, to determine if the substitution of Glu at positions −1 and +3 alter the predicted secondary structure. By the method of Garnier, four of six peptides do not change their predicted secondary structure, yet in each case the Glu mutants fail to acquire O-glycans. Similar findings are also observed for two of these peptides using the Chou-Fasman method (Table II).

Discussion

Despite intensive efforts, attempts to predict sites of O-glycosylation have met with only modest success (O'Connell *et al.*, 1991; Wilson *et al.*, 1991; Elhammer *et al.*, 1993; Chou *et al.*, 1995; Hansen *et al.*, 1995, 1996). However, the data-

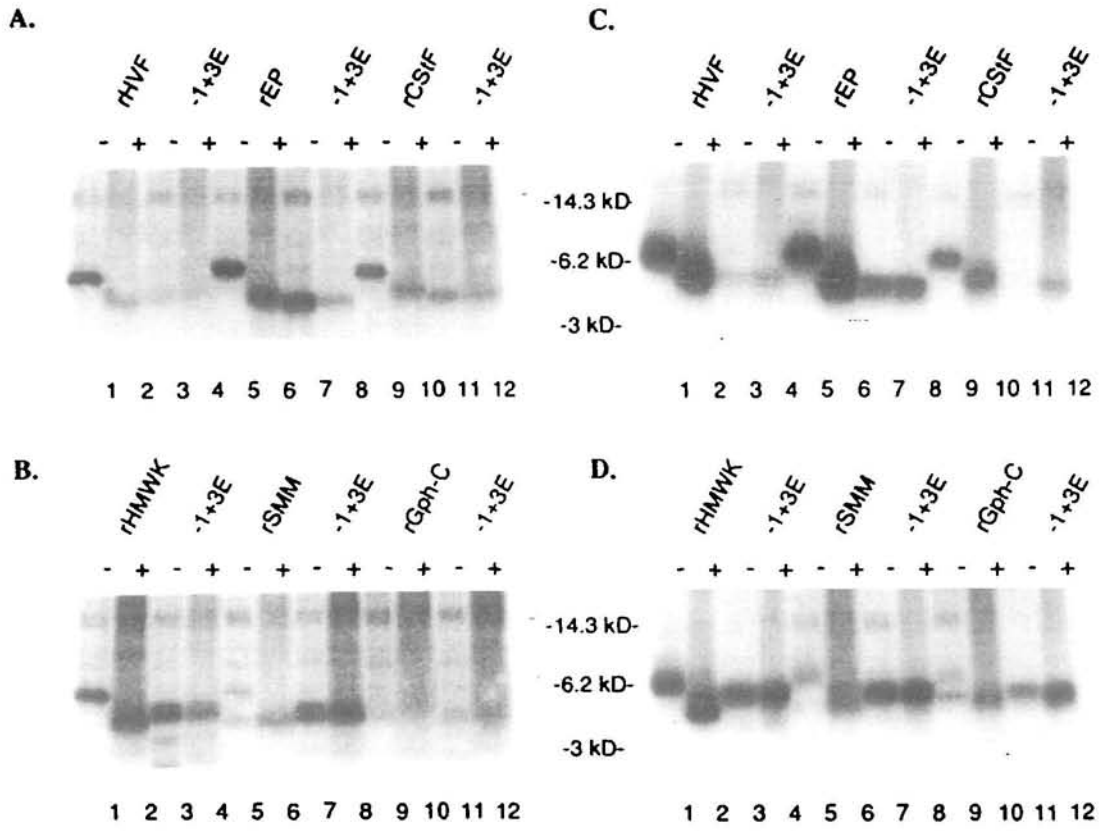

Fig. 8. Endoglycosidase digestions of reporter proteins produced from L6 and 10(3) cells. Immunoprecipitated, labeled reporter proteins from transfected L6 (A and B) or 10(3) (C and D) cells were treated as follows: no treatment, lanes 1, 3, 5, 7, 9, and 11, panels A, B, C, and D; (wild-type reporters) mild acid hydrolysis followed by O-glycanase digestion, lanes 2, 6, and 10, panels A, B, C, D; (mutant −1/+3E reporters) O-glycanase plus α-N-acetylgalactosaminidase digestion lanes 4, 8 and 12, panels A, B, C, and D. All materials were then analyzed by Tricine SDS–PAGE.

bases upon which these predictions depend are limited by: (1) experimental errors in assigning O-glycosylation sites and the percent occupancy of such sites (see Gooley and Williams, 1994); (2) the realization that there are multiple ppGaNTase isoforms (e.g., Hennet *et al.*, 1995; White *et al.*, 1995; Hagen *et al.*, 1997) which may have both unique as well as overlapping specificities; (3) the overall influence of secondary and tertiary structure which may render specific sites more or less accessible and cannot be readily ascertained from primary amino acid sequence (Aubert and Loucheux-Lefebvre, 1976; Elliot *et al.*, 1994; Yamada *et al.*, 1995; Sugahara *et al.*, 1996); and, (4) competing posttranslational modifications (Gerlitz *et al.*, 1993).

A basic assumption underlying this set of experiments is that different cell types of distinct lineages will express specific and varying levels of the various ppGaNTase isoforms. To date, four functional isoforms (-T1, -T2, -T3, and -T4) from mammalian cells have been identified (Hagen *et al.*, 1993, 1997; White *et al.*, 1995; Bennett *et al.*, 1996). The tissue distribution of these isoforms is quite specific, especially in the case of the -T3 and -T4 isoforms, but only one isoform-specific substrate has been identified (Bennett *et al.*, 1996). It follows from the overlapping specificities demonstrated *in vitro* that different isoforms of ppGaNTase may be capable of glycosylating the same substrate. Therefore, we have used Northern blot analysis with isoform specific probes to determine the form and abundance of ppGaNTase -T1, -T2, -T3, and -T4 in COS7, L6, and 10(3) cells. Since there is >90% nucleic acid conservation of a given isoform across mammalian species, we were able to probe cell lines derived from monkey, rat and mice. As shown in Figure 1, a different spectrum of transferase expression patterns was revealed in each cell type.

Despite the differences in ppGaNTase repertoire, the acquisition of O-glycans was demonstrated to be influenced by charge distribution of flanking amino acids. Although it is possible that the same transferase is responsible for the glycosylation of each of the reporters in all three cell types, it is more likely that several different ppGaNTase isoforms fail to O-glycosylate single sites which are flanked by charged residues at the −1 and +3 positions. Although we have not demonstrated the interchangeability of positive and negatively charged residues here, previously we determined that a double arginine replacement at the −1 and +3 sites in the HVF sequence has a similar inhibitory effect as that of glutamic acid (Nehrke *et al.*, 1996). Interestingly, none of the cell lines employed in this study appear to express appreciable levels of ppGaNTase-T3. This is significant because ppGaNTase-T3, thus far, is the only isoform which has been shown to glycosylate a substrate containing a charged residue at position +3, *in vitro* (Bennett *et al.*, 1996).

Given the small size of our reporters, it is likely that they assume little higher ordered structure. Nevertheless, we also tested the possibility that the presence of Glu at positions −1 and +3 could alter the secondary structure of the reporter. However, in several instances, we determined that the predicted structure was unaffected by these substitutions; yet, the mutated reporter fails to acquire O-glycan. Since each test sequence is presented within the same context, our results dem-

onstrate that where a substrate is normally fully occupied, the substitution of charge can abolish glycosylation *in vivo* without necessarily having impact on predicted secondary structure. Since, in a normal helical structure, the substituents on the amino acids at the −1 and +3 positions lie in the same outwardly facing plane, charged residues at both of these sites may constrain the entry point for ppGaNTase to the peptide backbone.

Our results also suggest that competition may be a factor controlling the onset of O-glycosylation; the rGph-C reporter contains a sequeon which, depending upon cell background, could be modified either with O-glycans or glycosaminoglycans (Gerlitz *et al.*, 1993).

Finally, we note that most O-glycosylation sites, especially within mucins, appear in clusters. Analysis of such sites is more difficult since changes in electrophoretic mobility are insufficient to distinguish among multiple potential sites. For that reason, we have begun studies to purify sufficient quantities of recombinant reporter to perform solid-phase microsequencing. This will permit assignment of quantitative occupancy levels at each potential site.

Materials and methods

Construction of plasmid vectors

A series of reporter proteins were constructed which consisted of short peptide sequences surrounding known O-glycosylation sites which were linked to an insulin secretory signal (which is cleaved prior to entry into the Golgi complex), a Ni^{2+} binding site and the FLAG octapeptide epitope (to facilitate purification) and a bovine heart muscle kinase site (to aid in detection of the reporter protein) using methods previously described (Nehrke *et al.*, 1996). Briefly, single site glycosylation sequeons were inserted into the background pKN4 vector (Nehrke *et al.*, 1996) that encodes the peptide MAQVTVGPL, which was derived from human von Willebrand factor. This was accomplished by first subcloning a PstI-KpnI fragment containing the fusion protein coding region into pBSIISK$^+$, making the construct pKN5. The NdeI-AvrII sites that flank the glycopeptide coding region (Figure 2) are only unique in the pKN5 vector. Kinased oligonucleotides corresponding to the sequences listed in Table I with cohesive NdeI-AvrII ends were inserted into the complementary sites in pKN5. These constructs were then digested with PstI and KpnI and the fragments containing the new fusion protein were cloned into the PstI-KpnI sites of pKN4, to create the constructs listed in Table I.

Cell culture and transfections

COS7 and 10(3) cells were maintained in Dulbecco's modified Eagle's medium (Life Technologies, Inc.) supplemented with 10% fetal bovine serum at 37°C in 5% CO_2, while L6 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 20% fetal bovine serum. Cells were transfected when reaching 70% confluency in 35 mm wells with either 1 μg (COS7 and L6 cells) or 3 μg (10(3) cells) of "Wizard Midipreps" DNA (Promega) using LipofectAmine (Life Technologies, Inc.) according to the manufacturer's protocol. Media were harvested 72 h posttransfection.

Immunoprecipitation, labeling, and gel analysis

To precipitate the reporter protein, 40 μl of 25% anti-FLAG agarose affinity gel (IBI) was added directly to 0.5 ml of cleared culture media. The mixture was rocked at 4°C overnight, pelleted, washed two times in phosphate-buffered saline containing 0.5% Tween-20, and then resuspended in 30 μl of HMK reaction buffer (20 mM HEPES, pH 7.0, 75 mM NaCl, and 15 mM MgCl2) containing 5 μCi of [γ^{32}P]-ATP (New England Nuclear) and 5 units of bovine heart muscle kinase (HMK, Sigma). Following incubation at 37°C for 1 h, the beads were washed two times as above and brought to a final volume of 25 μl in Tricine gel loading buffer (50 mM Tris pH 7.0, 12% glycerol, 4% SDS, 2% β-mercaptoethanol, and 0.01% Coomassie G-250). A sample (3 μl) was run on a 16.5%T/6%C resolving phase Tricine SDS–PAGE mini-gel containing both stacking and separating phases as described previously (Schagger and von Jagow, 1987). The gels were dried under vacuum and radiolabeled proteins were visualized by autoradiography and quantitated with a Molecular Dynamics PhosphorImager.

Acid hydrolysis and glycosidase digestions

We subjected 0.5 ml aliquots of cell culture media to mild acid hydrolysis (0.1 N sulfuric acid for 1 h at 80°C), neutralized with sodium hydroxide, centrifuged (14,000 × *g* for 5 min), and then incubated at 4°C with 40 μl of anti-FLAG agarose overnight to immunoprecipitate the reporter protein. Following washing and labeling as described above, 1 mU of O-glycanase (Genzyme) or 10 mU of chondroitinase ABC (Seikagaku Corp.) was added in a total volume of 20 μl according to the manufacturer's recommendations. The digestion products were eluted from the agarose beads by the addition of 0.25 volumes of 5× Tricine gel loading buffer and analyzed as described above.

Northern blot analysis

Total RNA from COS7, L6, and 10(3) cells was extracted according to the method described in Ausubel *et al.* (1995) for single step isolation of RNA from cultured cells. Electrophoresis in 1% formaldehyde-agarose gels and Northern blotting to Hybond-N membranes (Amersham) were performed according to Sambrook *et al.* (1989). Probes for ppGaNTase-T1, -T4 and 18S RNA were labeled as described previously (Hagen *et al.*, 1997) as was a -T3 specific probe (Zara *et al.*, 1996). A T2 specific fragment was generated by PCR amplification of a mouse T2 clone using primers mT2-S,d (TGGGGAAGTACGACATGATGA) and mT2-AS,d (CAACTCCATCCGCGAAGTGT). This amplification generates a ~340 bp product which was then asymmetrically labeled on the antisense strand by PCR using primer mT2-AS, as described previously (Bednarczuk *et al.*, 1991). All hybridizations were performed in 5× SSPE/50% formamide at 42°C with two final washes performed in 2× SSC/0.1% SDS at 65°C for 20 min.

Comparison of in vivo *glycosylation results with a neural net prediction method for O-glycosylation*

We submitted the sequence of each wild-type and mutant reporter to the NetOglyc network (Hansen *et al.*, 1995), including the eight amino acids on either side of the single potential O-glycosylation site. A score of 0.5 or above predicts that the sequence will be O-glycosylated, while a score of 0.5 or below predicts that it will not be.

Secondary structure predictions

These were performed using the Chou-Fasman (1978) and Garnier (1978) algorithms using the MacVector program (Oxford Molecular).

Acknowledgments

We are grateful to Dr. Arnold Levine for cell line 10(3) and Dr. Charles Thornton for the L6 cells. We also thank Pat Noonan for her help in preparing the manuscript. This work was supported in part by National Institutes of Health Grant DE-08108 (to L.A.T.). K.N. was supported by Training Grant T32 DE-07202 from the National Institutes of Health.

References

Aubert,J.-P. and Loucheux-Lefebvre,M.H. (1976) Conformational study of α_1-acid glycoprotein. *Arch. Biochem. Biophys.*, **175**, 400–409.

Bednarczuk,T.A., Wiggins,R.C. and Konat,G.W. (1991) Generation of high efficiency, single-stranded DNA hybridization probes by PCR. *Biotechniques*, **10**, 478.

Bennett,E.P., Hassan,H. and Clausen,H. (1996) cDNA cloning and expression of a novel human UDP-*N*-acetyl-α-D-galactosamine. *J. Biol. Chem.*, **271**, 17006–17012.

Carlsson,S.R., Lycksell,P.-O. and Fukuda,M. (1993) Assignment of O-glycan attachment sites to the hinge-like regions of human lysosomal membrane glycoproteins Lamp-1 and Lamp-2. *Arch. Biochem. Biophys.*, **304**, 65–73.

Chou,K.-C., Zhang,C.-T., Kézdy,F.J. and Poorman,R.A. (1995) A vector projection method for predicting the specificity of GalNAc-transferase. *Proteins Struct. Funct. Genet.*, **21**, 118–126.

Chou,P.Y. and Fasman,G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequences. *Adv. Enzymol.*, **47**, 45–147.

Elhammer,A.P., Poorman,R.A., Brown,E., Maggiora,L.L., Hoogerheide,J.G. and Kézdy,F.J. (1993) The specificity of UDP-GalNAc:polypeptide *N*-Acetylgalactosaminyltransferase as inferred from a database on *in vivo* substrates and from the *in vitro* glycosylation of proteins and peptides. *J. Biol. Chem.*, **268**, 10029–10038.

Elliott,S., Bartley,T., Delorme,E., Derby,P., Hunt,R., Lorenzini,T., Parker,V., Rohde,M.F. and Stoney,K. (1994) Structural requirements for addition of

O-linked carbohydrate to recombinant erythropoietin. *Biochemistry,* **33,** 11237–11245.

Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.,* **120,** 97–120.

Gerlitz,B., Hassell,T., Vlahos,C.J., Parkinson,J.F., Bang,N.U. and Grinnell,B.W. (1993) Identification of the predominant glycosaminoglycan-attachment site in soluble recombinant human thrombomodulin: potential regulation of functionality by glycosyltransferase competition for serine 474. *Biochem. J.,* **295,** 131–140.

Gooley,A.A. and Williams,K.L. (1994) Towards characterizing O-glycans: the relative merits of *in vivo* and *in vitro* approaches in seeking peptide motifs specifying O-glycosylation sites. *Glycobiology,* **4,** 413–417.

Hagen,F.K., Gregoire,C.A. and Tabak,L.A. (1995) Cloning and sequence homology of a rat UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase. *Glycoconjugate J.,* **12,** 901–909.

Hagen,F.K., Ten Hagen,K.G., Beres,T.M., Balys,M., VanWuyckhuyse,B.C. and Tabak,L.A. (1997) cDNA cloning and expression of a novel UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. *J. Biol. Chem.,* **272,** 13843–13848.

Hansen,J.E., Lund,O., Engelbrecht,J., Bohr,H., Nielsen,J.O., Hansen,J.-E. S. and Brunak,S. (1995) Prediction of O-glycosylation of mammalian protein: specificity patterns of UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase. *Biochem. J.,* **308,** 801–813.

Hansen,J.E., Lund,O., Nielsen,J.O., Hansen,H.-E.S., and Brunak,S. (1996) O-GLYCBASE: a revised database of O-glycosylated proteins. *Nucleic Acids Res.,* **24,** 248–252.

Hennet,T., Hagen,F.K., Tabak,L.A. and Marth,J.D. (1995) T cell-specific deletion of the polypeptide *N*-acetylgalactosaminyltransferase gene by site-directed recombination. *Proc. Natl. Acad. Sci. USA,* **92,** 12070–12074.

Jentoft,N. (1990) Why are proteins glycosylated? *Trends Biochem. Sci.,* **15,** 291–294.

Nehrke,K., Hagen,F.K. and Tabak,L.A. (1996) Charge distribution of flanking amino acids and Golgi transit time influence O-glycan acquisition *in vivo*. *J. Biol. Chem.,* **271,** 7061–7065.

Nishimori,I., Johnson,N.R., Sanderson,S.D., Perini,F., Mountjoy,K., Cerny,R.L., Gross,M.L. and Hollingsworth,M.A. (1994) Influence of acceptor substrate primary amino acid sequence on the activity of human UDP-*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosaminyltransferase. *J. Biol. Chem.,* **269,** 16123–16130.

O'Connell,B., Tabak,L.A. and Ramasubbu,N. (1991) The influence of flanking sequences on O-glycosylation. *Biochem. Biophys. Res. Commun.,* **180,** 1024–1030.

O'Connell,B., Hagen,F.K. and Tabak,L.A. (1992) The influence of flanking sequence on the O-glycosylation of threonine *in vitro*. *J. Biol. Chem.,* **267,** 25010–25018.

Oh-eda,M., Hasegawa,M., Hattori,K., Kuboniwa,H., Kojima,T., Orita,T., Tomonou,K., Yamazaki,T. and Ochi,N. (1990) O-linked sugar chain of human granulocyte colony-stimulating factor protects it against polymerization and denaturation allowing it to retain its biological activity. *J. Biol. Chem.,* **265,** 11432–11435.

Schagger,H. and von Jagow,G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.,* **166,** 368–379.

Sørenson,T., White,T., Wandall,H.H., Kristensen,A.K., Roepstorff,P. and Clausen,H. (1995) UDP-*N*-acetyl-α-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase. *J. Biol. Chem.,* **270,** 24166–24173.

Stadie,T.R.E., Chai,W., Lawson,A.M., Byfield,P.G.H. and Hanisch,F.-G. (1995) Studies on the order and site specificity of GalNAc transfer to MUC1 tandem repeats by UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase from milk or mammary carcinoma cells. *Eur. J. Biochem.,* **229,** 140–147.

Strub,J.-M., Sorokine,O., Van Dorsselaer,A., Aunis,D. and Metz-Boutigue, M.-H. (1997) Phosphorylation and O-glycosylation sites of bovine chromogranin A from adrenal medullary chromaffin granules and their relationship with biological activities. *J. Biol. Chem.,* **272,** 11928–11936.

Sugahara,T., Pixley,M.R., Fares,F. and Boime,I. (1996) Characterization of the O-glycosylation sites in the chorionic gonadotropin β subunit *in vivo* using site-directed mutagenesis and gene transfer. *J. Biol. Chem.,* **271,** 20797–20804.

Tabak,L.A. (1995) In defense of the oral cavity: structure, biosynthesis, and function of salivary mucins. *Annu. Rev. Physiol.,* **57,** 547–564.

White,T., Bennet,E.P., Takiol,K., Sorensen,T., Bonding,N. and Clausen,H. (1995) Purification and cDNA cloning of a human UDP-*N*-acetyl-α-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase. *J. Biol. Chem.,* **270,** 24156–24165.

Wilson,I.B., Gavel,Y. and von Heigne,G. (1991) Amino acid distributions around O-linked glycosylation sites. *Biochem. J.,* **275,** 529–534.

Yamada,T., Uyeda,A., Takao,T., Shimonishi,Y., Matsushima,M. and Kikuchi,M. (1995) O-glycosylation of the Thr70 residue of cell-adhesive lysozyme in yeast. *Eur. J. Biochem.,* **230,** 965–970.

Zara,J., Hagen,F.K., Ten Hagen,K.G., Van Wuyckhuyse,B.C. and Tabak,L.A. (1996) Cloning and expression of mouse UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase-T3. *Biochem. Biophys. Res. Commun.,* **228,** 38–44.

Received on June 10, 1996; revised on June 5, 1997; accepted on June 18, 1997