Glycoconjugate Journal (1995) **12**: 901–909

Cloning and sequence homology of a rat UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase

FRED K. HAGEN, CHRISTINE A. GREGOIRE and LAWRENCE A. TABAK*

Departments of Dental Research and Biochemistry, School of Medicine and Dentistry, University of Rochester, 601 Elmwood Ave., Box 611, Rochester, New York 14642, USA

Received 12 May 1995, revised 5 July 1995

A UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase (polypeptide GalNAc transferase) cDNA was amplified from rat sublingual, submandibular and parotid glands, brain, skeletal muscle, and liver, using the polymerase chain reaction (PCR) and sequences derived from bovine polypeptide GalNAc transferase-Type 1 (polypeptide GalNAc transferase-T1). The transcripts encoding the rat sublingual gland and bovine enzymes were 91% identical in nucleotide sequence, except in their 5′ and 3′ untranslated regions. The enzymes encoded by the rat and bovine cDNAs were 559 amino acids in length and were virtually identical (98% amino acid sequence identity and 99.5% homologous overall). Northern blot analysis indicates that the polypeptide GalNAc transferase-T1 transcripts are expressed in many tissues but at widely differing levels. Although the amino acid sequence of polypeptide GalNAc transferase-T1 is conserved among mammals, the pattern of tissue expression varies between rats and humans. For example, the steady-state level of polypeptide GalNAc transferase-T1 transcript is quite low in lung relative to other rat tissues, whereas high expression of this transcript is detected in human lung. Therefore, we surmise that isoforms of polypeptide GalNAc transferase must exist and that isoforms are expressed in a tissue-dependent fashion. Searches of the GenBank database have revealed homologous sequences for several isoforms derived from several human tissues. In addition, hypothetical proteins from *C. elegans* also display strong homology; evidence suggests six ancestral isoforms of polypeptide GalNAc transferases may exist in *C. elegans*.

Keywords: polypeptide GalNAc transferase, isoforms, O-glycosylation, mammals, nematodes, sequence homology, glycosyltransferase, mucin

Introduction

Biosynthesis of mucin-type O-linked glycans involves the step-wise transfer of monosaccharides to selected hydroxyamino acids of the polypeptide backbone. This process is initiated by action of a UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase (polypeptide GalNAc transferase) which catalyses the transfer of *N*-acetylgalactosamine (GalNAc) to hydroxyl groups of threonine and serine residues [1, 2]. Sites of O-glycosylation are rich in proline, serine, threonine, alanine, and glycine, and low in charged residues at specific positions relative to the glycosylated amino acid [3, 4]; however, no obvious consensus sequence has been derived from database analyses [5]. In addition, sites of O-glycosylation frequently exist as clustered arrangements of hydroxyamino acids, which are selectively and not fully glycosylated. The molecular basis for glycosylation site selection and heterogeneity occupancy is not clearly understood. *In vitro* glycosylation studies using purified enzymes from bovine colostrum or porcine submaxillary gland or from a single recombinant clone of the polypeptide GalNAc transferase demonstrated that a single isozyme can act on a diverse set of peptide acceptor sequences, yet the glycosylation rates of some peptides, particularly those containing serine, are low [1, 6–8]. It is therefore possible that more than one polypeptide Gal-

*To whom correspondence should be addressed.

NAc transferase may be involved in the O-glycosylation apparatus, and they may be differentially expressed in a given cell- or tissue-type.

The cDNA for a polypeptide GalNAc transferase has been isolated from bovine placenta [1] and intestine [9] and has been shown to hybridize to transcripts prepared from human heart, brain, placenta, lung, liver, skeletal muscle and pancreas, but not from kidney. The absence of message in the human kidney suggested that O-glycosylation must involve different polypeptide GalNAc transferase isoforms.

To advance our studies of mucin biosynthesis in rat model systems, we have cloned a rat polypeptide GalNAc transferase and compared it to other mammalian polypeptide GalNAc transferases and expressed sequence tags (EST) from human and *C. elegans* cDNA entries in the GenBank database. We designate this rat isoform as a polypeptide GalNAc transferase-Type 1 (polypeptide GalNAc transferase-T1) because it is essentially identical to the first polypeptide GalNAc transferase that was cloned [1, 9]. As an initial step towards understanding the mechanisms which control expression of the polypeptide GalNAc transferase-T1 gene, the tissue distributions of the rat and human enzymes have been compared. In addition, we have obtained evidence which suggests that the polypeptide GalNAc transferase is an ancient enzyme with several distinct, albeit homologous, forms detected in the *C. elegans* GenBank database.

Material and methods

Isolation of genomic sequences encoding the rat polypeptide GalNAc transferase

To obtain sequences encoding the amino and carboxyl terminus of the rat gene product, two million lambda Dash II phage from a Sprague-Dawley male rat genomic library were screened with either a full-length, 5′ partial, or 3′ partial cDNA probe encoding the secreted bovine polypeptide GalNAc transferase [1]. Positive clones for the genomic sequences encoding the ends of the gene product were either subcloned into pBSIISK+ or were digested with Sau3A and subcloned into M13SK (Stratagene). Each subclone obtained in this manner was sequenced on both strands using multiple M13 subclones. M13SK subclones hybridizing to cDNA probes were sequenced using Ladderman Dideoxy Sequencing Core Reagents (PanVera), an infrared fluorescent (IRD40)-labelled primer, and a Model 4000L LiCor Automated DNA Sequencer.

Reverse transcription-polymerase chain reaction (RT-PCR)

cDNA was synthesized using oligo dT18, a First Strand cDNA Synthesis Kit (Clonetech) and poly A+ mRNA from rat (Male, Sprague-Dawley) parotid, submandibular and sublingual glands, brain, liver, and skeletal muscle. To amplify the rat polypeptide GalNAc transferase cDNA sequence, the primers described in Fig. 1 were designed from genomic sequence data and used at 1 μM final concentration and amplified for 40 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 2 min. PCR products were either sequenced directly using a Sequenase PCR Product Sequencing Kit (United States Biochemical) or were cloned in a TA Cloning Vector pCRII (Invitrogen) and later subcloned into M13SK prior to DNA sequencing.

Northern blot analysis

A multiple tissue Northern blot (Clonetech) was prehybridized in 50% formamide, 5× SSPE, 7.2× Denhardt's solution, 2% SDS, 100 $\mu g\,ml^{-1}$ denatured salmon sperm DNA at 42 °C and hybridized in the presence of a 216 bp probe encompassing sequences 245–461 nt (Fig. 1). Blots were washed first in 2× SSC, 0.05% SDS at room temperature and then in 0.1× SSC, 0.1% SDS at 50 °C.

Preparation of tissues extracts, enzyme, and protein assays

Triton X-100 extracts of Wistar male rat tissues were used as a source of enzyme and prepared in the following manner: 30 mg tissue was homogenized for 2 min in a Mini-Beadbeater-8 Cell Disrupter (Biospec Products) using 1 ml of 1 mm glass beads and 1 ml homogenization medium (500 mM sucrose, 2 mM tris maleate, pH 6.4, 1 mM $MnCl_2$ and $MgCl_2$, 1% Dextran (250 000 MW), and 5 mM β-mercaptoethanol). This homogenate was diluted to 5.2 ml with H_2O and centrifuged for 1 h at 46 100 rpm in a SW 50.1 rotor. The pellet was resuspended in 300 μl extraction buffer (2.5% Triton X-100, 50 mM cacodylic acid, pH 6.6, 50 mM β-mercaptoethanol) for 1 h, pelleted at 14 000 × **g** for 15 min and then re-extracted in 150 μl of buffer. The supernatants were combined, aliquoted and frozen. Polypeptide GalNAc transferase enzyme assays were performed using 10–20 μl of extract and the following components in a final volume of 50 μl: 10 mM AMP to inhibit phosphatases, 50 μM UDP-[C^{14}]-GalNAc (50 000 cpm); 200 μM EA2 peptide substrate [Pro-Thr-Thr-Asp-Ser-Thr-Thr-Pro-Ala-Pro-Thr-Thr-Lys], 1% Triton X-100, 40 mM cacodylic acid, pH 6.6, 10 mM $MnCl_2$. Reaction products were purified by ion exchange chromatography using Dowex 1-×8 resin (BioRad). Phosphatase assays were performed as above; however, the peptide substrate (EA2) was omitted from the reaction. A modified Lowry protocol was used for protein determination.

GenBank sequence data

Bovine sequences were obtained from GenBank, accession numbers L17437, L16925 and L07780. The amino

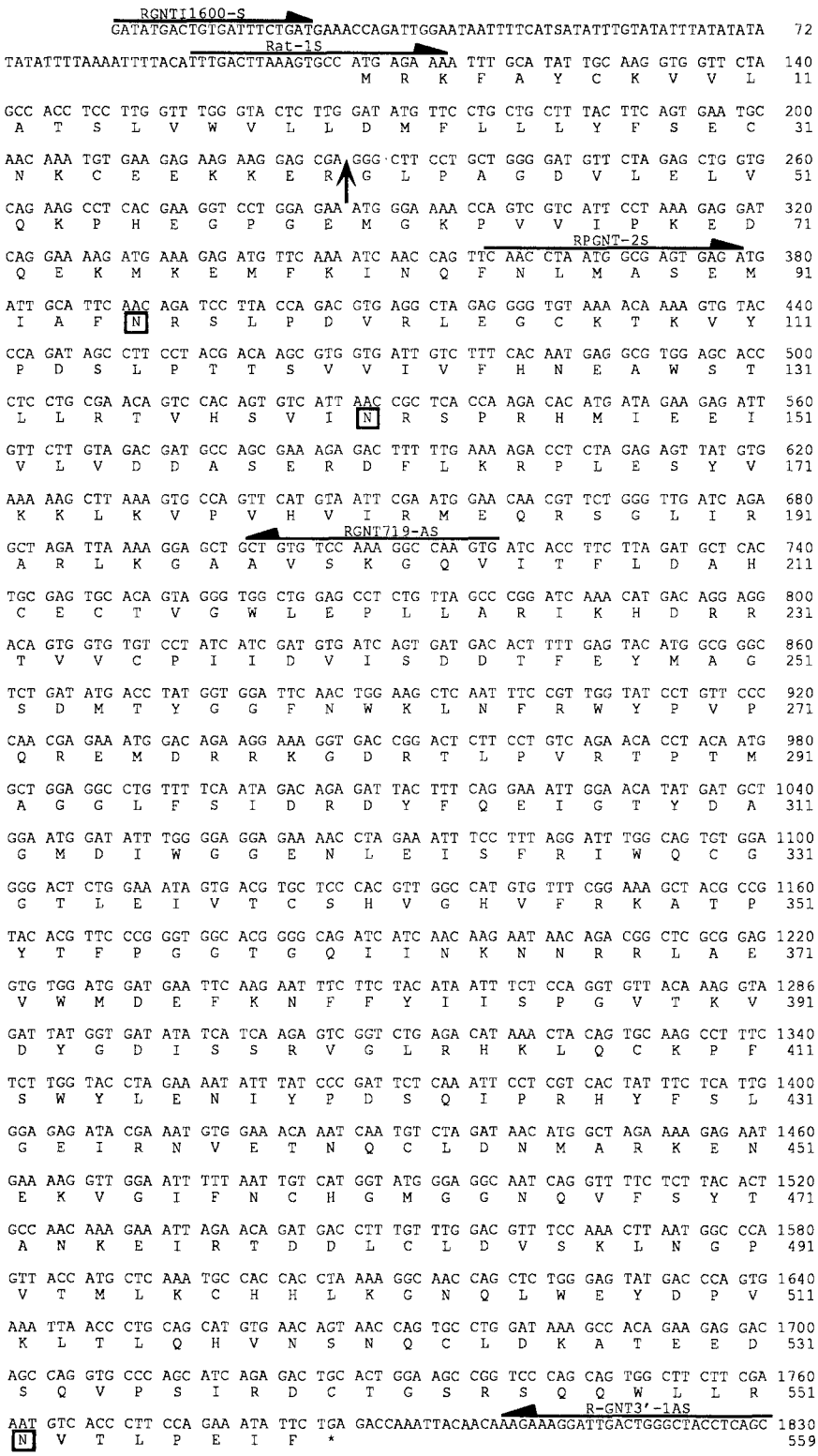

Figure 1. Nucleotide and amino acid sequence of a rat polypeptide GalNAc transferase-T1 cDNA. The cDNA sequence was obtained by amplifying sublingual gland cDNA in three overlapping segments, using the following primer pairs: (1) RGNT 1600-S and RGNT719-AS; (2) Rat-IS and RGNT719-AS; and (3) RPGNT-2S and R-GNT3′-IAS. Primer sequences used are indicated by horizontal arrows. The vertical arrow refers to the peptide cleavage site that generates a secreted catalytically active form of the enzyme. This cleavage site is inferred from bovine polypeptide GalNAc transferase-T1 cleavage data [1, 9]. Amino acids are indicated by single-letter notation. Putative N-glycosylation sites are boxed.

acid sequence for the full-length *C. elegans* cDNA (accession numbers L16621 and L18807) was conceptually translated and assembled using the GENEFINDER computer program [10]. A partial human polypeptide GalNAc transferase-T1 expressed sequence tag was derived from an infant brain (T08891), white blood cell (T34584), gall bladder (T32595), CD34+ cDNA (T25717), and an unannotated cDNA (F11419).

Results

Rat polypeptide GalNAc transferase cDNA

A rat cDNA containing the full coding region was amplified in three overlapping parts from parotid gland, brain, liver, skeletal muscle, submandibular gland, and sublingual gland cDNA. Identical size products were obtained from these tissues (data not shown) and the sublingual PCR products were cloned and sequenced (Fig. 1). Primer extension analysis revealed that the 5′ UTR is 295 nt (data not shown), indicating that this is a partial clone. The nucleic acid sequence displayed an overall 91% identity with cDNAs from the bovine placenta and intestinal polypeptide GalNAc transferase, except in their 5′ and 3′ untranslated region (UTR). The 5′ UTR obtained consists of 105 nt and contains two AUG codons followed by a total of seven termination codons, such that at least two stop codons exist in each reading frame. An in-frame termination codon UAA precedes the translation initiation site: 5′[*UAA*AGU*GC*-C*AUGA*]3′, which is homologous to most of but not the complete Kozak consensus sequence: [−5]CC*R*CC-*AUG*G[+4] [11]. The cytosines at −5 and −4 in the consensus sequence are not present in the rat GalNAc transferase translation initiation site, as is frequently observed in Kozak database when a purine is present at −3. The +4 position shows an adenine and not a guanine, which is only present in 40% of Kozak initiator codons. Conceptual translation of the rat cDNA revealed a 559 amino acid open reading frame, encoding a type 2 membrane protein. A potential transmembrane anchor is positioned nine amino acids from the N-terminus and is encoded by a run of 20 hydrophobic amino acids. A peptide cleavage site follows the transmembrane domain and is identical in amino acid sequence to the bovine cleavage site (Fig. 1) and thus must be responsible for producing the secreted form of the enzyme seen in rat tissue. The remaining 519 amino acids encode the functional catalytic domain of the enzyme.

Rat tissue distribution of polypeptide GalNAc transferase

Northern blot analysis indicated that multiple size transcripts were present in most tissues examined (Fig. 2). Two major transcripts were 4.65 kb and 4.16 kb in size and were present in most of the tissue tested, while a minor transcript of 3.52 kb was expressed predominantly

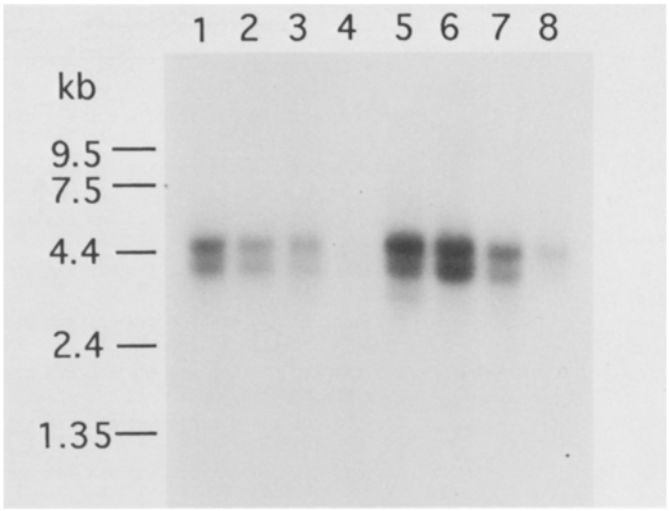

Figure 2. Northern blot analysis of rat poly A mRNA. A multiple tissue Northern blot for Clontech contained 2 μg of poly A^+ RNA from the following tissue sources in lanes 1–8: heart, brain, spleen, lung, liver, skeletal muscle, kidney and testis. Sequence from nucleotides 1649–1830 of the rat polypeptide GalNAc transferase-T1 cDNA were P^{32}-labelled and used as a hybridization probe.

in the liver, kidney and spleen. To ascertain if the different size transcripts represent alternative splice variants, the multiple tissue Northern blot was hybridized with sequences derived from either the 5′ or 3′ end of the gene product. Each probe displayed an identical pattern of expression as seen in Fig. 2. Reverse transcription-polymerase chain reaction (RT-PCR) experiments using primers derived from both ends of the cDNA revealed only a single size PCR product (data not shown). Therefore the differences in the sizes of the three transcripts appear to reside in their 5′ and/or 3′ untranslated (UTR) sequences. Paradoxically, the lung, which produces mucins, had barely detectable levels of transcript. We therefore assayed a panel of six tissues for polypeptide GalNAc transferase activity (Table 1). It is clear that the steady-state level of transcript does not correlate well with the total transferase activity of the tissue extracts.

Homology to mammalian expressed sequence tags

This rat polypeptide GalNAc transferase is virtually identical to the bovine placental and intestinal polypeptide GalNAc transferase enzyme (Fig. 3). All but two amino acid positions were homologous. Of ten positions that were not identical, eight were homologous. Over the entire protein, the rat and bovine amino acid sequences are 98% identical and 99.5% homologous. Therefore, we designate this form of transferase as a polypeptide GalNAc transferase-T1.

Table 1. Polypeptide GalNAc transferase enzyme activity and T-1 Northern blot data

Tissue	*Total activity*[a] *(pmol h^{-1})*	*Specific activity (pmol h^{-1} mg^{-1})*	*Northern T-1 data*[b] *(Peak value)*	*Normalized total activity*[c] *(%)*	*Normalized northern peak pixel value*[c] *(%)*
Kidney	6390	4581	7604	225	122
Spleen	4841	6147	4548	170	73
Liver	2696	3112	12272	95	197
Lung	1589	2692	1124	56	18
Testis	895	1658	3274	31	52
Brain	641	955	8634	23	138
Average	2842	3191	6243	100	100

[a]In an extract from 30 mg of tissue.
[b]Pixel value of major band in Northern blot (Fig. 2) was measured using a Molecular Dynamics Phosphoimager.
[c]The enzyme activity or Northern blot hybridization signals were displayed as a percentage of the average for these six tissues.

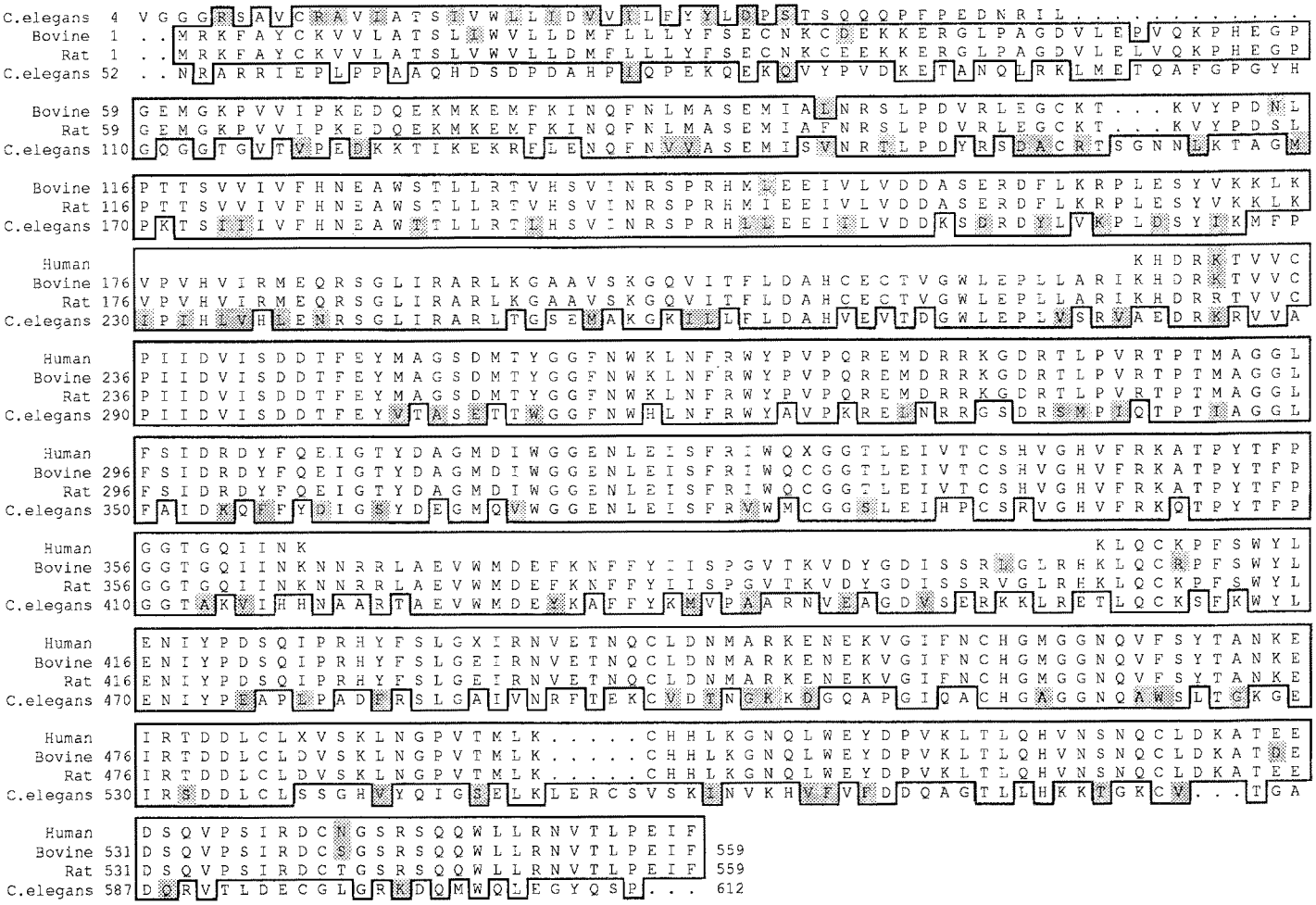

Figure 3. Amino acid homology of polypeptide GalNAc transferases. Predicted amino acid sequences were compiled from the GenBank database entries as described in the Material and methods. Only partial sequence was available for the human transferase. The displayed human sequences are encoded by five expressed sequence tags: (GenBank ID, amino acid position) T34584, position 227–323; T32595, position 233–324; T08891, position 254–364; F11419, position 405–522; T25717; position 488–559. Multiple sequence alignments were performed using the PileUp computer program in the GCG Sequence Analysis Software package. Gaps introduced to optimize alignments are indicated by periods. Identity with rat sequence is indicated in open boxes; similar residues are highlighted by shading and grouped as follows: [D, E]; [K, R]; [G, A]; [N, Q, S, T]; [A, L, I, V, M]; [W, Y, F]. The *C. elegans* protein is ZK688.8. Blank lines indicate that no sequence information is available.

A TBLASTN analysis [12] of the best expressed sequence tag and GenBank databases using the rat polypeptide GalNAc transferase-T1 amino acid sequence at high stringency identified more than five partial cDNA sequences from numerous tissues of human origin with up to 95% nucleic acid identity with the rat cDNA. Comparison of these partial amino acid sequences indicated that at least 293 amino acids of the human polypeptide GalNAc transferase-T1 enzyme are 98% identical with rat and bovine polypeptide GalNAc transferase-T1 enzyme (Fig. 3).

Homology to a C. elegans hypothetical protein

The rat polypeptide GalNAc transferase-T1 sequence displays low homology with numerous cDNA entries in the GenBank database. While most cDNAs displayed short (<20 amino acids) segments of modest homology (<70%), one of the entries with highest homology was a hypothetical 67.9 kDa protein (accession number Y028-CAEEL), which was compiled by the GENEFINDER Program using genomic sequence obtained from the *C. elegans* Genome Sequencing project [10]. Sequence alignments between this *C. elegans* protein (named ZK688.8) and mammalian proteins revealed numerous short (<10 amino acid) repeats with 100% identity and a central 400 aa region with an overall 78% homology (Fig. 4C). The hydrophobic membrane anchor/Golgi retention signal, however, is present and conserved in the N-terminus as a 22 amino acid segment with 50% identity and 91% homology between rat and nematodes (Fig. 3 top line). Hydropathy analysis illustrates the difference in placement of the membrane anchor relative to the catalytic domain (Fig. 4A and B). The spacing between most of the identical segments (especially between amino acid positions 109 and 496) is conserved between the nematode entry and mammalian polypeptide GalNAc transferase-T1 enzyme, except with the relative position of the N-terminal membrane anchor. The membrane anchor is 47 amino acids further from the conserved domain in *C. elegans* than in the mammalian GalNAc-T1 enzyme. This spacing difference accounts for most of the 53 amino acid size difference between the 612 amino acid *C. elegans* protein and 559 amino acid mammalian enzyme. The last 75 aa on the C-terminus of the *C. elegans* homologue are not conserved. Both enzymes possess three putative N-glycosylation sites (NXT or NXS); the two putative sites closest to the N-terminal membrane anchor are highly conserved in position and

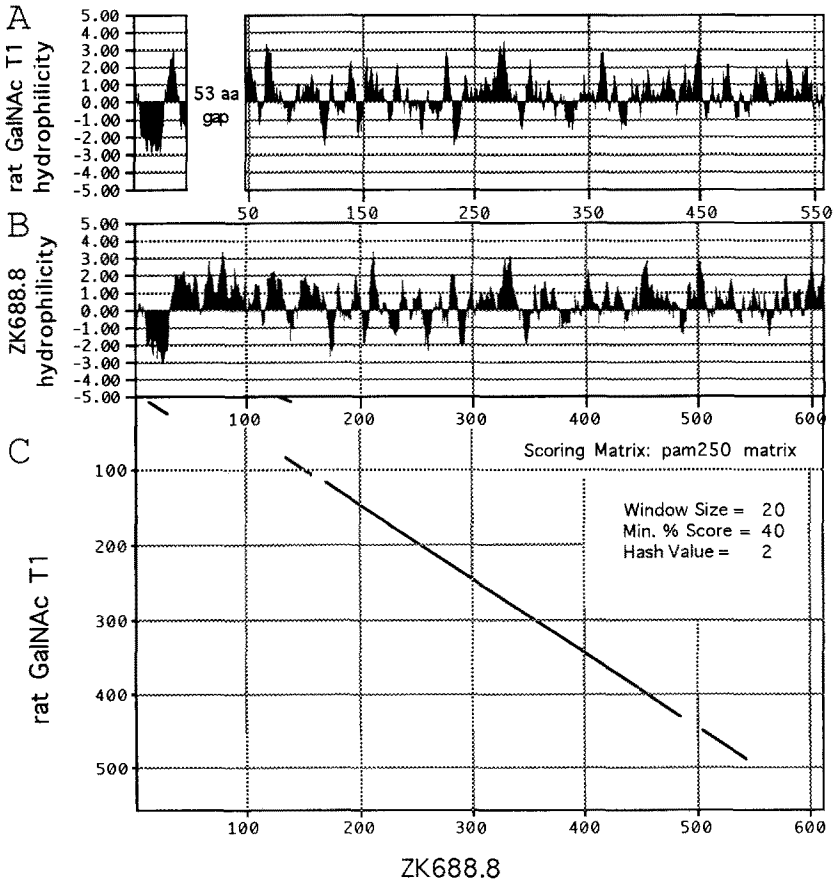

Figure 4. Comparisons of a rat polypeptide GalNAc transferase and *C. elegans* homologue. Kyte-Doolittle hydrophilicity analysis (A) rat polypeptide GalNAc transferase-T1 and (B) the *C. elegans* protein ZK688.8. (C) Pastell-DNA matrix comparisons.

sequence, while the position and sequence of the third sites are clearly different. Interestingly, Triton X-100 extracts of *C. elegans* contain polypeptide GalNAc transferase enzyme activity (data not shown).

A family of potential polypeptide GalNAc transferase homologues in C. elegans and humans

Additional clones from the expressed sequence tag (EST) database dbest and GenBank were identified using low stringency search criteria (Fig. 5). A comparison of a mammalian polypeptide GalNAc transferase-T1 enzyme with five putative *C. elegans* homologues is limited by the available sequence data for each EST but does suggest that these could represent at least five distinct enzyme isoforms. Further sequencing of two of these *C. elegans* clones revealed a range of 47 to 68% identity and 66 to 85% homologies with the mammalian ppGalNAc-T1 enzyme, numerous invariant amino acid residues, and highly conserved spacing between invariant amino acid positions and in the position of the termination codon (data not shown). In addition to the *C. elegans* clones, three human expressed sequence tags were identified using an initial screen parameter of greater than 65% homology over a window of 21 or more amino acids. One of these has been subcloned, further sequenced and shown to contain many of the same invariant amino acids as found in the *C. elegans* clones (data not shown).

Discussion

The rat genome encodes a polypeptide GalNAc transferase that is 98% identical to both the full-length bovine GalNAc polypeptide transferase-T1 enzyme and to the partial human sequence that is available in the GenBank database. This isozyme is expressed widely in the tissues examined in this study and thus suggests that this isoform may contribute to the biosynthesis of a large number of O-linked glycoproteins across numerous organ systems. However, the finding that polypeptide transferase activity does not correlate with steady-state levels of T1 transcript argues that different isoforms of the enzyme may also contribute. Amino acid sequence comparisons of mucin-like proteins have revealed strong distinctions in the potential O-glycosylation sites expressed in the

```
CLONE ID      HOMOLOGY  SEQUENCE
1.CE: D34980      83%   VYHNEAYSTLLRTVWSVIDRSPKELLKEIILVDDFSDREFLRYPXDTTLKPLPTDIKIIRSKERV..
                        V+HNEA+STLLRTV SVI+RSP+ +++EI+LVDD S+R+FL+ P ++ ++ L   + +IR ++R
Rat position 168         VFHNEAWSTLLRTVHSVINRSPRHMIEEIVLVDDASERDFLKRPLESYVEKLKVPVHVIRMEQRS..

2.CE: D27787      82%   PFRSPTHAGGLFAINRLWFKELGYYDEGLQIWGGEQYELSFKIWQCGGGIVFVPCSHVGHVYQKS
                        P R+PT AGGLF+I+R +F+E+G YD G+ IWGGE  E+SF+IWQCGG +  V CSHVGHV++K+
Rat position 285         PVRTPTMAGGLFSIDRDYFQEIGTYDAGMDIWGGENLEISFRIWQCGGTLEIVTCSHVGHVFRKA

3.CE: D36186      72%   LEPLLARITENXKXVVAPIIDVINVDNFNYVGASXDLRGGFDWTLVFRWEFMNEQLRKERHAHPT..
                        LEPLLARI  + + VV PIIDVI+ D F Y+  S    GGF+W L FRW  + ++    R    T
Rat position 219         LEPLLARIKHDRRTVVCPIIDVISDDTFEYMAGSDMTYGGFNWKLNFRWYPVPQREMDRRKGDRT..

4.CE: Z14442      72%   MISVHRTLPTNIXAECKTEKYNXNLPRTXVIICXHNEAWSVLLRTVHSVLERTPXHLLEEXVLVX..
                        MI+ +R+LP      CKT+ Y  +LP T V+I  HNEAWS LLRTVHSV+ R+P H++EE VLV
Rat position 70          MIAFNRSLPDVRLEGCKTKVYPDSLPTTSVVIVFHNEAWSTLLRTVHSVINRSPRHMIEEIVLVD..

5.CE: M89126      65%   LIRAKLAGAREAVGDIIVFLDSHCEANHGWLEPIVQRISDERTAIVCPMIDSISDNTLAYHGDWS..
                        LIRA+L GA  + G +I FLD+HCE   GWLEP++ RI  +R  +VCP+ID ISD+T  Y
Rat position 189         LIRARLKGAAVSKGQVITFLDAHCECTVGWLEPLLARIKHDRRTVVCPIIDVISDDTFEYMAGSD..

Human EST:

6.Hu: T11328      68%   GNFDWSLSFGWESLPDHEKQRRKXDETYPEGAPTFAGGLFSISXXYFEYIGSYDEEMEIWGG
                        G F+W L+F W  +P  E  +R+ D T P   PT AGGLFSI   YF+ IG+YD  M+IWGG
Rat position 257         GGFNWKLNFRWYPVPQREMDRRKGDRTLPVRTPTMAGGLFSIDRDYFQEIGTYDAGMDIWGG

7.Hu: R01653      79%   SRLDLRKNLRCQSFKWYLENIYPE
                        SR+ LR  L+C+ F WYLENIYP+
Rat position 398         SRVGLRHKLQCKPFSWYLENIYPD

8.Hu: R01477      71%   KRMKKLGIFNCHGYGGVIRFS
                        K  +K+GIFNCHG GG   FS
Rat position 449         KENEKVGIFNCHGMGGNQVFS
```

Figure 5. A family of potential *C. elegans* and human homologues of the polypeptide GalNAc transferase. Alignments were detected using TBLASTN. Database entries with 21 or more amino acids having greater than 65% homology are listed. The displayed sequence is truncated at 66 amino acids. The homology is calculated from the whole entry. The first amino acid of each EST aligns with the rat polypeptide GalNAc transferase-T1 at the position indicated in the left column.

human and rat airways [13–17] and thus may require different or more than one polypeptide GalNAc transferase isoform. The human MUC4, MUC5, MUC8 and AMN-22 apomucin are rich in threonine, proline, and serine, whereas one of the rat tracheal epithelial mucin-like proteins is rich in threonine but essentially lacks serine and proline. In general, repeated peptide sequence motifs in mucins are generally not conserved among mammals even in the same tissue types [18], and therefore the isozyme(s) responsible for glycosylating O-glycoproteins for a given tissue may differ among mammals. Coordinate transferase expression may be crucial, since some mucins are expressed in more than one organ system [19]. Questions regarding the substrate specificity of each isozyme and number of different isoforms expressed in each cell or tissue type will remain unanswered until functionally expressed cDNAs become available for individual polypeptide GalNAc transferase family members. The mechanisms which underlie the different levels of polypeptide GalNAc transferase-T1 transcript among the tissues surveyed must also be examined; this could include regulation at both the transcriptional and translational level.

Preliminary evidence for newly discovered members of the polypeptide GalNAc transferase family has been reported [20]; distinct polypeptide GalNAc transferase activities were detected in different tissues. GenBank database searches reveal many mammalian expressed sequence tags with up to 76% homology with the rat polypeptide GalNAc transferase-T1 enzyme presented in this study. The sequence variation of the GalNAc transferase isoforms in mammals may indeed be up to 50%, as the mammalian polypeptide GalNAc transferase-T1 and the evolutionary distant *C. elegans* GalNAc transferase homologues ZK688.8 share only 52% identity across the whole amino acid sequence.

GalNAc O-linked to a peptide core exists in organisms from distant phyla, including fungi [21], nematodes [22], amphibians [23] and mammals. The observation of polypeptide GalNAc transferase homologues in *C. elegans* is consistent with the observation that *C. elegans* detergent extracts contain the appropriate enzyme activity and is consistent with a recent report by Gems and Maizel [24] that one predominant class of transcripts encodes mucin-type glycoproteins, the peptide substrate of the enzyme. The function of mucins and O-linked oligosaccharides in nematodes is thought to be similar to that in mammals: to protect the organism from desiccation and to provide a protective mucin blanket lining epithelial surfaces in ducts and tracts – like the oral cavity, gastrointestinal tract, reproductive tract and excretory canal. In addition, the lubricating properties of O-linked mucin glycoproteins may be important for locomotion of worms. O-linked oligosaccharides and glycoproteins are also expressed on the cuticle (outer surface) of nematodes during specific larval stages [22]. Ectopic expression or secretion of mucins may be involved in mucin-mediated immune evasion by some parasitic nematodes [24]. The *C. elegans* model system provides an opportunity to investigate the functional and developmental significance of mucin-type O-linked glycans and to understand the significance of the redundancy in polypeptide GalNAc transferase family members.

Acknowledgement

This work was supported in part by National Institutes of Health grant DE-08108 (to L.A.T.). The nucleotide sequence(s) reported in the paper has been submitted to the Gen Bank™/EMBL Data Bank with accession number U35890.

References

1. Hagen FK, VanWuyckhuyse BC, Tabak LA (1993) *J Biol Chem* **268**: 18960–65.
2. Roth J, Wang Y, Eckhardt AE, Hill RL (1994) *Proc Natl Acad Sci USA* **91**: 8935–39.
3. O'Connell B, Takak LA, Ramasubbu N (1991) *Biochem Biophys Res Commun* **180**: 1024–30.
4. Wilson IBH, Gavel Y, von Heijne G (1991) *Biochem J* **275**: 529–34.
5. Gooley AA, Williams KL (1994) *Glycobiology* **4**: 413–17.
6. Elhammer AP, Poorman RA, Brown E, Maggiora LL, Hoogerheide JG, Kezdy FJ (1993) *J Biol Chem* **268**: 10029–38.
7. Wang Y, Argwal N, Eckhardt AE, Stevens RD, Hill RL (1993) *J Biol Chem* **268**: 22979–83.
8. Wragg S, Hagen FK, Tabak LA (1995) *J Biol Chem* **270**: 16947–54.
9. Homa FL, Hollander T, Lehman DJ, Thomsen DR, Elhammer AP (1993) *J Biol Chem* **268**: 12609–16.
10. Sulston J, Du Z, Thomas K, Wilson R, Hillier L, Staden R, Halloran N, Green P, Thierry-Wieg J, Qiu L, Dear S, Coulson A, Craxton M, Durbin R, Berks M, Metzstein M, Hawkins T, Ainscough R, Waterston R (1992) *Nature* **356**: 37–41.
11. Kozak M (1984) *Nucl Acids Res* **12**: 857–72.
12. Altschul SF, Warren G, Miller W, Myers EW, Lipman DJ (1990) *J Mol Biol* **215**: 403–10.
13. Porchet N, van Cong N, Dufosse J, Audie JP, Guyonnet-Duperat V, Gross MS, Denis C, Degand P, Bernheim A, Aubert JP (1991) *Biochem Biophys Res Commun* **175**: 414–22.
14. Meerzaman D, Charles P, Daskal E, Polymeropoulos MH, Martin BM, Rose MC (1994) *J Biol Chem* **269**: 12932–39.
15. Shankar V, Gilmore MS, Elkins RC, Sachdev GP (1994) *Biochem J* **300**: 295–98.
16. Ohmori H, Dohrman AF, Gallup M, Tsuda T, Kai H, Gum Jr JR, Kim YS, Basbaum CB (1994) *J Biol Chem* **269**: 17833–40.
17. Tsuda T, Gallup M, Jany B, Gum J, Kim Y, Basbaum C

(1993) *Biochem Biophys Res Commun* **195**: 363–73.
18. Verma M, Davidson EA (1994) *Glycoconj J* **11**: 172–79.
19. Gerard C, Eddy RL, Shows TB (1990) *J Clin Invest* **86**: 1921–27.
20. Sorenson T, White T, Wandel HH, Clausen H (1994) In 3rd International Workshop on Carcinoma-Associated Mucins, Imperial Cancer Res, Fund ed., Cambridge, UK. p. 47.
21. Kawaguchi N, Ohmori T, Takeshita Y, Kwanishi G, Katayaman S, Yamada H (1986) *Biochem Biophys Res Commun* **140**: 350–56.
22. Hemmer RM, Donkin SG, Chin KJ, Grenache DG, Bhatt H, Politz SM (1991) *J Cell Biol* **115**: 1237–47.
23. Probst JC, Gertzen EM, Hoffman W (1990) *Biochemistry* **29**: 6240–44.
24. Gems D, Maizel R (1993) *Worm Breeder's Gazette* **13**: 83.