Cloning and Expression of a Novel, Tissue Specifically Expressed Member of the UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferase Family*  

(Received for publication, March 30, 1998, and in revised form July 7, 1998)

Kelly G. Ten Hagen, Fred K. Hagen, Marlene M. Balys, Thomas M. Beres, Brian Van Wyckhuyse, and Lawrence A. Tabak†‡

From the Center for Oral Biology, Rochester Institute of Biomedical Sciences, University of Rochester Medical Center, Rochester, New York 14624

We report the cloning and expression of the fifth member of the mammalian UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase (ppGaNTase) family. Degenerate polymerase chain reaction amplification and hybridization screening of a rat sublingual gland (RSLG) cDNA library were used to identify a novel isoform termed ppGaNTase-T5. Conceptual translation of the cDNA reveals a novel, long tissue-specific region not observed for other members of this enzyme family. Recombinant proteins expressed transiently in COS7 cells displayed transferase activity in vitro. Relative activity and substrate preferences of ppGaNTase-T5 were compared with previously identified isoforms (ppGaNTase-T1, -T3, and -T4); ppGaNTase-T5 and -T4 glycosylated a restricted subset of peptides whereas ppGaNTase-T1 and -T3 glycosylated a broader range of substrates. Northern blot analysis revealed that ppGaNTase-T5 is expressed in a highly tissue-specific manner; abundant expression was seen in the RSLG, with lesser amounts of message in the stomach, small intestine, and colon. Therefore, the pattern of expression of ppGaNTase-T5 is the most restricted of all isoforms examined thus far. The identification of this novel isoform underscores the diversity and complexity of the family of genes controlling O-linked glycosylation.

O-Glycosidically linked oligosaccharides are responsible for the unique physical properties and extended structural conformation of molecules such as mucins and a number of membrane receptors (1). Additionally, O-glycans function as ligands for receptors mediating such diverse functions as sperm-egg adhesion (2) and leukocyte trafficking (3). The initiation of O-linked glycosylation occurs through the action of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase (ppGaNTase), EC 2.4.1.41, which catalyzes the transfer of GalNAc from the nucleotide sugar UDP-GalNAc to the hydroxyl group of either serine or threonine.

Four members of this enzyme family have been cloned and expressed thus far from mammals (ppGaNTase-T1 (4, 5), -T2 (6), -T3 (7, 8), and -T4 (9); one putative isoform has been cloned in mice with no obvious phenotype (10). Five additional isoforms have been identified in Caenorhabditis elegans (11). In vitro assays indicate that subtle differences in substrate preferences exist among the more widely expressed ppGaNTase-T1, -T2, and -T3, although there appears to be overlap (7, 8, 12). In contrast, ppGaNTase-T4 displays limited substrate preferences in vitro and a more restricted pattern of expression (9). Thus, while it remains unclear why so many isoforms of ppGaNTase have evolved, the conservation of this family of enzymes from nematode to humans suggests that there is functional importance to the diversity of the enzymes expressed. In the present study we have cloned and expressed a new form of ppGaNTase, which we designate ppGaNTase-T5. Conceptual translation of the cDNA that encodes this isoform reveals the unusual structural feature of a long putative stem region. In addition, both the pattern of expression and in vitro substrate preferences are more restricted in nature.

EXPERIMENTAL PROCEDURES

Construction of Rat Sublingual Gland cDNA Library—Poly(A)+ RNA was isolated from Wistar RSLGs using the QIAGEN Rapid Poly(A)+ mRNA Isolation System (5 Prime→3 Prime, Inc., Boulder, CO). All mRNA was doubly poly(A)+ selected using two oligo(dT)-cellulose columns. The integrity of the purified mRNA was verified by Northern analysis using multiple RSLG probes (data not shown). Prior to cDNA synthesis, 5 μg of poly(A)+ mRNA was treated with 10 μl CH3HgOH for 1 min to eliminate secondary structure. cDNA synthesis was performed using the Stratagene cDNA Synthesis Kit. cDNA size fractionation, quantitation, and cloning into the Uni-Zap XR vector was performed according to the manufacturer’s instructions as was packaging, amplification, and titering of the resultant Uni-Zap XR library.

Isolation of ppGaNTase-T5 Probes and Full-length cDNAs—The conserved amino acid regions EEWGGN and WWMDEYK were used to design sense and antisense PCR primers, d(GARATHTGGGNNNGARAA) (321-S) and d(CTTATYCRTCCATCCNAC) (379-AS), respectively. These primers were then employed in PCR reactions using RSLG cDNA as the template under conditions described previously (9). The expected 200-bp PCR products were cloned into the SfiI site of an M13 cloning vector and screened for putative ppGaNTase fragments using plaque hybridization with the 32P-end-labeled primers mentioned above. From this screen, 24 positive clones were obtained and sequenced using infrared fluorescent dye-labeled primers on a LI-COR DNA 4000L DNA sequencer. The insert from one unique clone was used to generate a 140-bp PCR fragment using the oligonucleotides 2-6-S (d(AATATGGAGCTATCATTCAAGGTCT)) and 2-6-A (d(AGATTCCGTTAGTTGACT)) and 2-6-A (d(AGATTCGGTCCACCGTCTC)). This 140-bp fragment was labeled by asymmetric trophoresis; HIV, human immunodeficiency virus; Tricine, N, N'-hydroy-1,1-bis(hydroxymethyl)methylglycine.
PCR with oligo 2-6-A as described previously (23) and subsequently used to screen 1 × 10^6 plaques from the oligo(dT)-primed Uni-Zap XR RSFLG cDNA library according to standard procedures (13). Six positive clones were obtained during this screen. One clone was sequenced to entirety on both strands. The N-terminal transmembrane domain was determined by a Kyte-Doolittle hydrophobicity plot. Sequence alignments were performed using the Clustal method of Megalign (DNASTAR) and began at the conserved region FN127750. Numbering of the cDNA begins with the initiation codon.

**Amino Acid Similarity Determinations**—Amino acid sequences were aligned, one pair at a time, using the pairwise ClustalW (1.4) algorithm in MacVector (Oxford Molecular Group). The following alignment parameters were used: slow alignment, open gap penalty = 10, extended gap penalty = 0.1, similarity matrix = blosum, delay divergence = 40%, and no hydrophilic gap penalty. The percent aa similarity displayed in Tables I and II represents the sum of sequence similarity for certain isoforms. The percent identities and similarities. Sequences comprising the conserved domain used in Table II begin with the first aa in Fig. 2 and end in the putative lumenal domain.

**Generation of Secretion Constructs for ppGaNTase-T5**—The full-length 3.5-kb cDNA of ppGaNTase-T5 was cloned into the M13 vector. The secretion construct of ppGaNTase-T5 was generated as described previously (9) using peptides in the concentration range of 5–2500 μM for substrates which gave a measurable enzyme activity.

Enzyme activity was measured against a panel of peptides including: EA2 (PTTDSTTPAPTTK), from the rat submandibular gland mucin repeat (17); HIV (RGPGRAFVTIGKIGNMR), from HIV gp120 mucin (19); MUC1b (PDTRPAPGSTAPPAC), from human Muc1 protein (7); MUC2 (PTTTPISTTTMVTPTPTPTC), from human intestinal mucin (18); and mG-MUC (QTSSPNTGKTSTISTT), from mouse gastric mucin (21). To generate the plasmid pF1-rT5, and 8 μl of LipofectAMINE™ (Life Technologies Inc.) was used to transfect a 35-mm well of COS7 cells as described previously (9). Reombinant enzymes were purified from the culture media of transfected cells using anti-FLAG™ M2 affinity gel (Eastman Kodak Co.) as described previously (16). Gels were dried under vacuum and exposed to film (XAR-Kodak) or quantitated on a PhosphorImager (Molecular Dynamics). A dilution series of certain isoforms was labeled and separated by Tricine/SDS-PAGE to verify the accuracy and linearity of gel quantitation.

**Kinetic Analyses**—Kinetic constants for the ppGaNTases were determined using anion exchange chromatography (AG 1X-8, Bio-Rad). Reaction products were analyzed by Tricine/SDS-PAGE (15) after labeling with [γ-^32P]ATP using heart muscle kinase (HMK) as described previously (16). Densiometric measurements. All assays were performed in duplicate at 37 °C for 1 h as described previously (9). Reaction products were purified using anion exchange chromatography (AG 1X-8, Bio-Rad). Reactions lacking peptide yielded a background value which was averaged for each isoform and subtracted from each duplicate value. Adjusted duplicate measurements were then averaged to give the values shown in Fig. 4. Enzyme activity is expressed as cpm/h/densitometric unit of gel.
1–250 μM. In an effort to ensure that less than 10% of the substrate UDP-[14C]GalNAc was consumed in each assay, amounts of each enzyme used in these assays were not normalized relative to one another due to substantial differences in activity between isoforms. Assays to determine kinetic parameters of UDP-GalNAc were performed with saturating concentrations of EA2 (500 μM), and assays to determine parameters of EA2 and mG-MUC were performed with 500 μM UDP-GalNAc. All assays were performed in duplicate. 

*Km* and *V*max values were estimated using Lineweaver-Burk and Hanes plots.

**Northern Blot Analysis**—Total RNA was extracted from Wistar rat tissues according to the single step isolation method described in Ausubel et al. (22). Following electrophoresis in a 1% formaldehyde agarose gel, rat total RNA samples were transferred to Hybond-N membranes (Amersham Pharmacia Biotech) according to Sambrook et al. (13). pp-GaNTase-T5 transcripts were detected by asymmetric PCR labeling (23) of the 1-kb MscI/NheI fragment from the ppGaNTase-T5 cDNA.

![Fig. 2](image)

**Fig. 2** Amino acid sequence alignments of ppGaNTase-T1, T2, T3, T4, and T5 from human, murine, and rat clones. A consensus sequence is shown below alignment blocks for positions that are 60% conserved among isoforms shown (only one of the three T1 isoforms was used in this determination). Alignments began at the conserved region, FNXSD, in the putative luminal domain of the enzymes. Segments of amino acid sequences that were used to make hybridization probes or PCR primers are boxed and underlined with a bold arrow.

| Table I | Amino acid similarity between ppGaNTase-T5 and other ppGaNTase isoforms |
|---------|---------------------------------------------------------------------|
| %       |                                                     |
| rT1     | 57                                                   |
| hT2     | 53                                                   |
| mT3     | 52                                                   |
| mT4     | 54                                                   |

| Table II | Amino acid similarity between ppGaNTase isoforms within a 340-aa conserved domain |
|----------|-----------------------------------------------------------------------------------|
| %        |                                                                                   |
| hT1      | NA 98 99 70 71 70 68                                                             |
| mT1      | 98 99 70 71 70 68                                                                 |
| rT1      | 99 99 70 70 70 68                                                                 |
| mT3      | 71 71 71 66 68 66                                                               |
| mT4      | 70 70 70 68 69 69                                                              |
| rT5      | 68 68 68 65 64 65                                                             |

1–250 μM. In an effort to ensure that less than 10% of the substrate UDP-[14C]GalNAc was consumed in each assay, amounts of each enzyme used in these assays were not normalized relative to one another due to substantial differences in activity between isoforms. Assays to determine kinetic parameters of UDP-GalNAc were performed with saturating concentrations of EA2 (500 μM), and assays to determine parameters of EA2 and m-G-MUC were performed with 500 μM UDP-GalNAc. All assays were performed in duplicate. *Km* and *V*max values were estimated using Lineweaver-Burk and Hanes plots.

**Northern Blot Analysis**—Total RNA was extracted from Wistar rat
clone using the antisense oligonucleotide, d(CAGAATGTGGTTGGGTT- TAG). ppGaNTase-T1 and -T4 were detected as described previously (9). The antisense oligonucleotide for the 18 S ribosomal subunit (dTATTGGAGCTGGAATTACCGCGGCTGCTGG) was 32P-end-labeled as described (13) and hybridized in 5 molar excess to the Northern blots, to normalize for sample loading. All hybridizations were performed in 5 × SSPE, 50% formamide at 42 °C with two final washes in 2 × SSC, 0.1% SDS at 65 °C for 20 min.

**RESULTS**

**cDNA Cloning of ppGaNTase-T5 and Sequence Analysis—**

PCR amplification was performed on RSLG cDNA using degenerate primers designed from within a 420-amino acid conserved region of the ppGaNTase family of proteins. A novel PCR product was identified that shared homology with previously characterized isoforms and was subsequently used as a probe to screen a RSLG cDNA library. One cDNA clone containing a complete open reading frame was isolated, sequenced, and given the designation, ppGaNTase-T5 (GenBank™/EBI Data Bank accession number AF049344). The cDNA that encodes ppGaNTase-T5 contains a 3519-bp insert (Fig. 1) which can be conceptually translated into a 930-aa type II membrane protein consisting of a 13-aa N-terminal cytoplasmic region, a 24-aa hydrophobic region, a 416-aa stem region, and a 477-aa luminal region. The 416-aa stem region is 4–7 times longer than those of other isoforms. As shown in Fig. 2, ppGaNTase-T5 is distinct from previously identified mammalian isoforms yet shares many blocks of sequence similarity or identity between consensus aa 454 and 930. Amino acid similarity comparisons were performed between rat ppGaNTase-T5 and other known isoforms (Tables I and II). Similarity values range between 52 and 57% when looking at the full-length proteins (Table I) and...
increase to 64–68% when comparisons are performed within an approximately 340-aa conserved domain (Table II).

During this screen, a partial cDNA clone was isolated that shared complete homology with ppGaNTase-T5 except for an additional 100-bp region inserted at nucleotide position 2410. The conceptually translated sequence of this region revealed a stop codon that would yield a 3’ truncation of the protein. Based on preliminary sequence data and Southern analysis, this clone may represent a splice variant of ppGaNTase-T5 (data not shown).

**Functional Expression**—The coding region of ppGaNTase-T5 was cloned downstream of the insulin secretion signal, HMK site, and FLAG™ epitope tag of the vector pIKMF1 (9). The ppGaNTase-T5 expression vector, as well as similar constructs containing the previously identified ppGaNTase-T1, -T3, and -T4 cDNAs, were transfected into COS7 cells, and the expressed products were enriched by immunoadsorption to an anti-FLAG affinity matrix. Enriched material for each isoform was γ-32P-labeled and resolved by Tricine/SDS-PAGE (Fig. 3). Equivalent amounts of each partially purified isoform (as judged by densitometric scanning of Tricine/SDS-PAGE gels) were used to assess substrate preferences in vitro (Fig. 4). Both ppGaNTase-T4 and -T5 showed restricted preferences; ppGaNTase-T5 demonstrated significant transferase activity against substrate EA2, with lower levels of glycosylation seen with the rMUC-2, mG-MUC, and EPO-T peptides. Furthermore, mixing experiments between ppGaNTase-T4 and -T5 resulted in an additive effect, demonstrating that these recombinant preparations do not contain differing levels of transferase inhibitors (data not shown). In contrast, ppGaNTase-T1 and -T3 were more promiscuous and incorporated sugar in virtually all peptides tested, albeit at varying levels. When we attempted to express the putative splice variant described above, neither secreted protein nor enzymatic activity could be detected.

Kinetic parameters for ppGaNTase-T1, -T3, -T4, and -T5 were determined for peptide substrates, EA2 and mG-MUC. Amounts of enzymes used in this analysis were not normalized relative to one another due to substantial differences in activity between isoforms. \( K_m \) and \( V_{max} \) values obtained are shown in Table III. All four enzymes displayed similar \( K_m \) values for UDP-GalNAc, with ppGaNTase-T4 and -T5 having the highest values (0.069 mM and 0.051 mM, respectively). All enzymes, with the exception of ppGaNTase-T4, seemed to prefer EA2 as a substrate over mG-MUC. \( K_m \) values for ppGaNTase-T5 with the other peptides shown in Fig. 4 were not measurable.

**Northern Blot Analysis**—Northern blots of rat total RNA from multiple tissues were hybridized with specific probes for ppGaNTase-T1, -T4, and -T5 (Fig. 5). In contrast to 3.7- and 4.3-kb ppGaNTase-T1 transcripts, which were present in all samples examined, the expression of the 4.4-kb ppGaNTase-T5 message was restricted primarily to the SLG, with lower levels seen in stomach, small intestine, and colon. A minor transcript band was also detected at ~8 kb. The distribution of the 4.0-kb and 2.8-kb ppGaNTase-T4 transcripts was similar to ppGaNTase-T5, detected principally in the SLG, digestive tract, and additionally, in the female reproductive tract.

**DISCUSSION**

In an effort to further define the repertoire of ppGaNTases, we have employed degenerate PCR to clone a novel isofrom from a rat cDNA library. ppGaNTase-T5 has the attributes of a type II membrane protein, characteristic of previously identified glycosyltransferases (Fig. 6). The N-terminal, hydrophobic, and catalytic regions of ppGaNTase-T5 are similar in length to ppGaNTase-T1, -T2, -T3, and T4. However, ppGaNTase-T5 has an unusually large stem region (416 aa), which is ~4–7 times longer than other known isoforms. Many type II membrane proteins are converted into soluble, circulating species after cleavage within the stem region by a group of enzymes referred to as "secretases" (24). Of interest is the large number of hydroxyamino acids and potential N-glycosylation sites found in the ppGaNTase-T5 stem region. The presence of O- or N-glycans in this region could inhibit the action of secretases, thereby keeping this isoform resident within the Golgi complex. Alternatively, a heavily O-glycosylated stem may function as a extended stalk (1), inserting the catalytic region of this enzyme further into the Golgi lumen.

The aa conservation of an individual ppGaNTase isoform is greater than 90% across mammalian species, whereas different isoforms within the same species show more variability. ppGaNTase-T5 demonstrates that highest degree of similarity to ppGaNTase-T1; they share 68% aa similarity within the conserved region and 57% aa similarity across the full-length protein (Tables I and II). The reduced similarity observed when comparing full-length proteins is most likely the result of di-
minished conservation within the N terminus and the long stem region of ppGaNTase-T5. Previously identified isoforms (ppGaNTase-T1, -T2, -T3, and -T4) have regions of highly conserved or invariant sequences within the luminal domain. Although ppGaNTase-T5 contains most of these conserved regions, it demonstrates the lowest aa similarity to other isoforms and has a very poorly conserved CHGXGGNQ block (which has been used in the past to identify novel isoforms) (9). This result points out the limitation of using a restricted set of conserved sequences to identify new isoforms; the clones recovered by such methods may be biased against the population of transferases lacking these blocks. When taken together with the results of a site-directed mutagenesis study demonstrating that the histidine residue in this sequence is non-essential (25), it further suggests that this conserved block per se may not play a significant functional role in the transferase activity of these enzymes.

Highly tissue specifically regulated isoforms of ppGaNTases may be uniquely programmed to act only on a limited subset of substrates from a particular cell or tissue type. This may account for the relatively modest activities seen for ppGaNTase-T5 and -T4 against our limited panel of peptides. Thus, it is plausible that ppGaNTase-T5 acts on a substrate unique to the RSLG. Future studies will focus on the identification of substrates whose expression patterns mimic those of a given isoform of ppGaNTase-T5. The highly restricted expression and unique structural features of ppGaNTase-T5 further emphasize the diversity and complexity of this family of enzymes. It is unclear why so many distinct isoforms are needed in vivo. It remains possible that some isoforms serve more general, housekeeping functions while others act on specific molecules only. However, an extensive network of transferases may be in place with enough redundancy to ensure appropriate glycosylation. Through the continued identification of novel isoforms and gene ablation studies on the diverse repertoire of members available currently, we can begin to dissect the biological significance of this family of enzymes.

Acknowledgment—We thank Meng Qian for her help in preparing this manuscript.

REFERENCES
1. Jenetoff, N. (1990) Trends Biochem. Sci. 15, 291–294
2. Kinloch, R. A., Sakai, Y. & Wasserman, P. M. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 263–267
3. Dowbenko, D., Andalibi, A., Young, P. E., Lusis, A. J. & Lasky, L. A. (1993) J. Biol. Chem. 268, 4525–4529
4. Hagen, F. K., Van Wuyckhuyse, B. & Tabak, L. A. (1993) J. Biol. Chem. 268, 18960–18965
5. Hom, F. L., Hollander, T., Lehman, D. J., Thomsen, D. R. & Elhammer, A. P. (1993) J. Biol. Chem. 268, 12669–12676
6. Sørensen, T., White, T., Wendall, H. H., Kristense, A. K., Roepstorff, P. & Clausen, H. (1995) J. Biol. Chem. 270, 24166–24173
7. Bennett, E. P., Hassan, H. & Clausen, H. (1996) J. Biol. Chem. 271, 17006–17012
8. Zara, J., Hagen, F. K., Ten Hagen, K. G., VanWuyckhuyse, B. C. & Tabak, L. A. (1996) Biochem. Biophys. Res. Commun. 228, 38–44
9. Hagen, F. K., Ten Hagen, K. G., Beres, T. M., Balya, M. M., VanWuyckhuyse, B. C. & Tabak, L. A. (1997) J. Biol. Chem. 272, 13843–13848
10. Hennett, T., Hagen, F. K., Tabak, L. A. & Marth, J. D. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 12070–12074
11. Hagen, F. K. & Nehrke, K. (1998) J. Biol. Chem. 273, 8268–8277
12. Wandall, H. H., Hasson, H., Mørgodskaya, E., Kristensens, A. K., Roepstorff, P., Bennett, E. P., Nielsen, P. A., Hollingworth, M. A., Burchell, J., Taylor-Papadimitrion, J. & Clausen, H. (1997) J. Biol. Chem. 272, 23503–23514
13. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
14. Kunkle, T. A. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 488–492
15. Schagger, H. & von Jagow, G. (1987) Anal. Biochem. 166, 368–378
16. Nehrke, K., Hagen, F. K. & Tabak, L. A. (1996) J. Biol. Chem. 271, 7061–7065
17. Albone, E. F., Hagen, F. K., VanWuyckhuyse, B. C. & Tabak, L. A. (1994) J. Biol. Chem. 269, 16945–16952
18. Gum, J. R., Jr., Hickes, J. W., Toribara, N. W., Siddikhi, B. & Kim, Y. S. (1994) J. Biol. Chem. 269, 2440–2446
19. Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Dubrig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E. N. & Wilson, D. (1999) J. Biol. Chem. 265, 15286–15293
20. Ohmori, H., Dohrman, A. F., Gallup, M., Tsuda, T., Kai, H., Gum, J. R., Jr., Kim, Y. S. & Basham, C. B. (1994) J. Biol. Chem. 269, 17833–17840
21. Shekels, L. L., Lyftogt, C., Kieliszewski, M., Filie, J. D., Kozak, C. A. & Ho, S. B. (1995) Biochem. J. 311, 775–785
22. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (eds) (1995) Current Protocols in Molecular Biology, pp. 4.2.4–4.2.8, Wiley, New York
23. Bednarzuk, T. A., Wiggins, R. C. & Konat, G. W. (1991) BioTechniques 10, 478–479
24. Hooper, N. M., Karran, E. H. & Turner, A. J. (1997) Biochem. J. 321, 265–279
25. Wragg, S., Hagen, F. K. & Tabak, L. A. (1997) Biochem. J. 328, 193–197