Mucin biosynthesis: Molecular cloning and expression of mouse mucus-type core 2 β1,6 N-acetylgalactosaminyltransferase

Mitsuyoshi Hashimoto²,³, Shuhua Tan³, Naoyoshi Mori², Helen Cheng³, and Pi-Wan Cheng¹,³,⁴

¹Department of Pathology, Nagoya University Graduate School of Medicine, Nagoya, Japan; ²Department of Biochemistry and Molecular Biology, College of Medicine, University of Nebraska Medical Center and ³Eppley Cancer Center for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE 68198-5870, USA

Received on April 27, 2007; revised on June 15, 2007; accepted on June 17, 2007

Secrected mucins protect the underlying epithelium by serving as the major determinant of the rheological property of mucus secretion and the receptors for pathogens. These functions can be affected by the three branch structures, including core 2, core 4, and blood group I, which are synthesized by the mucus-type core 2 β1,6 N-acetylgalactosaminyltransferase (C2GnT-M). Decreased activity of this enzyme and expression of this gene have been found in colorectal cancer, which supports the important role of this enzyme in the protective functions of secreted mucins. We cloned full-length mouse (m) C2GnT-M cDNAs and showed that the deduced amino acid sequence was homologous to those of other C2GnT-Ms. The recombinant protein generated by mC2GnT-M cDNA exhibited core 2, core 4, and blood group I enzyme activities with a ratio of 1.00:0.46:1.05. We identified two different size transcripts by rapid amplification of cDNA ends and RT-PCR. Derived from the 6.6 kb mC2GnT-M gene composed of three exons and two introns, these two transcripts were intronless and resulted from the 6.6 kb mC2GnT-M gene composed of three exons.

Keywords: C2GnT-M gene /C2GnT-M gene expression/mucin biosynthesis/mucus cell IHC staining/mucin glycan branch enzyme

Introduction

Mucins are large molecular weight glycoproteins found in mucous secretions or tethered on the cell surface (Hollingsworth and Swanson 2004; Rose and Vojnov 2006). Carbohydrate constitutes 60–90% of the molecular mass of mucins. For secreted mucins the conjugated carbohydrate not only contributes to the rheological properties of mucus, which covers the epithelial surface, but also serves as ligands to trap and remove airborne or injected pathogens. For mucins tethered on the cell surface, the conjugated carbohydrates are responsible for cell–cell recognitions affecting inflammatory processes and differentiation, maturation, and activation of T and B cells (Tsuboi and Fukuda 2001), as well as cancer metastasis (Kannagi 1997). Since the alteration of mucin carbohydrate structures can affect the biological functions of mucins under normal and diseased conditions, it is critically important to understand how mucin carbohydrates are synthesized and how their synthesis is regulated.

Mucin carbohydrate is assembled in the Golgi apparatus following the synthesis of mucin polypeptide in the endoplasmic reticulum. Synthesis of mucin carbohydrate is template-independent, a process different from the synthesis of DNA/RNA and proteins. The sugar is added one at a time as immature mucins travel from cis-Golgi to trans-Golgi network. This type of synthetic mechanism leads to the production of mucins with heterogeneous carbohydrate structures, a hallmark of mucin properties. Synthesis of mucin glycan is initiated by the transfer of GalNAc from UDP-GalNAc to serine or threonine in the protein backbone as catalyzed by a group of polypeptidyl GalNAc transferases (Ten Hagen et al. 2003). The glycosyltransferases involved in subsequent glycosylation steps can be classified as chain termination and elongation enzymes (Cheng and Bona 1982; Schachter et al. 1989). The chain termination enzymes can either terminate or limit carbohydrate chain extension. They include fuscosyltransferases, sialyltransferases, blood group A/B, secretor enzymes, etc. The chain elongation enzymes promote continued synthesis of carbohydrate chains. These enzymes include β-N-acetylgalactosaminyltransferases, β-galactosyltransferases, etc. Competition of chain termination glycosyltransferases against chain elongation glycosyltransferases contributes to the heterogeneity of mucin carbohydrate structures (Cheng and Bona 1982; Schachter et al. 1989).

Among the mucin glycan elongation enzymes, β GalNAc transferases are unique because they can form branch structures to expand the number of terminal sugar chains, building the complexity and functional potential of mucins. The three major branch structures found in mucin glycan are core 2 (Galβ3(GlcNAcβ6)GalNAc), core 4 (GlcNAcβ3(GlcNAcβ6)GalNAc), and blood group I (GlcNAcβ3(GlcNAcβ6)Gal) (Beum and Cheng 2001; Inaba et al. 2003). Core 2 structure is synthesized from core 1 (Galβ3GalNAc) by catalyzing by three core 2 GlcNAc transferases (C2GnT), which include leukocyte-type C2GnT-L (or 1) (Bierhuizen et al. 1992; Li et al. 1998), mucus-type C2GnT-M (or 2) (SchwieteK et al. 1999; Yeh et al. 1999; Vandenplasschen et al. 2000; Korekane et al. 2003; Choi et al. 2004).

\[
\text{UDP-GlcNAc + Galβ3GlcNAc2Ser/Thr} \rightarrow \text{Galβ3(GlcNAcβ6)GlcNAc2Ser/Thr + UDP}
\]

2. C4GnT activity (conversion of core 3 to core 4): C2GnT-M

\[
\text{UDP-GlcNAc + GlcNAcβ3GlcNAcSer/Thr} \rightarrow \text{GlcNAcβ3(GlcNAcβ6)GlcNAcSer/Thr + UDP}
\]

3. c/dIGnT activity (conversion of blood group i to blood group 1): cGnT or C2GnT-M

\[
\text{UDP-GlcNAc + (±Galβ4)GlcNAcβ3Galβ} \rightarrow (±\text{Galβ4})\text{GlcNAcβ3(GlcNAcβ6)Galβ} + \text{UDP}
\]

Since our previous report (Ropp et al. 1991) describing purification and demonstration of multi-functional properties of a bovine β1,6-N-acetylgalcosaminyltransferase from tracheal epithelial tissue, cDNAs coding a recombinant protein exhibiting similar enzymatic specificity have been cloned from human (Schwientek et al. 1999; Yeh et al. 1999), bovine herpes virus type 4 (Vanderplasch et al. 2000), bovine (Choi et al. 2004), and rat (Korekane et al. 2003). In addition, we showed that expression of C2GnT-M gene was subject to regulation. Th2 cytokines and retinoic acid upregulate the expression of C2GnT-M gene (Beum et al. 2005) while EGF downregulates its expression (Beum et al. 2003). Downregulation of C2GnT-M results in shifting mucin glycan structures from core 2 to core 1. Also, C2GnT-M gene expression is cyclically regulated in the hamster oviduct during estrus cycle (McBride et al. 2005). Furthermore, this enzyme has been shown to be downregulated in colorectal cancer and its restoration causes growth inhibition of these cancer cells (Brockhausen 2006; Huang et al. 2006). Although the biological importance of mucin glycans is well recognized and expression of C2GnT-M gene can be regulated (Beum et al. 2003, 2005), very little is known about the cells, which express this enzyme. In this communication, we report the cloning of mouse (m) C2GnT-M cDNA, and the characterization of its genomic structure and mucus cell-specific expression of this enzyme.

Results

Cloning and characterization of mC2GnT-M cDNA

The nucleotide sequence of the longest mC2GnT-M transcript (4301 bp) is shown in Figure 1. The amino acid sequence deduced from the ORF of mC2GnT-M cDNA contained 437 amino acid residues (50.7 kDa molecular weight). It displayed a typical type II transmembrane protein composed of a short N-terminal cytoplasmic domain, a transmembrane region, a short stem, and a long catalytic domain. The domain structure is characteristic of Golgi glycosyltransferases (Paulson and Colley 1989). There were two consensus N-glycosylation sites at N-69 and N-288. It also contained nine cysteines (C70, C110, C161, C182, C209, C227, C381, C390, C422) conserved among C2GnT family and cIGnT (Beum and Cheng 2001). The cDNA sequence has been submitted to GenBank (accession EF202835).

To determine if the protein encoded by the cloned cDNA displayed the characteristic multi-acceptor specificity of C2GnT-M (Ropp et al. 1991; Beum and Cheng 2001; Choi et al. 2004), the recombinant protein secreted into the conditioned medium by CHO cells stably transfected with pSecTag-mC2GnT-M181 was measured for core 2 (C2GnT), core 4 (C4GnT), and distally acting blood group I (dIGnT) activities. As shown in Figure 2, the recombinant mC2GnT-M protein exhibited all three enzyme activities with a ratio of 1.00/0.46/1.05 for C2GnT/C4GnT/dIGnT.

Characterization of mC2GnT-M transcripts and genomic structure

mC2GnT-M transcript as shown in Figure 1 was identified in colon tissues by RACE and RT-PCR (Figure 3). 5′ RACE with antisense gene-specific primers (GSP) located at the 5′ end of the ORF yielded a 605 bp DNA fragment (-508 to 97). Additional 5′ RACE using another antisense GSP located at the 5′ UTR upstream of the other GSP generated many clones with the same 5′ ends (-508 to -464) and several clones with four more nucleotides (-512 to -464). 3′ RACE yielded two different cDNA fragments, including one of 561 bp (1180–1740) and another one of 2610 bp (1180–3789). RT-PCR using a forward primer located at the 5′ end of 5′ UTR and a reverse primer in the middle portion of the long 3′ UTR yielded a 2.6 kb DNA product (-485 to 2115), indicating that the short 3′ RACE product was a part of the long 3′ RACE product. The results suggested that the lengths of the long and short mC2GnT-M transcripts were about 4.3 kb and 2.25 kb, respectively (Figure 3).

To characterize the genomic structure and intron-exon utilization of mC2GnT-M gene, the nucleotide sequence of the long transcript was aligned with the mouse genomic DNA sequence in GenBank. As a result, a 6593 bp mC2GnT-M gene containing three exons and two introns was identified (Figure 4). Sizes of various exons and introns of the mC2GnT-M gene were: exon 1, 116 bp; intron 1, 998 bp; exon 2, 342 bp; intron 2, 1294 bp; and exon 3 for the long transcript, 3843 bp. The size of exon 3 for the short transcript was 1794 bp. Further, exon 3 was composed of 54 bp 5′ UTR, 1311 bp ORF, and 429 bp 3′ UTR for the short transcript and 2478 bp 3′ UTR for the long transcript. The sizes of these two transcripts were confirmed by northern blot analysis (Figure 4). Probes A and B detected both long and short transcripts while probe C detected only the long transcript.

Expression of mC2GnT-M gene in various tissues

Analysis of mC2GnT-M gene expression in various tissues by northern blotting showed the highest expression in the colon, which was followed closely by the stomach and distantly by the small intestine (Figure 5A). Expression of this gene was also detected in testis although the sizes of the transcripts were different from those found in the gastrointestinal tract. The mC2GnT-M transcripts were not detectable in other tissues, including brain, heart, lung, liver, spleen, kidney, esophagus, thymus, and submandibular glands.
M Hashimoto et al.

Fig. 1. Nucleotide sequences of mC2GnT-M cDNAs and deduced amino acid sequences of mC2GnT-M ORF. The cDNAs were cloned from colon tissue by 5′ and 3′ RACEs and RT-PCR. The 5′ UTR of the cDNAs included exon 1, exon 2 and parts of exon 3 with the 5′ end junctions of exons 1 and 2, and exons 2 and 3 indicated by ▼. The 3′ UTRs have two different 3′ ends, which include one indicated by †. Variant polyadenylation signal CATAAA (in nucleotides 1722–1727) for the short transcript and polyadenylation signal AATAAA (in nucleotides 3768–3773) for the long transcript is located 13 and 16 nucleotides upstream the poly(A) tail, respectively. The transmembrane domain is underlined by a double line. The consensus N-glycosylation sites are indicated by * and the nine cysteines conserved among all βGlcNAc transferases by ^. The locations of primers used for the construction of p3xFLAG-C2GnT-M expression vector are underlined. The previously reported mC2GnT-M cDNA (1841 bp) covered −74 to +1767 bp, and +118 bp was assigned as the translation start site. The updated cDNA sequence is accessible in GenBank with an accession number of EF202835.

The results of 5′ RACE and RT-PCR experiments suggested possible heterogeneous 5′ UTR of mC2GnT-M transcripts, which prompted us to examine this phenomenon in detail by RT-PCR using primers located at the 5′ ends of both exon 1 (sense primer: 5′-TTCTCACTTTCCAGGTGCT-3′) and exon 3 (antisense primer: 5′-CCTCTGGCCAAGGATCTCCT-3′). As shown in Figure 5B, three RT-PCR products of 503, 347, and 277 bp were generated from stomach, small intestine, and colon. These three products contained the same exon 1 (89 bp) and exon 3 (72 bp) fragments but different exon 2. The two large size exon 2 products had the same 3′ end while the longest and the shortest products had the same 5′ end. The 3′ end of exon 1 was linked to GT sequence in intron 1, and all three exon 2s were flanked by AG and GT splicing signature sequences located at the 3′ end of intron 1 and 5′ end of intron 2, respectively.

Detection of mC2GnT-M protein by immunohistochemical staining

Anti-mC2GnT-M antibodies generated with a 19-amino acid peptide at the C-terminus of mC2GnT-M were used to perform western blot analysis of recombinant and endogenous mC2GnT-M, and immunohistochemical staining of mucus-secretory tissues. Both anti-mC2GnT-M and anti-FLAG antibodies detected the recombinant protein generated by transient transfection of Panc1 cells with p3xFLAG-mC2GnT-M (Figure 6A). The size (51 kDa) of the protein band as estimated from the western blot matched that calculated from the amino acid sequence of the protein.
mC2GnT-M cDNA and expression in mucus cells

Fig. 2. Enzyme activities of recombinant mC2GnT-M. The conditioned medium of pSecTag-mC2GnT181 stably transfected CHO cells was concentrated 30-fold with a 30 kDa molecular weight cutoff membrane after the cells had been treated overnight with 2 mM sodium butyrate, and then assayed for C2GnT, C4GnT, and dIGnT activities. The ratio of these three activities was C2GnT:C4GnT:dIGnT = 1.0:0.46:1.05 (n = 3).

Fig. 3. Genomic structure of the mC2GnT-M gene and the splicing events for generation of the transcripts. Exons are denoted by boxes in proportion to their sizes and numbered. The location of open reading frame (ORF) is shaded. The exons and introns were identified by alignment of the nucleotide sequence of the longest transcript identified with the mouse genomic DNA sequence. The sizes of these two transcripts were confirmed by 3′ RACE and RT-PCR. mC2GnT-M gene contained exon 1 (116 bp), intron 1 (998 bp), exon 2 (342 bp), intron 2 (1294 bp), and exon 3 (1794 bp and 3843 bp). Both exon 3 contained 5′ UTR, 1311 bp ORF, and 429 bp 3′ UTR, but the longer exon 3 contained 2049 bp extra sequence.

cDNA insert. The anti-mC2GnT-M antibodies also detected endogenous mC2GnT-M protein in colon tissue (Figure 6B). The size of the protein band was close to that (50.7 kDa) calculated from the amino acid composition of the full-length mC2GnT-M cDNA.

Although this mC2GnT-M gene was highly expressed in mucus-secreting tissues and assumed to be expressed in mucus cells, this assumption has never been validated. To address this question, we performed immunohistochemical staining of these mucus secretory tissues using affinity-purified polyclonal antibodies against mC2GnT-M. In the colon, a strong immunohistochemical staining was detected in goblet cells at the base of the crypts (Figure 7A). Only selected goblet cells near the apex were stained. Although all goblet cells were positive with PAS staining, those near the base exhibited stronger immunohistochemical stain than those at the apex of the crypts. Neither PAS nor mC2GnT-M staining was observed in the other epithelial and interstitial cells. Similar to the colon, the small intestine displayed a uniform and strong immunostaining in goblet cells (Figure 7B). All goblet cells were stained with both anti-mC2GnT-M antibodies and PAS, and there was no distinct gradient of staining intensity along the crypt-villous axis in the small intestine. Interestingly, moderate immunoreactivity was observed in Paneth cells. Further, mC2GnT-M protein was highly expressed in duodenal Brunner’s glands (Figure 7C). No immunoreactivity was detected in other epithelial, interstitial...
Fig. 5. (A) Northern blot analysis of mC2GnT-M in various mouse tissues. High-level expression was detected in colon and stomach, and moderate level expression in small intestine. Very low-level expression was observed in testis although the band sizes were different from those of the others. The membrane was hybridized with the $^{32}$P-labeled probe B used in Figure 4. The migration positions of RNA size markers are indicated on the left. (B) RT-PCR analysis of exon 2 utilization in gastrointestinal tissues. Forward primer located at 5′ end of exon 1 and reverse primer located at 5′ end of exon 3 were used to perform RT-PCR in three gastrointestinal tissues, including colon, small intestines, and stomach. The products contain an 87 bp exon 1, three different sizes of exon 2 fragments, including A, 342 bp; B, 186 bp; and C, 116 bp, and 72 bp exon 3.

cells, and white blood cells. In the stomach, strong immunostaining with anti-mC2GnT-M antibodies was detected in pyloric glands (Figure 7D) and mucous neck cells (Figure 7E), and to a lesser degree in foveolar epithelial cells (Figure 7D and E). The positive signals in the foveolar epithelium were weak, however, distinct gradient in the staining intensity was observed. Similar to the colon, a stronger expression of C2GnT-M in foveolar epithelium was observed at the base of pit corresponding to the neck region of gastric mucosa. The expression of C2GnT-M in the surface foveolar epithelial cells was faint. No immunoreactivity was observed in the interstitial cells and other epithelial cells such as chief and parietal cells. We observed the same immunohistochemical staining pattern using frozen sections fixed with 4% paraformaldehyde (data not shown). In all immunohistochemical staining experiments, none of the negative control sections showed positive signals (data not shown). The results of immunohistochemical staining of gastrointestinal epithelia are summarized in Table I.

### Discussion

In this report, we describe the cloning of cDNAs that encode mC2GnT-M, characterization of the recombinant protein, transcripts, and genomic structure of this gene, and show its

<table>
<thead>
<tr>
<th>Tissues</th>
<th>Staining intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stomach (fundus)</td>
<td></td>
</tr>
<tr>
<td>Foveolar epithelium</td>
<td>+a</td>
</tr>
<tr>
<td>Mucous neck cells</td>
<td>+++</td>
</tr>
<tr>
<td>Chief and parietal cells</td>
<td>-</td>
</tr>
<tr>
<td>Stomach (antrum)</td>
<td></td>
</tr>
<tr>
<td>Foveolar epithelium</td>
<td>++a</td>
</tr>
<tr>
<td>Pyloric gland cells</td>
<td>+++</td>
</tr>
<tr>
<td>Duodenum</td>
<td></td>
</tr>
<tr>
<td>Surface epithelium</td>
<td></td>
</tr>
<tr>
<td>Goblet cells</td>
<td>+++</td>
</tr>
<tr>
<td>Nongoblet cells</td>
<td>-</td>
</tr>
<tr>
<td>Brunner’s gland cells</td>
<td>+++</td>
</tr>
<tr>
<td>Jejunum and ileum</td>
<td></td>
</tr>
<tr>
<td>Surface epithelium</td>
<td></td>
</tr>
<tr>
<td>Goblet cells</td>
<td>+</td>
</tr>
<tr>
<td>Nongoblet cells</td>
<td>-</td>
</tr>
<tr>
<td>Paneth cells</td>
<td>++</td>
</tr>
<tr>
<td>Colon</td>
<td></td>
</tr>
<tr>
<td>Surface epithelium</td>
<td></td>
</tr>
<tr>
<td>Goblet cells</td>
<td>+</td>
</tr>
<tr>
<td>Nongoblet cells</td>
<td>-</td>
</tr>
</tbody>
</table>

Staining intensity: -, negative; +, weak; ++, moderate; ++++, strong. 
aBasal side of the gastric and colonic mucosa.
mC2GnT-M cDNA and expression in mucus cells

Fig. 6. Western blot analysis. (A) Recombinant mC2GnT-M protein. Anti-mC2GnT-M and anti-FLAG M2 antibodies were used to detect FLAG-tagged recombinant mC2GnT-M secreted into the conditioned medium. (B) Endogenous mC2GnT-M protein. Anti-mC2GnT-M antibody was used to detect endogenous mC2GnT-M protein in colon tissue.

The physiological significance of expressing intronless transcripts for mC2GnT-M gene is not clear although this type of transcript is considered more favorable for translation when compared to transcripts containing long intron(s) at the 5′ UTR according to the ribosome scanning mechanism (Kozak 1996, 2000, 2002). Another difference between bovine and mouse C2GnT-M transcripts is that the most abundant mC2GnT-M transcripts are the long transcripts, which are 2049 bp longer at the 3′ UTR than the short transcripts. Increased stability of the long transcripts may be one explanation for the observation. As has been previously reported for liver Galβ1,4GlcNAcα2,6-sialyltransferase mRNA, the half-life of this transcript can be extended by a 3′ UTR-specific binding protein (Beaudoing et al. 2000).

The expression of mouse and human C2GnT-M transcripts is not tissue-specific. The expression pattern of C2GnT-M in the mucus cells of gastrointestinal epithelia. The assignment of the cloned cDNA as mC2GnT-M cDNA was based on the high homology of the deduced amino acid sequence to those of other C2GnT-Ms (Beum and Cheng 2001; Korekane et al. 2003; Choi et al. 2004) and the C2GnT-M-characteristic multi-disaccharide acceptor specificity exhibited by the recombinant protein (Figure 2) (Ropp et al. 1991). Like other Golgi glycosyltransferases, mC2GnT-M is a type II membrane protein (Toth and Colley 1989). It contains the nine cysteines conserved among all C2GnT and cIGnT enzymes (Beum and Cheng 2001). Further, as compared to C2GnT-L, C2GnT-T, and cIGnT, mC2GnT-M contains four additional cysteines, including C8, C18, C35, and C52, which are also found in other C2GnT-Ms (Figure 8). C8 is localized in the cytoplasmic tail, C18 in the transmembrane domain, and C35 and C52 in the stem region. The biological significance of these four cysteines is not known, although some of them may be involved in the retention of this enzyme in the Golgi compartment by forming complexes with itself or other members of the Golgi glycosyltransferases in the same Golgi stacks (Colley 1997). This glycosyltransferase has a 89% amino acid sequence identity to that of rat C2GnT-M (GenBank accession number: AB98520) (Korekane et al. 2003), but only a 70% sequence identity to those of bovine (GenBank accession number: AY283763) (Choi et al. 2004), bovine herpes virus (Vanderplasschen et al. 2000) (GenBank accession number: AF231105) and human (GenBank accession number: NP004742) (Yeh et al. 1999; Schwientek et al. 2000) C2GnT-Ms. Among these three C2GnT-Ms, bovine and bovine herpes virus C2GnT-Ms have a 94% sequence identity as compared to a 79% sequence identity between two human C2GnT-Ms. Therefore, based on the amino acid sequence, mC2GnT-M is closer to rat C2GnT-M than to human and bovine counterparts.

Similar to bovine (Choi et al. 2004) and human (Tan and Cheng 2007) C2GnT-M genes, mC2GnT-M gene contains three exons and two introns although size distribution of the transcripts and exon-intron utilization are different. Bovine C2GnT-M gene (6.3 kb) is made of 128 bp exon 1, 1427 bp intron 1, 186 bp exon 2, 1771 bp intron 2, and 1798 bp exon 3 (Choi et al. 2004). Human C2GnT-M gene (8.26 kb) is made of 69–198 bp exon 1, 4.5 kb intron 1, 333–401 bp exon 2, 1297 bp intron 2, and 1864 bp exon 3 (Tan and Cheng 2007) (GenBank accession number: EF152283). The sizes of the two mC2GnT-M transcripts are 2.25 and 4.3 kb, which are intronless and differ by the size of the 3′ UTR. Both transcripts are longer than the length (1841 bp) of the reported mC2GnT-M cDNA fragment (GenBank accession number: BC018297). It should be noted that these authors assigned the translation start site, ATG, at +118 (Figure 1 and GenBank accession numbers: EF202835 and BC018297). This previously reported cDNA fragment is composed of a 74 bp (instead of 191 bp as claimed) 5′ UTR, 1314 bp (instead of 1197 bp as claimed) ORF, and 453 bp 3′ UTR. Our findings extend the reported length of mC2GnT-M transcript by exon 1, an additional 5′ UTR of exon 2, and a long transcript with a significantly longer 3′ UTR. Closer examination of the 5′ UTRs identified by 5′ RACE and RT-PCR reveals that these transcripts have the same exon 1 but various sizes of exon 2 (Figure 5B). All exon 2s of mC2GnT-M transcripts are flanked by the signature splicing sequences, AG at the 3′ end of intron 1 and GT at the 5′ end of intron 2 (Padgett et al. 1986), suggesting that they are derived from alternative splicing. Further, intron-exon utilization of mC2GnT-M gene is different from those of human (Tan and Cheng 2007) and bovine (Choi et al. 2004) C2GnT-M genes. The sizes of the three human C2GnT-M transcripts (Yeh et al. 1999; Tan and Cheng 2006) are 6.8–7.0 kb, 3.6–3.8 kb, and 2.3–2.5 kb. All transcripts contain all three exons, but one has no intron, one has intron 1, and one has intron 2. The bovine C2GnT-M transcripts contain two different size groups, one with and one without intron 2. Within each group of similar size transcripts, one does and one does not have intron 1. Expression of exon 1 is tissue-specific for bovine C2GnT-M gene in that intron one containing transcripts are expressed only in the trachea and the testis (Choi et al. 2004). However, the expression of mouse and human C2GnT-M transcripts is not tissue-specific.

The physiological significance of expressing intronless transcripts for mC2GnT-M gene is not clear although this type of transcript is considered more favorable for translation when compared to transcripts containing long intron(s) at the 5′ UTR according to the ribosome scanning mechanism (Kozak 1996, 2000, 2002). Another difference between bovine and mouse C2GnT-M transcripts is that the most abundant mC2GnT-M transcripts are the long transcripts, which are 2049 bp longer at the 3′ UTR than the short transcripts. Increased stability of the long transcripts may be one explanation for the observation. As has been previously reported for liver Galβ1,4GlcNAcα2,6-sialyltransferase mRNA, the half-life of this transcript can be extended by a 3′ UTR-specific binding protein (Beaudoing et al. 2000). In addition, 3′ UTR has been viewed as a regulatory region essential for appropriate expression of many genes (Conne et al. 2000). These long and short transcripts contain the canonical polyadenylation signal (AATAAA) and the variant
polyadenylation signal (CATAAA) near their 3′ ends (Garige et al. 2006), respectively. It has been proposed that the variant polyadenylation signal is not processed as efficiently as the canonical signal. Therefore, the long transcripts may be more efficiently utilized as a template for protein synthesis.

We further show that mC2GnT-M gene is highly expressed in mucus-secreting tissues, especially colon, stomach, and small intestine (Figure 5A), which is in agreement with previous reports on human (Yeh et al. 1999), bovine (Choi et al. 2004), and rat (Korekane et al. 2003) C2GnT-Ms. However, the cell types which express this gene have never been identified although they have been assumed to be mucus cells. By immunohistochemical staining employing antibodies against the C-terminal peptide of mC2GnT-M, we demonstrate that mC2GnT-M is expressed in mucus-secreting cells such as the goblet cells of the colon and the small intestine, foveolar epithelial cells, mucus neck cells, pyloric glands of the stomach, Brunner’s glands and Paneth cells of the small intestine. The detection of mC2GnT-M in Paneth cells is a surprising finding because these cells have never been regarded as mucus cells despite the presence of PAS-positive granules. Further studies are warranted to see if these PAS-positive materials are secreted mucins. The detection of C2GnT-M protein in the mucus cells of gastrointestinal epithelia suggests the important role C2GnT-M plays in the production of secreted mucins, the major function of which is to protect the gastrointestinal tract. The highest expression of mC2GnT-M gene is consistently detected in mucin-secreting cells that express MUC2 (Ho et al. 1993; Chang et al. 1994) and MUC6 (Bartman et al. 1998), and to a lesser degree, MUC5AC (Ho et al. 1995a, 1995b), which are secreted mucins containing cysteine-rich motifs responsible for oligomerization of these mucins to form a gel layer covering the mucosal surface. Localization of secreted mucins and C2GnT-M in mucin-secretory cells, and the presence of core 2, core 4, and blood group I structures in secreted mucins (Breg et al. 1987; Hounsell et al. 1989; Lamblin et al. 1991) support the notion that C2GnT-M is involved in the synthesis of these mucins.

The one-enzyme–one-product principle predicts that each glycosyltransferase forms only a single glycosidic linkage. This is generally true, but with a few exceptions. These include Lewis blood group Galβ1,4/3GlcNAc:α1,3/4fucosyltransferase, Galβ1,4/3GlcNAc: (Sialic acid-Gal)α2,3-sialyltransferase (Prieels et al. 1981; Kukowska-Latallo et al. 1990; Wen et al. 1992), and C2GnT-M (Ropp et al. 1991; Schwientek et al. 1999; Yeh et al. 1999; Vanderplasschen et al. 2000; Korekane et al. 2003; Choi et al. 2004). These enzymes can make more than one glycosidic bond. For example, C2GnT-M can form all three mucin-type β6 GlcNAc branch
**mC2GnT-M cDNA and expression in mucus cells**

### Fig. 8.

#### (A) Alignment of deduced amino acid sequences of C2GnT-M from bovine (B, AY283763); bovine herpes virus (BV, AF231105); human (H, AF102542); mouse (M, EF202835); and rat (R, AB098520). The nine cysteines conserved among all β6GlcNAc transferases are indicated with ▼. The four additional cysteines conserved only in C2GnT-Ms are C8, C18, C35 and C52 (based on mC2GnT-M sequence) are indicated by ^. The identical sequences are indicated with a black background and homologous sequences with a gray background. Additionally, under “similarity or Sim;” the asterik (∗) indicates identical amino acids; the colon (:)) indicates conserved substitutions; the point (.) indicates semi-conserved substitutions.

#### (B) The cladogram of this enzyme from four different species as predicted using ClustalW software (European Bioinformatics Institute).

---

<table>
<thead>
<tr>
<th>B</th>
<th>BY283763</th>
</tr>
</thead>
<tbody>
<tr>
<td>BV</td>
<td>AF231105</td>
</tr>
<tr>
<td>H</td>
<td>AF102542</td>
</tr>
<tr>
<td>M</td>
<td>EF202835</td>
</tr>
<tr>
<td>R</td>
<td>AB098520</td>
</tr>
</tbody>
</table>

---

**Fig. 8.** (A) Alignment of deduced amino acid sequences of C2GnT-M from bovine (B, BY283763); bovine herpes virus (BV, AF231105); human (H, AF102542); mouse (M, EF202835); and rat (R, AB098520). The nine cysteines conserved among all β6GlcNAc transferases are indicated with ▼. The four additional cysteines conserved only in C2GnT-Ms are C8, C18, C35 and C52 (based on mC2GnT-M sequence) are indicated by ^. The identical sequences are indicated with a black background and homologous sequences with a gray background. Additionally, under “similarity or Sim;” the asterik (∗) indicates identical amino acids; the colon (:)) indicates conserved substitutions; the point (.) indicates semi-conserved substitutions. (B) The cladogram of this enzyme from four different species as predicted using ClustalW software (European Bioinformatics Institute).
structures (Ropp et al. 1991; Schwientek et al. 1999; Yeh et al. 1999; Vanderplasschen et al. 2000; Korekane et al. 2003; Choi et al. 2004). This one-enzyme–multiple-products phenomenon would ensure a rapid production of all these structures within a short notice in order to meet such a need. Another interesting feature of glycan biosynthesis is that the same carbohydrate linkage can be formed by several different glycosyltransferases.

For example, the mucin glycan core 2 structure can be generated by three different C2GnTs, which include C2GnT-L, C2GnT-M, and C2GnT-T. While C2GnT-M can form core 2, core 4, and blood group I branch structures, and C2GnT-L and C2GnT-T can form only core 2 structures. These three isozymes are encoded by three different genes and have different tissue distribution. C2GnT-L is ubiquitously distributed (Bierhuizen et al. 1992; Beum and Cheng 2001) while C2GnT-T is highly restricted to thymus (Schwientek et al. 2000) and C2GnT-M to mucin-secreting cells. Tissue-specific expression of these enzymes ensures that only the enzyme in a particular tissue is generated when needed.

While C2GnT-L and C2GnT-T are involved in immune functions (Ellies et al. 1998; Tsuibo and Fukuda 1997, 1998, 2001), C2GnT-M is involved in the first line of host defense against airborne and ingested pathogens. These epithelia are rich in mucin-secreting cells. These mucins are high in carbohydrate content and very heterogeneous in carbohydrate structures. Mucin with high carbohydrate content is important for retention of water and the maintenance of mucus rheological properties. Heterogeneous carbohydrate structures for secreted mucins would ensure that they bind as many different pathogens as possible. It has been shown that mucins secreted from the airways contained hundred of glycans with different mucin-type carbohydrate structures (Lamblin et al. 1991; Xia et al. 2005) as compared to less than 20 different oligosaccharides detected on the surface membrane-tethered mucin of T cells (Piller et al. 1988; Saitoh et al. 1991). The repertoire of nonreducing terminal carbohydrate structures of secreted mucins can be affected by C2GnT-M because this enzyme can make all three β6GlCNac structures which can be decorated with different carbohydrates. Therefore, modulation of C2GnT-M gene expression can have a profound effect on the functions of secreted mucins. It has been shown that the C2GnT-M enzyme activity (Yang et al. 1994) and gene expression (Huang et al. 2006) are in colorectal cancer and that restoration of its expression suppresses the growth of cancer cells (Huang et al. 2006). Downregulation of core 3 synthase, which produces core 3 to serve as an acceptor for C2GnT-M, has also been observed in colorectal cancer (Iwai et al. 2005). Decreased expression of either or both genes may lead to the production of underglycosylated mucins, thus reducing the protective function of these mucins. This condition may mimic the one found in the Muc2 gene knockout mouse, which develops colorectal cancer (Velcich et al. 2002) and colitis (Van der Sluis et al. 2006). It would be of interest to see if the C2GnT-M gene knockout mouse (Stone et al. 2006) develops colorectal cancer and colitis.

We previously showed that the C2GnT-M gene expression is subject to regulation: Th2 cytokines and retinoic acid upregulate (Beum et al. 2005) while the epidermal growth factor inhibits C2GnT-M gene expression (Beum et al. 2003). Inhibition of the expression of this gene converts mucin core structures from core 2 to core 1 (Beum et al. 2003). A cyclical change in the C2GnT-M gene expression during estrous cycle has also been reported in the oviduct of the golden hamster (McBride et al. 2005). To understand the biological functions of this enzyme, we plan to characterize the C2GnT-M gene knockout mouse when it becomes available. The information provided by this report should facilitate these efforts.

Materials and methods

Materials

The materials used in this study were purchased from the following suppliers: Galβ1–3GalNAca-benzyl and GlcNacβ1–3GalNAca-p-nitrophenyl, Toronto Research Chemical (Toronto, Canada); GlcNacβ1–3Galβ-methyl, Maackia amurensis agglutinin and Protease Inhibitor Cocktail (P8340), Sigma-Aldrich (St. Louis, MO); UDP-[3H]GlcNAc, American Radiolabeled Chemicals (St. Louis, MO); Transferrin (iron-saturated), Collaborative Biomedical Products (Bedford, MA); and Bond Elute C18 cartridges, Varian (Sunny Vale, CA).

Animals and tissue preparation

C57BL/6 mice (7–9-week-old) were sacrificed by exsanguination through the abdominal aorta after deep anesthesia with diethyl ether or carbon dioxide asphyxiation. Fresh organs were resected, frozen immediately in liquid nitrogen, and stored at −80°C until the RNA isolation. Tissues for immunohistochemical staining were fixed in 10% buffered formalin for less than 24 h and embedded in paraffin using standard methods. The protocol was approved by the Institutional Animal Care and Use Committee at the University of Nebraska Medical Center and Nagoya University Graduate School of Medicine.

RNA isolation

Total RNAs were extracted from mouse frozen organs using TRI reagent (GIBCO-BRL, Gaithersburg, MD), Poly(A)+ RNA was then isolated from total RNA using the PolyAtract messenger RNA (mRNA) isolation system III (Promega, Madison, WI).

Identification and cloning of mC2GnT-M complementary DNA (cDNA)

mC2GnT-M cDNA sequence was retrieved from the GenBank database with the coding sequence of the human C2GnT-M cDNA (GenBank accession number: NP004751) as a query sequence using the blastn algorithm at the National Center for Biotechnology Information. A cDNA fragment of 1841 bp (GenBank accession number: BC018297) of the mouse homologue of human C2GnT-M was found. The cDNA fragment contained a fragment of the 5′ untranslated region (5′ UTR), open reading frame (ORF), and a 3′ untranslated region (3′ UTR). To obtain the complete sequences of both 5′ and 3′ UTRs, rapid amplification of cDNA ends (RACE) was performed using SMART RACE cDNA Amplification Kit (CLONTECH, Palo Alto, CA) with 1 µg of Poly(A)+ RNA isolated from mouse colon and genespecific primers (GSPs) according to the user’s manual. Antisense GSPs (5′-GTCCTGGAGGCAAGTCTTTCAGGCGCAAG -3′ and 5′-TGAGCACCCTGGGAAAGTGAAAGGCA-3′) and sense GSP (5′-TCAGGGGACCTGCACTGGTGACTTAA-3′) were used for 5′ RACE and 3′ RACE, respectively, under the recommended condition (5 cycles: 94°C for 30 s, 72°C for
nts between 5’ RACE and 3’ RACE products. RT-PCR (sense primer: 5’-TTCTCACTTTCCAGGTGCT-3’, antisense primer: 5’-TTTCACTCCCCGACCTAATG-3’, 35 cycles: denaturation 94°C for 30 s, annealing 58°C for 30 s, extension 72°C for 4 min) with Takara LA Taq polymerase (Takara) was carried out to obtain cDNA fragments between 5’ RACE and 3’ RACE products. RT-PCR (sense primer: 5’-TTTCTCACTTTCCAGGTGCT-3’, antisense primer: 5’-TTTCACTCCCCGACCTAATG-3’, 35 cycles: denaturation 94°C for 15 s, annealing 58°C for 30 s, extension 72°C for 45 s) was also performed to confirm three different exon sequences. RACE and RT-PCR products were purified by QIAquick Gel Extraction Kit (QIAGEN) and cloned into the pGEM-T Easy vector (Promega, Madison, WI), with at least three different clones of each product being sequenced.

The complete genomic sequence of the mC2GnT-M gene was also obtained from the GenBank database employing the sequences generated by 5’ and 3’ RACE.

**Northern blot analysis**

Northern blot analysis was performed using the standard methods (Sambrook et al. 2001). Briefly, total RNA (20 µg/lane) and RNA size markers (Wako Chemicals, Richmond, VA) were fractionated by 1% agarose gel containing 2.2 M formaldehyde and transferred to a nylon membrane. RNA was fixed to the membrane by baking at 80°C for 2 h. The membrane was prehybridized in a solution containing 50% formamide, 5× saline sodium citrate (SSC), 5× Denhardt’s solution, 5.0 mM EDTA, 0.1% sodium dodecyl sulfate (SDS), and 100 µg/mL denatured salmon sperm DNA at 42°C for 4 h. Hybridization was carried out in a solution containing 50% formamide, 5× SSC, 5× Denhardt’s solution, 5.0 mM EDTA, 0.1% SDS, 100 µg/mL denatured salmon sperm DNA, 10% dextran sulfate, and 32P-labeled probes at 42°C for 18 h. The membrane was then washed twice with 1× SSC, 0.1% SDS at room temperature for 15 min and twice with 0.25× SSC, 0.1% SDS at 58°C for 15 min and visualized by autoradiography.

Three different probes located in 5’ UTR and 3’ UTR of mC2GnT-M were generated by digestion of the plasmid vectors containing the RACE products with SpeI, XbaI, and EcoRI. The probes were purified using QIAquick Gel Extraction Kit (QIAGEN) and labeled with [α32P]dCTP using Prime-It II Random Primer Labeling Kit (Stratagene, La Jolla, CA). Glycerinaldehyde 3-phosphate dehydrogenase (GAPDH) cDNA probe was used as an internal control.

**Antibodies**

Anti-mC2GnT-M polyclonal antibodies were generated by immunization of rabbits with a keyhole limpet hemocyanin-disulfide bond conjugated peptide (VLQCLEEYLRHKAIYGTEL) located at the carboxyl-terminus of mC2GnT-M (GenBank accession number BC018297 and EF020285) (0.5 mg/2 mL/injection, five times at two-week intervals). The antibodies (1.9 mg/mL) were then affinity purified from the immune sera. Biotinylated swine anti-rabbit immunoglobulin was purchased from Dako (Glostrup, Denmark). Anti-FLAG M2 monoclonal antibody was obtained from Sigma-Aldrich.

**Construction of Expression Vectors Encoding mC2GnT-M**

A 1214 bp fragment (+97 to +1311) encoding the stem and catalytic domain of mC2GnT-M was obtained by PCR (24 cycles: denaturation 94°C for 1 min, annealing 54°C for 1 min, extension 72°C for 1 min 30 s) with Pfu polymerase (Stratagene, La Jolla, CA) using sense primer (5’-CCCAAGCTTCTGAAATGCGACTTCAAGC-3’) and antisense primer (5’-ATAGAATGGCAGCGCTTCTCAAGTTCGACTCATAAGTCCATAG-3’) which contained HindIII and NotI restriction sites plus extra nucleotides at their 5’ ends, respectively. The antisense primer included the stop codon of ORF. The PCR product was ligated into HindIII and NotI sites of the mammalian expression vector p3xFLAG-myc-CMV25 (Sigma-Aldrich), designated as p3xFLAG-C2GnT-M, and sequenced in full to confirm the sequence integrity. This plasmid was used for expression of the recombinant protein in a transient transfection experiment performed in Panc1 cells.

A 1128 bp mC2GnT-M cDNA fragment (+181 to +1308 bp) encoding the catalytic domain was also cloned from mC2GnT-M cDNA by PCR using the High-fidelity Taq DNA polymerase and the following primers:

- AGGGCCACGGGCCAAGCTGAGCCAGAAGAGTTCCATCC (forward)
- AGGGCCCTAGGTCAGTCAGCCATAGTGCTTTGTGTG (reverse)

The PCR condition was as follows: 30 cycles at 94°C for 30 s; 50°C for 30 s; 70°C for 1 min 30 s, and followed by 1 cycle at 72°C for 5 min. The PCR product was cloned into pCRII vector and confirmed by sequencing. Subsequently, the PCR product was subcloned into the pSecTag2A secretory expression vector behind Igκ-chain leader sequence at the Sfi/Apal sites. The resultant construct encoded a secreted protein composed of 406 amino acids, which included a hexapeptide spacer at the N-terminus, 377 amino acids (residues 61–437) of mC2GnT-M followed by a 17 amino acid spacer and then 6-His at the C-terminus and was named pSecTag-mC2M181. The plasmid DNA was used for construction of a stable clone in CHO cells for characterization of the substrate specificity of the recombinant enzyme secreted into the conditioned medium.

**Transient and stable expression of mC2GnT-M in Panc1 and CHO cells**

Panc1 cells (American Type Culture Collection, Manassas, VA) were grown in DMEM with 10% fetal bovine serum and antibiotics (50 units/mL penicillin and 50 µg/mL streptomycin). Transient transfection of Panc1 cells with p3xFLAG-mC2GnT was carried out under serum-free conditions using a transfection-assisted lipofection protocol as previously described (Cheng 1996). The transfection solution was replaced with standard culture medium without serum after 6 h. Forty-two hours later, the conditioned medium was concentrated by Centricron-10 (Amicon), and used for Western blot analysis. As a negative control, p3xFLAG-myc-CMV25 vector alone, which was designated as p3xFLAG-Empty, was used.

CHO cells stably transfected with pSecTag-mC2M181 were prepared as described below. Briefly, CHO cells cultured to 50% confluency in F-12 medium containing 5% FBS were transfected with 8 mL transfection solution containing 80 µg Maackia
amurensis agglutinin, 80 μg DMRIE-C, 17 μg pSecTag-mC2M181, 2 mL PBS, and 6 mL serum-free F-12 medium prepared as described (Yanagihara and Cheng 1999). The stable clones were selected by resistance to 500 μg/mL Zeocin and sorted for single cell into 96-well plates by a cell sorter. The positive clones were identified by ELISA assay of the conditioned medium using anti-His-Tag antibodies after 2 mM butyrate treatment for 24 h. These clones were further characterized by measurement of C2GnT activity in the cells cultured in T-25 flasks treated with 2 mM butyrate for 24 h. The clone that yielded the highest C2GnT activity was chosen for characterization of the disaccharide acceptor specificity.

**Characterization of endogenous and recombinant mC2GnT-M protein**

To characterize the endogenous mC2GnT-M protein, colon tissue was crushed in liquid nitrogen, mixed with lysis buffer (25 mM Tris–HCl, pH 7.5, 125 mM NaCl, 2.5 mM EDTA, 10% glycerol, 1 mM DTT, 1% Triton X-100), Protease Inhibitor Cocktail), briefly sonicated on ice, and centrifuged at 10000 × g for 3 min. The supernatant was subject to Western blot analysis as previously described (Choi et al. 2004). Briefly, protein samples (recombinant C2GnT-M: 0.5 μg, colon tissue: 20 μg) were mixed with 2× sample buffer (250 mM Tris–HCl pH 6.8, 4% SDS, 20% glycerol, 0.006% bromphenol blue, 10% β-mercaptoethanol) and heated in boiling water for 3 min. The samples were separated by SDS-13% polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto the PVDF membrane. The membrane was blocked with 5% skim milk in Tris buffered saline-Tween (TBS-T) for 1 h at room temperature, then incubated with anti-mC2GnT-M antibodies (at a dilution of 1:500) or anti-FLAG M2 antibody (at a dilution of 1:500) in TBS-T containing 5% skim milk for 1 h at room temperature. Signals were detected using horseradish peroxidase (HRP)-conjugated anti-rabbit or mouse IgG and Lumi-LightPLUS Western blotting substrate (Roche).

**Assay of C2GnT, C4GnT, and dIGnT activities**

CHO cells stably transfected with pSecTag-mC2M181 were cultured in T-75 flasks to confluence in F-12 containing 5% FBS before medium was switched to 15 mL of F-12 medium containing 1% FBS and 2 mM sodium butyrate and then cultured for 40 h. The conditioned medium was harvested and concentrated about 30-fold with a 30 kDa molecular weight cutoff filtration membrane. The concentrated medium was assayed for C2GnT, C4GnT, and dIGnT activities as described previously (Ropp et al. 1991) using 2 mM UDP-[3H]GlcNAc (1500 dpm/nmol) as the donor substrate and 2 mM Galβ1-3GlcNAcα-O-Benzyl, 4 mM GlcNAcβ1–3GalNAcα-O-PNP, or 4 mM GlcNAcβ1–3Galβ1-methyl as the acceptor substrates, respectively. The products of C2GnT and C4GnT assays were isolated by solid phase extraction on C18 cartridges with 2 mL of methanol after rinsing with 10 × 3 mL of 0.1 M Tris–HCl, pH 7.5. The isolated products were dried by SpeedVac System (Savant, Holbrook, NY) and resuspended in 300 μL H2O. The product of dIGnT assay was isolated in the run-through and water washing of AG1 Dowex 1-Cl- column. Radioactivity associated with the product was measured with a liquid scintillation counter (Packard, Meriden, CT). The enzyme activity was calculated by subtracting the endogenous activity measured without exogenous acceptor from total activity and was expressed as nmol of sugar donor transferred per hour. The data were expressed as ratios of C4GnT and dIGnT activities to C2GnT activity.

**Immunohistochemistry (IHC)**

Immunohistochemical staining was performed utilizing the polyclonal antibodies specific for mC2GnT-M antigen and a biotin–avidin peroxidase complex method on formalin-fixed and paraffin-embedded sections, as previously described (Hsu et al. 1981). The first antibody was an anti-mC2GnT-M polyclonal antibody (at a dilution of 1:200) and the secondary antibody was a biotinylated swine anti-rabbit immunoglobulin (at a dilution of 1:300). Color was developed with diaminobenzidine after treatment with avidin peroxidase. No antigen retrieval step was required prior to immunohistochemical staining of paraffin sections. The treated paraffin sections were counterstained with hematoxylin. Sections processed in the absence of first or second antibody were used as negative controls. In addition, we confirmed that working dilution of the anti-mC2GnT-M antibodies preabsorbed using the synthetic C2GnT-M peptide (10 μg/mL) resulted in negative staining.

**Funding**

Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan to M.H. and research grants from National Institute of Health (RO1 HL48282), Cystic Fibrosis Foundation, and State of Nebraska (NRI-Cancer Glycobiology Program and LB506) to P.W.C.

**Acknowledgements**

We thank Dr. David Klinkebiel for proofreading the manuscript. We also thank Yoriko Tanaka, Yousuke Murase, Emiri Miura, Toshiyasu Sakai, and Nobuaki Misawa from the Department of Pathology, Nagoya University Graduate School of Medicine, Nagoya, Japan, and Helen Cheng at UNMC for excellent technical assistance.

**Conflict of interest statement**

None declared.

**Abbreviations**

C2GnT Core 2 β1,6 N-acetylgalcosaminyltransferase; C2GnT-L (or 1) C2GnT leukocyte-type; C2GnT-M (or 2), C2GnT mucus-type; C2GnT-T (or 3), C2GnT thymus-type; cGnT, centrally acting blood group I α, 6N-acetylgalcosaminyltransferase; dGnT, distally acting blood group I α, 6N-acetylgalcosaminyltransferase; IHC, immunohistochemistry; PAS, periodic acid-Schiff; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription polymerase chain reaction.

**References**


