FEBS Letters 562 (2004) 134–140 FEBS 28189

Molecular cloning and characterization of β1,4-*N*-acetylgalactosaminyltransferases IV synthesizing *N*,*N*′-diacetyllactosediamine[1]

Masanori Gotoh[a,b], Takashi Sato[a], Katsue Kiyohara[a], Akihiko Kameyama[a,b], Norihiro Kikuchi[a,c], Yeon-Dae Kwon[a,c], Yasuko Ishizuka[a,d], Toshie Iwai[a], Hiroshi Nakanishi[d], Hisashi Narimatsu[a,*]

[a] *Glycogene Function Team, Research Center for Glycoscience, National Institute of Advanced Industrial Science and Technology (AIST), Open Space Laboratory Central-2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan*
[b] *Amersham Biosciences KK, 3-25-1, Hyakunincho, Shinjuku-ku, Tokyo 169-0073, Japan*
[c] *Mitsui Knowledge Industry Co., Ltd., Honcho 1-Chome, Nakano-ku, Tokyo 164-8721, Japan*
[d] *Biological Information Research Center, AIST, Central-6, 1-1 Higashi, Tsukuba, Ibaraki 305-8568, Japan*

Received 19 January 2004; revised 2 February 2004; accepted 18 February 2004

First published online 4 March 2004

Edited by Masayuki Miyasaka

Abstract A sequence highly homologous to β1,4-*N*-acetylgalactosaminyltransferase III (β4GalNAc-T3) was found in a database of human expressed sequence tags. The full-length open reading frame of the gene, *β4GalNAc-T4* (GenBank accession number AB089939), was cloned using the 5′ rapid amplification of cDNA ends method. It encodes a typical type II transmembrane protein of 1039 amino acids having 42.6% identity with β4GalNAc-T3. The recombinant enzyme transferred *N*-acetylgalactosamine to *N*-acetylglucosamine-β-benzyl with a β1,4-linkage to form *N*,*N*′-diacetyllactosediamine as did β4GalNAc-T3. In specificity toward oligosaccharide acceptor substrates, it was quite similar to β4GalNAc-T3 in vitro, however, the tissue distributions of the two enzymes were quite different. These results indicated that the two enzymes have similar roles in different tissues.

Key words: Glycosyltransferase; *N*-Acetylgalactosaminyltransferase; *N*-Glycan; *N*,*N*′-Diacetyllactosediamine; Glycodelin; Lutropin

1. Introduction

N-Glycan is a general term for a sugar chain covalently bound to Asn on proteins. *N*,*N*′-Diacetyllactosediamine (GalNAcβ1-4GlcNAc, LacdiNAc) is an unusual constituent of *N*-glycans in mammals. It has been found on glycoproteins and glycohormones produced in certain tissues. Lutropin (LH), thyrotropin [1], glycodelin [2] and tissue factor pathway inhibitor [3] are all glycoproteins carrying the LacdiNAc structure. LH is a pituitary glycoprotein hormone which is essential for the regulation of follicular maturation, ovulation and the secretion of estradiol and progesterone. Both LH and follicle stimulating hormone, consisting of a common α-subunit and specific β-subunit [4], are produced in the same cells, gonadotrophs of the anterior pituitary, however, they differ in the terminal structure of their *N*-glycans. The SO_4-4GalNAcβ1-4GlcNAcβ1-2Man structure on LH [5,6] is recognized by a receptor in hepatic endothelial and Kupffer cells to be cleared rapidly from blood [7].

Glycodelin (Gd) is expressed in various cells of reproductive organs and erythroid precursor cells of bone marrow [8–12]. The Gd isoform in amniotic fluid is called GdA, and the corresponding isoform in seminal plasma GdS [13]. GdA carries LacdiNAc, sialylated LacdiNAc (NeuAcα2-6GalNAcβ1-4GlcNAc) and a LacdiNAc analog of Lewis X (Le^x, GalNAcβ1-4(Fucα1-3)GlcNAc) at the non-reducing terminus, in addition to LacNAc (Galβ1-4GlcNAc), sialylated LacNAc and Le^x (NeuAcα2-6Galβ1-4(Fucα1-3)GlcNAc) [2], whereas the major non-reducing epitopes on GdS are Le^x and Le^y (Fucα1-2Galβ1-4(Fucα1-3)GlcNAc) [14]. These isoforms of Gd exhibit different biological functions directed by the carbohydrate epitopes. GdA, immunologically indistinguishable from GdS, was reported to inhibit sperm–zona pellucida binding in a hemizona assay [15]. This may reflect an important role for GdA in fetomaternal defense [16]. In contrast, GdS does not have this function [15]. Recently, Dell et al. reported that zona pellucida 3 derived from mouse eggs expresses *O*-glycans containing LacdiNAc [17].

Very recently, we cloned and characterized a β1,4-*N*-acetylgalactosaminyltransferase, β4GalNAc-T3, which synthesizes the LacdiNAc structure in both *N*- and *O*-glycans [18]. β4Gal-

*Corresponding author. Fax: (81)-29-861 3201.
E-mail address: h.narimatsu@aist.go.jp (H. Narimatsu).

[1] The nucleotide sequence reported in this paper has been registered in the GenBank™/EBI Data Bank with accession numbers AB089939 and AB114827.

Abbreviations: LacdiNAc, *N*,*N*′-diacetyllactosediamine (GalNAcβ1-4GlcNAc); GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; LH, lutropin; Man, mannose; Gd, glycodelin; NeuAc, *N*-acetylneuraminic acid; Gal, galactose; Le, Lewis; Le^x, Galβ1-4(Fucα1-3)GlcNAc; Fuc, fucose; Le^y, Fucα1-2Galβ1-4(Fucα1-3)GlcNAc; β4GalNAc-T, β1,4-*N*-acetylgalactosaminyltransferase; β4GT, β1,4-glycosyltransferase; EST, expressed sequence tag; ORF, open reading frame; RACE, rapid amplification of cDNA ends; HEK, human embryonic kidney; GlcA, glucuronic acid; *p*Np, *para*-nitrophenyl; Bz, benzyl; Glc, glucose; Xyl, xylose; MS, mass spectrometry; MALDI-TOF, matrix-assisted laser desorption/ionization-time of flight; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; HPLC, high performance liquid chromatography

0014-5793/04/$30.00
doi:10.1016/S0014-5793(04)00219-4

NAc-T3 was expressed in stomach, colon and testis at high levels, but in brain at a very low level. Woodworth et al. reported that tenascin-R associated with Purkinje cell bodies and their dendrites in the molecular layer of the cerebellum bears *N*-glycans terminating with SO_4-4GalNAcβ1-4GlcNAc [19]. A high level of β4GalNAc-T activity was observed in other regions of the brain in addition to the pituitary gland and cerebellum [20]. This suggested the existence of an unknown β4GalNAc-T which is responsible for the synthesis of LacdiNAc in brain.

In this paper, we report the cloning and characterization of a novel β4GalNAc-T, named β4GalNAc-T4, that is the second such enzyme synthesizing the LacdiNAc structure. Considering the similar substrate specificities of β4GalNAc-T4 and T3 in transfected cells and their different tissue distribution, we discuss their roles in vivo.

2. Materials and methods

2.1. Isolation of human β4GalNAc-T4 cDNAs

We performed a BLAST search of the GenBank™ database using β4GT motifs, such as β4GalNAc-T3 (GenBank™ accession number AB089940), as query sequences and identified an expressed sequence tag (EST) with accession number N48738, which contained a partial open reading frame (ORF) but showed high homology to the C-terminal region of β4GalNAc-T3. To obtain the complete ORF, the 5′ rapid amplification of cDNA ends (5′-RACE) method was employed using a Marathon-Ready™ cDNA Amplification Kit (Clontech, Palo Alto, CA, USA). Two reverse primers, 5′-GCTCCTGCAGCTCCAGCTCCA-3′ for the first polymerase chain reaction (PCR) and 5′-AAGCGACTCCCTCGCGCCGAGT-3′ for the nested PCR, were employed for the extension. The approximately 0.6 kb fragment amplified did not contain a transmembrane domain, therefore, another EST search was performed using the amino acid sequence of the amplified fragment as a query, and an EST with accession number BF058197 was found. The two sequences, N48738 and BF058197, were not contiguous, therefore, it was confirmed they exist on a single mRNA by reverse transcription (RT) PCR with two primers, 5′-ATGCCGCGGCTCCCGGTGAAGAAG-3′ and 5′-AAGCGACTCCCTCGCGCCGAGT-3′. The sequences of the DNA fragments obtained by the 5′-RACE method and RT-PCR were determined using a DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences, Amersham, UK). Finally, a cDNA sequence encoding the ORF was obtained by PCR using the Marathon-Ready™ cDNA of human brain (Clontech) as a template.

2.2. Construction and purification of β4GalNAc-T4 and T3 proteins fused with FLAG peptide

The putative catalytic domain of β4GalNAc-T4 (amino acids 62–1039) was expressed as a secreted protein fused with a FLAG peptide in human embryonic kidney HEK293T cells. An approximately 2.9 kb DNA fragment was amplified by PCR using the Marathon-Ready™ cDNA derived from human whole brain as a template, and two primers, 5′-CCCAAGCTTCGGGGGGTCCACGCTGCGCCAT-3′ and 5′-GCTCTAGACTCAAGACGCCCCCGTGCGAGA-3′. The amplified fragment was digested with the restriction endonucleases *Hin*dIII and *Xba*I, then inserted into the *Hin*dIII-*Xba*I site of pFLAG-CMV-1 (Sigma, St. Louis, MO, USA) to construct pCMV/β4GalNAc-T4. The construction method of the β4GalNAc-T3 expression vector was described in detail in our previous paper [18].

The mouse *β4GalNAc-T4* (*mβ4GalNAc-T4*) gene encoding a putative catalytic domain (amino acids 45–1034) was amplified with 5′-CCCAAGCTTCGCCTGGGCTACGGGCGAGAT-3′ and 5′-GCTCTAGACTCAGGATCGCTGTGCGCGGGCA-3′ using the cDNA derived from mouse stomach as a template. The amplified 3.0 kb fragment was digested with the restriction endonucleases *Hin*dIII and *Xba*I, then inserted into pFLAG-CMV-1.

The catalytic domains of β4GalNAc-T4 and T3 were expressed in HEK293T cells. A 50 ml volume of culture medium was mixed with anti-FLAG M1 antibody resin (Sigma) and incubated with rotation at 4°C overnight. The resin was washed twice with 50 mM Tris-buffered saline (50 mM Tris–HCl, pH 7.4, 150 mM NaCl) containing 1 mM $CaCl_2$ and suspended in 100 μl of the assay buffer [18].

2.3. Assay for glycosyltransferase activity

To determine the enzyme activity, UDP-GalNAc, UDP-Gal, UDP-GlcNAc, UDP-glucuronic acid (GlcA), GDP-mannose (Man), and GDP-fucose (Fuc) (Sigma) were utilized as donor substrates. For acceptor substrates, monosaccharide acceptors (Table 1, numbers 1–13) were purchased from Calbiochem (La Jolla, CA, USA) and Sigma. *N*- and *O*-glycan-related acceptor substrates were purchased from Seikagaku, Takara (Shiga, Japan) and Honen (Tokyo, Japan). For the GalNAc-T assay the method described in our previous paper was utilized [18].

2.4. Determination of products of β4GalNAc-T4 with mass spectrometry (MS) and ^1H nuclear magnetic resonance (NMR) spectra

The MS and NMR analyses were described in detail in our previous paper [18]. A matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) MS (Reflex IV, Bruker Daltonics, Billerica, MA, USA) was used for the analysis. A one-dimensional ^1H NMR spectrum was recorded with a DMX750 (Bruker, Germany, 750.13 MHz for ^1H nucleus) spectrometer at 25°C.

Table 1
Substrate specificity of β4GalNAc-Ts

	Acceptor substrate	Relative activity (%)	
		β4GalNAc-T4	β4GalNAc-T3
1.	GlcNAcβ-Bz	100[a]	100[a]
2.	GlcNAcα-Bz	nd	nd
3.	Galα-*p*Np	nd	nd
4.	Galβ-*o*Np	nd	nd
5.	GalNAcα-Bz	nd	nd
6.	GalNAcβ-Bz	nd	nd
7.	Glcα-*p*Np	nd	nd
8.	Glcβ-*p*Np	nd	nd
9.	GlcAβ-*p*Np	nd	nd
10.	Fucα-*p*NP	nd	nd
11.	Manα-*p*Np	nd	nd
12.	Xylα-*p*Np	nd	nd
13.	Xylβ-*p*Np	nd	nd
14.	GlcNAcβ1-6(Galβ1-3)GalNAcα-*p*Np (core 2-*p*Np)	15.2	11.4
15.	GlcNAcβ1-3GalNAcα-*p*Np (core 3-*p*Np)	20.0	32.3
16.	GlcNAcβ1-6GalNAcα-*p*Np (core 6-*p*Np)	190.7	220.4

[a]The absolute activity of β4GalNAc-T4 and T3 was 2.43 and 8.28 nmol/ml medium/h, respectively.

2.5. *Quantitative analysis of* β4GalNAc-T4 *and* T3 *transcripts in human tissues by real-time PCR*

For the quantification of *β4GalNAc-T4* and *T3* transcripts, we employed the real-time PCR method, as described in detail previously [21,22]. The primer set and the probe for the *β4GalNAc-T4* gene were as follows: the forward primer was 5′-GCTGCAGGCGGGAGTG-3′, the reverse primer 5′-TGCCGTCCAGGATGTTGG-3′ and the probe 5′-GCGGTAGAGGACGCC-3′ with a minor groove binder [23]. The primer set and the probe for the *β4GalNAc-T3* gene were shown in our previous paper [18]. The relative amount of transcript was normalized to the amount of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript in the same cDNA.

3. Results

3.1. *Determination of nucleotide and putative amino acid sequence of β4GalNAc-T4*

We determined a novel full-length cDNA sequence by the 5′-RACE method and registered it in the GenBank™ database with the accession number AB089939. The nucleotide sequence and putative amino acid sequence are shown in Fig. 1. The gene is located at 11p15 in the human genome database, however, information on the genomic structure is insuf-

```
ATGCCGCGGCTCCCGGTGAAGAAGATCCGTAAGCAGATGAAGCTGCTGCTGCTGCTGCTGCTGCTGAGCTGCGCCGCGTGGCTCACCTACGTGCACCTGGGCCTGGTGCGCCAGGGACGC    120
M  P  R  L  P  V  K  K  I  R  K  Q  M  K  L  L  L  L  L  L  L  L  S  C  A  A  W  L  T  Y  V  H  L  G  L  V  R  Q  G  R      40

GCGCTGCGCCAGCGCCTGGGCTACGGGCGAGATGGTGAGAAGCTGACCAGTGAGACCGACGGCCGGGGGGTCCACGCTGCGCCATCCACACAGAGGGCTGAGGACTCCAGTGAGAGCCGT    240
A  L  R  Q  R  L  G  Y  G  R  D  G  E  K  L  T  S  E  T  D  G  R  G  V  H  A  A  P  S  T  Q  R  A  E  D  S  S  E  S  R      80

GAAGAGGAGCAAGCGCCCGAAGGTCGGGACCTAGACATGCTGTTTCCTGGGGGGGCTGGGAGGCTGCCACTGAACTTCACCCATCAGACACCCCCATGGCGGGAGGAGTACAAGGGGCAG    360
E  E  E  Q  A  P  E  G  R  D  L  D  M  L  F  P  G  G  A  G  R  L  P  L  Ⓝ  F  T  H  Q  T  P  P  W  R  E  E  Y  K  G  Q    120

GTGAACCTGCACGTGTTTGAGGACTGGTGTGGGGGCGCCGTGGGCCACCTGAGGAGGAACCTGCACTTCCCGCTGTTCCCTCATACGCGCACCACCGTGAAGAAGTTGGCCGTGTCCCCC    480
V  N  L  H  V  F  E  D  W  C  G  G  A  V  G  H  L  R  R  N  L  H  F  P  L  F  P  H  T  R  T  T  V  K  K  L  A  V  S  P    160

AAGTGGAAGAACTATGGACTCCGTATTTTTGGTTTCATCCACCCGGCGAGGGACGGAGACGTCCAGTTTTCTGTGGCCTCAGACGACAACTCGGAGTTCTGGCTGAGTCTGGACGAGAGC    600
K  W  K  N  Y  G  L  R  I  F  G  F  I  H  P  A  R  D  G  D  V  Q  F  S  V  A  S  D  D  N  S  E  F  W  L  S  L  D  E  S    200

CCTGCTGCTGCCCAGCTTGTGGCCTTTGTGGGCAAGACTGGCTCCGAGTGGACAGCGCCTGGAGAATTCACCAAGTTCAGCTCCCAGGTGTCCAAGCCCAGGCGGCTCATGGCCTCCCGG    720
P  A  A  A  Q  L  V  A  F  V  G  K  T  G  S  E  W  T  A  P  G  E  F  T  K  F  S  S  Q  V  S  K  P  R  R  L  M  A  S  R    240

AGGTACTACTTTGAGTTGCTGCACAAGCAGGACGACCGCGGCTCGGACCACGTGGAAGTGGGCTGGCGAGCTTTCCTGCCCGGCCTGAAGTTCGAGGTCATCAGCTCTGCTCACATCTCC    840
R  Y  Y  F  E  L  L  H  K  Q  D  D  R  G  S  D  H  V  E  V  G  W  R  A  F  L  P  G  L  K  F  E  V  I  S  S  A  H  I  S    280

CTGTACACAGATGAGTCAGCCTTGAAGATGGACCACGTGGCGCACGTCCCCCAGTCTCCAGCCAGCCACGTGGGGGGGCGTCCGCCGCAGGAGGAGACCAGCGCAGACATGCTGCGGCCA    960
L  Y  T  D  E  S  A  L  K  M  D  H  V  A  H  V  P  Q  S  P  A  S  H  V  G  G  R  P  P  Q  E  E  T  S  A  D  M  L  R  P    320

GATCCCAGGGATACCTTTTTCCTCACTCCACGCATGGAATCTTCGAGCCTGGAGAACGTGCTGGAGCCCTGCGCCTACGCCCCCACCTACGTGGTCAAGGACTTCCCGATCGCCAGATAC   1080
D  P  R  D  T  F  F  L  T  P  R  M  E  S  S  S  L  E  N  V  L  E  P  C  A  Y  A  P  T  Y  V  V  K  D  F  P  I  A  R  Y    360

CAGGGCCTGCAATTTGTGTACCTGTCCTTCGTTTATCCCAACGACTACACTCGCCTCACCCACATGGAGACGGACAACAAGTGCTTCTACCGCGAGTCTCCGCTGTATCTGGAGAGGTTT   1200
Q  G  L  Q  F  V  Y  L  S  F  V  Y  P  N  D  Y  T  R  L  T  H  M  E  T  D  N  K  C  F  Y  R  E  S  P  L  Y  L  E  R  F    400

GGGTTCTATAAATACATGAAGATGGACAAGGAGGAGGGGGATGAGGATGAAGAAGACGAGGTGCAGCGCCGAGCCTTCCTCTTCCTCAACCCGGACGACTTCCTGGACGACGAGGACGAG   1320
G  F  Y  K  Y  M  K  M  D  K  E  E  G  D  E  D  E  E  D  E  V  Q  R  R  A  F  L  F  L  N  P  D  D  F  L  D  D  E  D  E    440

GGGGAGCTGCTCGACAGCCTGGAGCCCACCGAGGCGGCCCCGCCCAGGAGCGGCCCCCAGTCCCCCGCCCCAGCAGCCCCCGCCCAGCCCGGAGCCACCCTCGCCCCGCCGACCCCTCCC   1440
G  E  L  L  D  S  L  E  P  T  E  A  A  P  P  R  S  G  P  Q  S  P  A  P  A  A  P  A  Q  P  G  A  T  L  A  P  P  T  P  P    480

CGCCCCCGGGACGGGGGGACCCCCAGGCACTCCCGGGCCCTGAGCTGGGCCGCCAGGGCCGCCCGCCCTTTGCCGCTCTTCTTGGGCCGAGCTCCGCCCCCGCGCCCTGCAGTGGAGCAG   1560
R  P  R  D  G  G  T  P  R  H  S  R  A  L  S  W  A  A  R  A  A  R  P  L  P  L  F  L  G  R  A  P  P  P  R  P  A  V  E  Q    520

CCGCCCCCAAAGGTGTACGTGACCAGGGTGCGGCCGGGACAGCGGGCATCCCCCCGGGCCCCAGCGCCGCGTGCGCCCTGGCCGCCCTTCCCTGGCGTCTTCCTGCACCCCAGGCCTCTG   1680
P  P  P  K  V  Y  V  T  R  V  R  P  G  Q  R  A  S  P  R  A  P  A  P  R  A  P  W  P  P  F  P  G  V  F  L  H  P  R  P  L    560

CCCAGAGTGCAGCTGCGGGCGCCCCCACGCCCACCCCGGCCCCACGGCCGCAGGACCGGCGGCCCCCAGGCCACACAGCCGAGGCCCCCAGCCCGGGCGCAGGCCACCCAAGGGGGCCGG   1800
P  R  V  Q  L  R  A  P  P  R  P  P  R  P  H  G  R  R  T  G  G  P  Q  A  T  Q  P  R  P  P  A  R  A  Q  A  T  Q  G  G  R    600

GAGGGCCAGGCGCGCACGCTGGGACCTGCGGCGCCCACAGTGGACTCAAACTTGTCCTCCGAAGCGCGGCCCGTGACCTCCTTCCTGAGCTTGTCCCAGGTGTCCGGGCCGCAGCTGCCC   1920
E  G  Q  A  R  T  L  G  P  A  A  P  T  V  D  S  Ⓝ  L  S  S  E  A  R  P  V  T  S  F  L  S  L  S  Q  V  S  G  P  Q  L  P    640

GGGGAGGGCGAAGAGGAGGAGGAAGGGGAGGACGATGGGGCCCCGGGCGACGAGGCCGCGTCGGAGGACAGCGAGGAGGCCGCGGGCCCGGCGCTCGGACGCTGGCGTGAGGACGCCATC   2040
G  E  G  E  E  E  E  E  G  E  D  D  G  A  P  G  D  E  A  A  S  E  D  S  E  E  A  A  G  P  A  L  G  R  W  R  E  D  A  I    680

GACTGGCAGCGCACGTTCAGCGTGGGCGCCGTGGACTTCGAGCTGCTGCGCTCGGACTGGAACGACCTGCGATGCAACGTTTCGGGGAACCTGCAGCTGCCGGAGGCGGAGGCCGTGGAC   2160
D  W  Q  R  T  F  S  V  G  A  V  D  F  E  L  L  R  S  D  W  N  D  L  R  C  Ⓝ  V  S  G  N  L  Q  L  P  E  A  E  A  V  D    720

GTGACCGCTCAGTACATGGAGCGGCTGAACGCGCGCCACGGCGGGCGCTTCGCGCTTCTGCGCATCGTGAACGTGGAGAAGCGCCGGGACTCGGCGCGAGGGAGTCGCTTCCTGCTGGAG   2280
V  T  A  Q  Y  M  E  R  L  N  A  R  H  G  G  R  F  A  L  L  R  I  V  N  V  E  K  R  R  D  S  A  R  G  S  R  F  L  L  E    760

CTGGAGCTGCAGGAGCGCGGGGGCGGCCGCCTGCGACTGTCCGAGTACGTCTTCCTGCGGCTGCCGGGAGCCCGCGTAGGGGATGCAGACGGAGAAAGTCCCGAACCCGCTCCCGCCGCC   2400
L  E  L  Q  E  R  G  G  G  R  L  R  L  S  E  Y  V  F  L  R  L  P  G  A  R  V  G  D  A  D  G  E  S  P  E  P  A  P  A  A    800

TCCGTGCGCCCCGACGGCCGCCCCGAGCTCTGCCGGCCACTGCGCCTGGCCTGGCGCCAGGACGTGATGGTTCACTTCATCGTGCCAGTGAAAAACCAGGCACGGTGGGTGGCACAGTTC   2520
S  V  R  P  D  G  R  P  E  L  C  R  P  L  R  L  A  W  R  Q  D  V  M  V  H  F  I  V  P  V  K  N  Q  A  R  W  V  A  Q  F    840

CTGGCGGACATGGCTGCGCTGCACGCGCGCACCGGGGACTCGCGTTTCAGCGTCGTCCTGGTGGATTTCGAGAGCGAGGATATGGACGTGGAGCGGGCCCTGCGCGCCGCGCGCCTGCCC   2640
L  A  D  M  A  A  L  H  A  R  T  G  D  S  R  F  S  V  V  L  V  D  F  E  S  E  D  M  D  V  E  R  A  L  R  A  A  R  L  P    880

CGGTACCAGTACCTGAGACGAACCGGGAACTTCGAGCGCTCCGCCGGGCTGCAGGCGGGAGTGGACGCGGTAGAGGACGCCAGCAGCATCGTGTTCCTCTGCGACCTGCACATCCACTTC   2760
R  Y  Q  Y  L  R  R  T  G  N  F  E  R  S  A  G  L  Q  A  G  V  D  A  V  E  D  A  S  S  I  V  F  L  C  D  L  H  I  H  F    920

CCACCCAACATCCTGGACGGCATCCGCAAGCACTGCGTGGAGGGCAGGCTGGCCTTCGCGCCCGTGGTCATGCGCCTGAGCTCGCGGAGCTCGCCCCGGGACCCCCACGGTTACTGGGAG   2880
P  P  N  I  L  D  G  I  R  K  H  C  V  E  G  R  L  A  F  A  P  V  V  M  R  L  S  C  G  S  S  P  R  D  P  H  G  Y  W  E    960

GTGAACGGCTTTGGCCTTTTTGGGATCTACAAGTCGGACTTTGACCGGGTTGGAGGAATGAACACGGAGGAGTTCCGAGACCAGTGGGGGGGTGAAGACTGGGAGCTCCTGGACAGGGTC   3000
V  N  G  F  G  L  F  G  I  Y  K  S  D  F  D  R  V  G  G  M  N  T  E  E  F  R  D  Q  W  G  G  E  D  W  E  L  L  D  R  V   1000

CTGCAGGCAGGGCTGGAGGTGGAGCGGCTCCGACTGCGGAATTTCTATCACCACTACCACTCCAAGAGGGGCATGTGGAGCGTCCGCAGCAGGAAGGGCTCTCGCACGGGGGCGTCTTGA   3120
L  Q  A  G  L  E  V  E  R  L  R  L  R  N  F  Y  H  H  Y  H  S  K  R  G  M  W  S  V  R  S  R  K  G  S  R  T  G  A  S  *    1039
```

Fig. 1. Nucleotide and amino acid sequences of β4GalNAc-T4. The putative transmembrane domain is underlined. The DXH motif is written in bold. The β1,4-glycosyltransferase (β4GT) motif is boxed. Possible *N*-glycosylation sites are indicated by open circles.

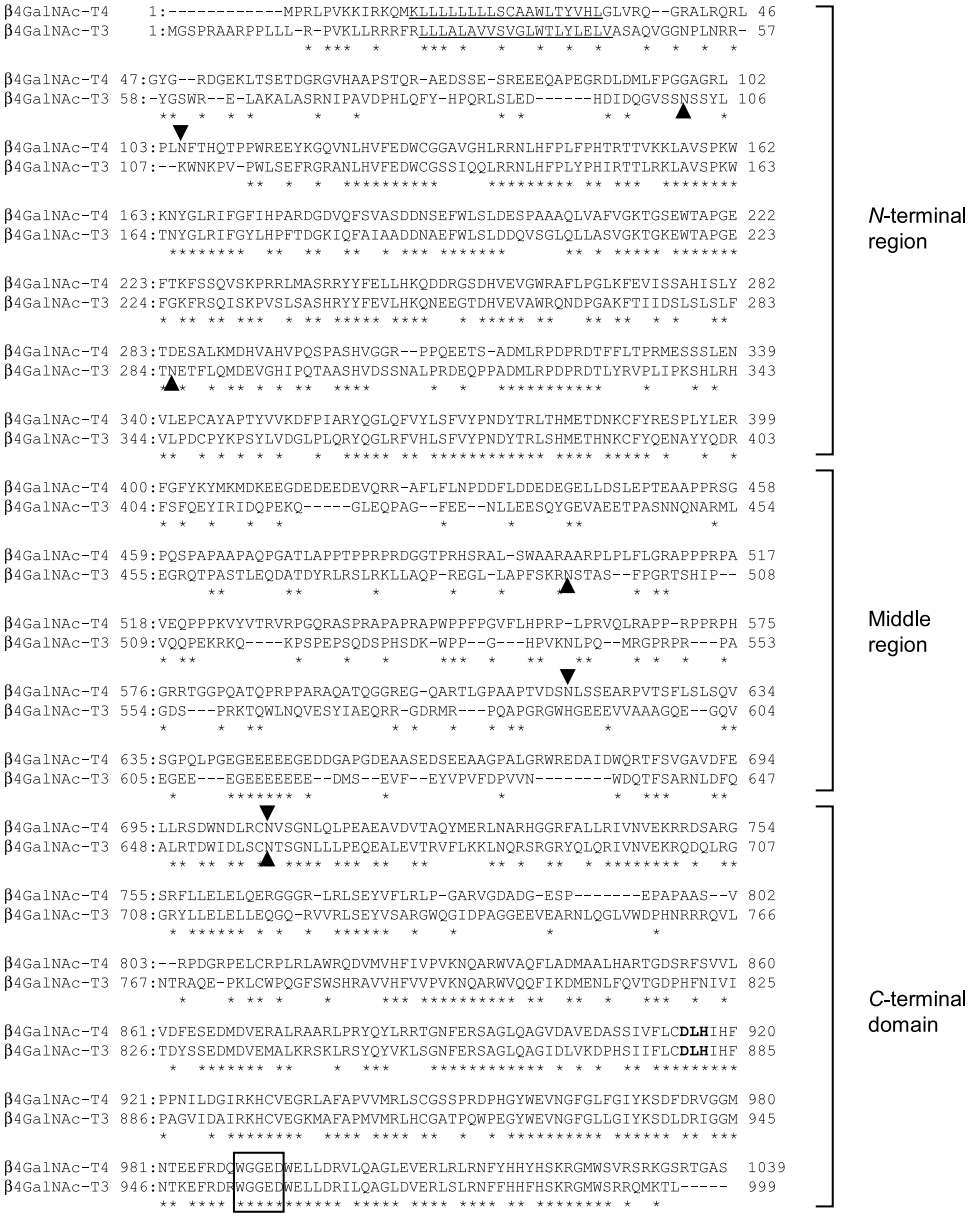

Fig. 2. Multiple alignment of amino acid sequences of β4GalNAc-T4 and T3 performed with GENETYX. Introduced gaps are shown with hyphens. The putative transmembrane domains are underlined. The D*X*H motifs are written in bold. The β4GT motif, WGGED, is boxed. Identical amino acids are shown with asterisks. Possible *N*-glycosylation sites are indicated by arrowheads.

ficient for an analysis of the exon/intron structure. This ORF consisted of 3120 bp encoding a predicted 1039 amino acid protein with a typical type II topology, which is common in glycosyltransferases. It contained three potential *N*-glycosylation sites and a D*X*H sequence, which is conserved in UDP-GalNAc:polypeptide-GalNAc-Ts, and is thought to participate in divalent cation binding. In amino acid sequence, β4GalNAc-T4 showed 42.6% identity with β4GalNAc-T3, with highly conserved sequences in the N- and C-terminal regions (Fig. 2). The middle region contained many acidic amino acids and proline-rich stretches, and showed a relatively low homology between the two enzymes. This region, however, showed no homology with any identified proteins except for β4GalNAc-T3. The β4GT motif, WGGED, is present in the C-terminal region. A homologous gene was found in the mouse (GenBank™ accession number AB114827) (data not shown). It has 83.7% identity with human β4GalNAc-T4. The mouse gene is probably orthologous to the human *β4GalNAc-T4* gene, and encoded a hypothetical 1042 amino acid protein carrying three possible *N*-glycosylation sites, a D*X*H sequence and a β4GT motif.

3.2. Substrate specificity of β4GalNAc-T4

We determined the substrate specificity of the truncated and soluble β4GalNAc-T4 expressed in HEK293T cells. Utilizing a variety of UDP donors and monosaccharide acceptors with a *para*-nitrophenyl (*p*Np) or benzyl (Bz) group, donor and acceptor substrates for β4GalNAc-T4 were screened with high performance liquid chromatography (HPLC). As summarized in Table 1, β4GalNAc-T4 transferred GalNAc to the non-reducing end of GlcNAcβ as did β4GalNAc-T3. The relative levels of activity toward various acceptors were

Table 2
Substrate specificity of β4GalNAc-Ts

	Acceptor substrate	Relative activity (%)	
		β4GalNAc-T4	β4GalNAc-T3
1.	GlcNAcβ1-2Manα1–6 GlcNAcβ1-2Manα1–3Manβ1-4GlcNAcβ1-4 GlcNAc-PA	100	100
2.	Fucα1–6 GlcNAcβ1-2Manα1–6 GlcNAcβ1-2Manα1–3Manβ1-4GlcNAcβ1-4 GlcNAc-PA	76.8	87.1
3.	Galβ1-4GlcNAcβ1-2Manα1–6 GlcNAcβ1-2Manα1–3Manβ1-4GlcNAcβ1-4 GlcNAc-PA	26.2	45.0
4.	Galβ1-4GlcNAcβ1-2Manα1–6 Fucα1–6 GlcNAcβ1-2Manα1–3Manβ1-4GlcNAcβ1-4 GlcNAc-PA	26.7	51.7
5.	GlcNAcβ1-2Manα1–6 Galβ1-4GlcNAcβ1-2Manα1–3Manβ1-4GlcNAcβ1-4 GlcNAc-PA	16.2	21.6
6.	Fucα1–6 Galβ1-4GlcNAcβ1-2Manα1–6 Galβ1-4GlcNAcβ1-2Manα1–3Manβ1-4GlcNAcβ1-4 GlcNAc-PA	3.4	5.0

similar between the two enzymes. A peak appeared at 24.8 min, in addition to the acceptor substrate peak at 30.3 min (data not shown) when UDP-GalNAc and GlcNAcβ-Bz were used as a donor and an acceptor substrate, respectively. To confirm that the additional peak was the real reaction product of β4GalNAc-T4, it was isolated by reversed-phase HPLC and identified with MALDI-TOF MS. In the positive-ion mode, peaks of 515.221, 537.193 and 553.165 *m/z* appeared. They were the same molecular mass as GalNAc-GlcNAc-Bz, and GalNAc-GlcNAc-Bz with Na^+ and with K^+, respectively (data not shown). However, no activity was observed with any other donor and monosaccharide acceptor combination (Table 1, numbers 1–13).

To determine the glycosidic linkage of the reaction product of β4GalNAc-T4, ^1H NMR spectroscopy was employed. In the previous study, we determined the structure of the β4GalNAc-T3 product as GalNAcβ1-4GlcNAc-*O*-Bz by means of NMR spectroscopy [18]. In the present study, the chemical shifts from the β4GalNAc-T4 product, GalNAc-GlcNAc-*O*-Bz, were compared to those of GalNAcβ1-4GlcNAc-*O*-Bz produced by β4GalNAc-T3. The NMR signals obtained from the β4GalNAc-T4 product were almost identical to those from GalNAcβ1-4GlcNAc-*O*-Bz. Thus, we concluded that the product of β4GalNAc-T4 has a β1-4 linkage like the β4GalNAc-T3 product (data not shown).

3.3. Comparison of acceptor substrates

In the previous paper, we determined the specificity of β4GalNAc-T3 for *O*- and *N*-glycans [18]. In the present study, we compared the specificities of two β4GalNAc-Ts toward *O*- and *N*-glycan acceptors. In brief, the enzymes showed very similar specificities for *O*- and *N*-glycans, although the activity of β4GalNAc-T3 was 3.4-fold higher than that of T4 (Table 1, numbers 14–16, and Table 2). Among the *O*-glycan acceptor substrates examined, the most effective was core 6-*p*Np, although it is a rare structure in *O*-glycans. The activity of β4GalNAc-T4 toward core 6-*p*Np was 1.9-fold higher than that toward GlcNAcβ-Bz. Core 2-*p*Np (Galβ1-3(GlcNAcβ1-6)GalNAcα1-*p*Np) and core 3-*p*Np (GlcNAcβ1-3GalNAcα1-*p*Np) could be acceptors, even with a relatively low efficiency, for β4GalNAc-T4 as for T3. Among the *N*-glycans examined, the most efficient substrate for both enzymes was a non-fucosylated bi-antennary form (number 1 in Table 2). The efficiency decreased as the number of antennas increased (data not shown). The presence of an α1,6 Fuc residue had little effect in decreasing the activity of the enzymes (compare results between numbers 1 and 2, numbers 3 and 4, and numbers 5 and 6). β4GalNAc-T4 preferred the GlcNAcβ1-2Manα1-3 antenna to the GlcNAcβ1-2Manα1-6 antenna as an acceptor substrate as did β4GalNAc-T3 (numbers 3 and 5, and 4 and 6).

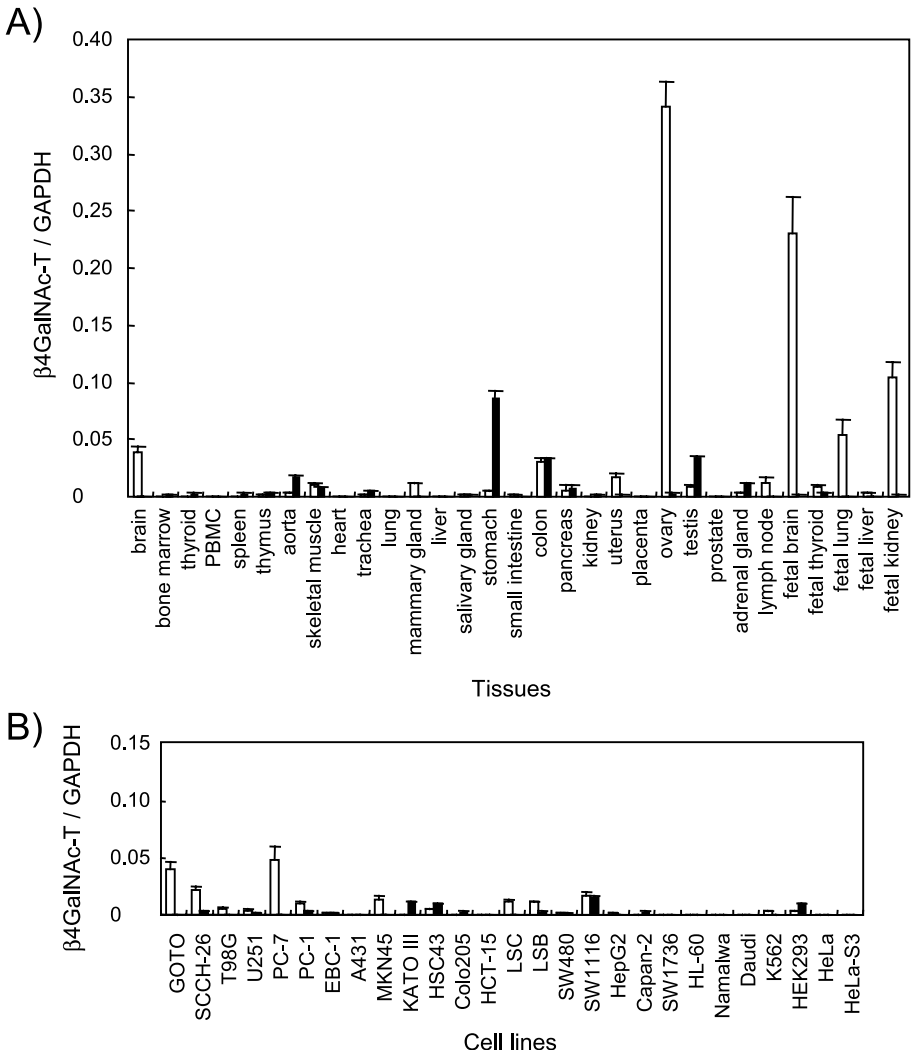

Fig. 3. Quantitative real-time PCR analysis of the *β4GalNAc-T4* and *T3* transcripts in human tissues and cell lines. Standard curves for β4GalNAc-T4, T3 and GAPDH were generated by serial dilution of each plasmid DNA. The expression levels of the *β4GalNAc-T4* (open bars) and *T3* (closed bars) transcripts were normalized to the level of *GAPDH* transcript, which was measured in the same cDNAs from human tissues (A) and cell lines (B). Data were obtained from triplicate experiments and are indicated as the mean ± S.D. *PBMC*, peripheral blood mononuclear cells; *GOTO* and *SCCH-26*, neuroblastomas; *T98G* and *U251*, glioblastomas; *PC-7*, lung adenocarcinoma; *PC-1* and *EBC-1*, lung squamous cells; *A431*, esophagus cancer; *MKN45*, *KATO III* and *HSC43*, stomach cancers; *Colo205*, *HCT15*, *LSC*, *LSB*, *SW480* and *SW1116*, colorectal cancer; *HepG2*, hepatocarcinoma; *Capan-2*, pancreas cancer; *SW1736*, thyroid cancer; *HL-60*, promyelocytic leukemia; *Namalwa*, B cell lymphoma; *Daudi*, B cell (Burkitt's) lymphoma; *K562*, erythroid leukemia; *HEK293*, embryonic kidney cell; *HeLa* and *HeLa-S3*, cervix cancers.

3.4. *Quantitative analysis of the* β4GalNAc-T4 *and* T3 *transcripts in human tissues and cell lines by real-time PCR*

In the present study, we determined the tissue distribution and expression levels of the *β4GalNAc-T4* transcript by the real-time PCR method, and compared them with those of the *β4GalNAc-T3* transcript reported in our previous study. The results are shown as the relative amount versus the *GAPDH* transcript in Fig. 3. Interestingly, β4GalNAc-T3 transcript was found to be very faint, while T4 was significantly expressed in adult and fetal brain (Fig. 3A). β4GalNAc-T4 was most highly expressed in ovary followed by fetal brain and various adult brain tissues. It was also highly expressed in fetal kidney and lung. β4GalNAc-T3 was expressed at an extremely low level in these tissues, while it was highly expressed in adult stomach, colon and testis. The expression levels of both enzymes in a variety of cell lines were comparatively low (Fig. 3B), although some cell lines, PC-7 (lung adenocarcinoma) and GOTO (neuroblastoma) cells, showed a relatively high level of β4GalNAc-T4.

4. Discussion

We found a novel β4GalNAc-T, named β4GalNAc-T4, carrying a β4GT motif, WGGED, in its C-terminus. Its amino acid sequence was homologous to β4GalNAc-T3 with 42.6% identity. The truncated protein produced in HEK293T cells showed β4GalNAc-T activity toward GlcNAc, resulting in the synthesis of LacdiNAc, as expected. Its ortholog was also found in mouse. The specificity of β4GalNAc-T4 toward oligosaccharides was quite similar to that of β4GalNAc-T3, however, the tissue distribution of the two was different. This is the second report of the molecular cloning and characterization of a mammalian LacdiNAc synthase.

The *β4GalNAc-T4* gene was found to be located at 11p15, i.e. on a different chromosomal locus from the *β4GalNAc-T3* gene at 12p13.3. The genomic information for *β4GalNAc-T4* was insufficient for the analysis of the exon/intron structure. The *β4GalNAc-T4* gene encodes a putative 1039 amino acid protein which is unusually long in comparison with the other glycosyltransferases. Comparing amino acid sequences between the two β4GalNAc-Ts, the N- and C-terminal regions were found to be more homologous than the middle region (Fig. 2). The middle of both β4GalNAc-T4 and T3 contains an unusual sequence in which numerous Pro residues and acidic amino acids, Glu and Asp, are present. The catalytic activity of both β4GalNAc-Ts is probably directed by the C-terminal region which contains the β4GT motif. Functions of the N-terminal and middle regions remain to be elucidated.

This enzyme showed GalNAc-T activity toward GlcNAcβ-Bz as did β4GalNAc-T3 [18]. The product of β4GalNAc-T4, GalNAc-GlcNAcβ-Bz, showed the same spectrum as β4GalNAc-T3 on one-dimensional ^{1}H NMR spectroscopy as reported previously by us [18]. Thus, the reaction product was determined as GalNAcβ1-4GlcNAc-*O*-Bz (data not shown). Comparing the acceptor specificity between the two β4GalNAc-Ts, their preference for oligosaccharides was remarkably similar, that is, core 6-*p*Np and GlcNAcβ-Bz were the preferred acceptors for both enzymes (Table 1). Furthermore, all oligosaccharides containing non-reducing terminal GlcNAcβ residues could be acceptors for both β4GalNAc-Ts. Core 6 was preferable as an acceptor to core 2. The Gal residue of core 2 might be obstructive to β4GalNAc-Ts. The similar preference has been reported in β4Gal-Ts. Ujita et al. reported that core 2 branch was not a suitable acceptor for β4Gal-T1 [24]. Thus, the activities of these β4GTs, i.e. β4Gal-Ts and β4GalNAc-Ts, might be influenced by steric hindrance due to the presence or absence of the Gal residue. Core 6 differs from core 3 only in the linkage between GlcNAc and GalNAc. The former forms GlcNAcβ1-6GalNAc whereas the latter forms GlcNAcβ1-3GalNAc. Both β4GalNAc-Ts can recognize this difference, and prefer core 6 to core 3. Core 6 is not a major core structure of *O*-glycan, in contrast to core 2 and 3 which are widely found as major core structures in digestive organs, such as stomach and colon. However, the core 6 structure and core 6 synthesizing activity have been found in ovarian cyst fluid, seminal fluid and meconium [26–28]. Core 6 in such tissues may be a physiological substrate for β4GalNAcTs, because β4GalNAc-Ts were found to be expressed in ovary and testis in the present study. In the case of mouse, core 2 carrying LacdiNAc has been found in mouse eggs [17]. This structure might be synthesized by β4GalNAc-T3 in ovary. Very recently, we cloned β3GalNAc-T2, which transfers GalNAc to GlcNAc with a β1-3 linkage, and prefers core 2 to core 3 or core 6 as an acceptor [25]. It is possible that β3GalNAc-T2 competes with β4GalNAc-Ts for core 2 containing acceptors in ovary and some tissues, because both enzymes are expressed in such tissues. However, the GalNAcβ1-3GlcNAc structure has not been found in mammalian tissues.

For the oligosaccharide acceptors of *N*-glycans, the GalNAc transfer efficiency of β4GalNAc-T4 and T3 decreased as the number of antennas of *N*-glycan increased (data not shown). The GlcNAcβ1-2Manα1-3 antenna was preferred to the GlcNAcβ1-2Manα1-6 antenna. Dell et al. recently found the LacdiNAc structures on *N*-glycans, however, they did not determine which antenna has the structures [2]. The differential tissue distribution of β4GalNAc-T4 and T3 suggested that T4 is responsible for the in vivo synthesis of LacdiNAc on GdA and LH. In adult and fetal brain, β4GalNAc-T4 was expressed at high levels, while T3 was not. β4GalNAc-T4 is probably responsible for the synthesis of LacdiNAc in the *N*-glycan of tenascin-R which is present in the central nervous system, including the cerebellum [19]. β4GalNAc-T4 was also expressed in some fetal tissues, such as kidney and lung, while T3 was expressed in stomach, colon and testis. In such tissues, the proteins carrying LacdiNAc have yet to be identified. It is of interest to find the proteins carrying LacdiNAc in these tissues, and to investigate their specific functions.

Acknowledgements: This work was performed as part of the R&D Project of the Industrial Science and Technology Frontier Program (R&D for Establishment and Utilization of a Technical Infrastructure for Japanese Industry) supported by the New Energy and Industrial Technology Development Organization.

References

[1] Hiyama, J., Weisshaar, G. and Renwick, A.G. (1992) Glycobiology 2, 401–409.
[2] Dell, A. et al. (1995) J. Biol. Chem. 270, 24116–24126.
[3] Smith, P.L., Skelton, T.P., Fiete, D., Dharmesh, S.M., Beranek, M.C., MacPhail, L., Broze Jr., G.J. and Baenziger, J.U. (1992) J. Biol. Chem. 267, 19140–19146.
[4] Pierce, J.G. and Parsons, T.F. (1981) Annu. Rev. Biochem. 50, 465–495.
[5] Smith, P.L. and Baenziger, J.U. (1988) Science 242, 930–933.
[6] Green, E.D. and Baenziger, J.U. (1988) J. Biol. Chem. 263, 25–35.
[7] Fiete, D., Srivastava, V., Hindsgaul, O. and Baenziger, J.U. (1991) Cell 67, 1103–1110.
[8] Julkunen, M., Koistinen, R., Sjoberg, J., Rutanen, E.M., Wahlstrom, T. and Seppala, M. (1986) Endocrinology 118, 1782–1786.
[9] Kamarainen, M., Riittinen, L., Seppala, M., Palotie, A. and Andersson, L.C. (1994) Blood 84, 467–473.
[10] Morrow, D.M. et al. (1994) Am. J. Pathol. 145, 1485–1495.
[11] Koistinen, H., Koistinen, R., Kamarainen, M., Salo, J. and Seppala, M. (1997) Lab. Invest. 76, 683–690.
[12] Seppala, M., Koistinen, H. and Koistinen, R. (2001) Trends Endocrinol. Metab. 12, 111–117.
[13] Koistinen, H. et al. (1996) Mol. Hum. Reprod. 2, 759–765.
[14] Morris, H.R. et al. (1996) J. Biol. Chem. 271, 32159–32167.
[15] Oehninger, S., Coddington, C.C., Hodgen, G.D. and Seppala, M. (1995) Fertil. Steril. 63, 377–383.
[16] Clark, G.F., Oehninger, S., Patankar, M.S., Koistinen, R., Dell, A., Morris, H.R., Koistinen, H. and Seppala, M. (1996) Hum. Reprod. 11, 467–473.
[17] Dell, A. et al. (2003) Proc. Natl. Acad. Sci. USA 100, 15631–15636.
[18] Sato, T. et al. (2003) J. Biol. Chem. 278, 47534–47544.
[19] Woodworth, A., Fiete, D. and Baenziger, J.U. (2002) J. Biol. Chem. 277, 50941–50947.
[20] Dharmesh, S.M., Skelton, T.P. and Baenziger, J.U. (1993) J. Biol. Chem. 268, 17096–17102.
[21] Iwai, T. et al. (2002) J. Biol. Chem. 277, 12802–12809.
[22] Gotoh, M. et al. (2002) J. Biol. Chem. 277, 38179–38188.
[23] Kutyavin, I.V. et al. (2000) Nucleic Acids Res. 28, 655–661.
[24] Ujita, M., McAuliffe, J., Schwientek, T., Almeida, R., Hindsgaul, O., Clausen, H. and Fukuda, M. (1998) J. Biol. Chem. 273, 34843–34849.
[25] Hiruma, T. et al. (2004) J. Biol. Chem. (in press).
[26] Feeney, J., Frenkiel, T.A. and Hounsell, E.F. (1986) Carbohydr. Res. 152, 63–72.
[27] Wu, A.M. (1988) Adv. Exp. Med. Biol. 228, 351–394.
[28] Yazawa, S., Abbas, S.A., Madiyalakan, R., Barlow, J.J. and Matta, K.L. (1986) Carbohydr. Res. 149, 241–252.