Genomic Organization and Chromosomal Assignment of the Human β1,4-N-Acetylgalactosaminyltransferase Gene

IDENTIFICATION OF MULTIPLE TRANSCRIPTION UNITS*

(Received for publication, December 11, 1995, and in revised form, June 5, 1996)

Keiko Furukawa**, Hidenobu Soejima†, Norio Niikawa§, Hiroshi Shiku¶, and Koichi Furukawa‡

From the Departments of †Oncology and ‡Human Genetics, Nagasaki University School of Medicine, 1-12-5 Sakamoto, Nagasaki 852 and the §Department of Internal Medicine II, Mie University School of Medicine, Edohashi, Tsu, Mie 514, Japan

The β1,4-N-acetylgalactosaminyltransferase (β1,4GalNAc-T) (EC 2.4.1.92) gene is expressed in normal brain tissues and in various malignant transformed cells, such as malignant melanoma, neuroblastoma, and adult T cell leukemia. To analyze the regulatory mechanisms of gene expression, we determined the genomic organization of the β1,4GalNAc-T gene. The gene consists of at least 11 exons and spans >8 kilobase pairs. The coding region is located in exons 2-11. To determine the transcription initiation sites, 5'-rapid amplification of cDNA ends analysis and ribonuclease protection assays were performed using RNA obtained from the human melanoma cell line SK-MEL-31. Consequently, we defined three transcription initiation sites and the alternative usage of three exons. Exons 1a and 1b partially overlap; the latter is part (3'-side) of the former and corresponds to the 5'-noncoding region of the cDNA clone previously isolated. The third transcript, exon 1c, corresponds to nucleotides -520 to -412 (position +1 = A of ATG of β1,4GalNAc-T cDNA), which are considered to be in intron 1 based on the cloned cDNA sequence. Ribonuclease protection assays revealed the corresponding protection bands in samples of the gene-expressing cell lines. 5'-Flanking regions of individual initiation sites showed promoter activity when analyzed by chloramphenicol acetyltransferase assay in SK-MEL-31 cells. The multiple transcription initiation sites and their promoters/enhancers identified here might be differentially involved in the cell type-specific expression of the β1,4GalNAc-T gene. This gene was assigned to human chromosome 12q13.3 by means of fluorescence in situ hybridization.

Glycosphingolipids are amphipathic molecules consisting of a hydrophilic carbohydrate moiety and lipophilic ceramides (1). They are expressed mainly on the cell membrane and have various roles in cell-cell or cell-extracellular matrix recognition and in the regulation of signaling events (2). Although carbohydrate structures on the glycosphingolipids as well as those on glycoproteins are enormously diverse, they are usually characterized by different stages of development, a distinct cell lineage, and various steps of malignant transformation (3). These cell type-specific profiles of glycolipid components are determined by a combination of the glycosylation machineries expressed in individual tissues and cells (4). However, little is known about the mechanisms regulating glycosylation systems (5).

Studies of the regulation of carbohydrate synthesis and expression have been hampered by the lack of cloned glycosyltransferase genes. However, the isolation of cDNAs of these genes (6) has improved the understanding of the genomic organization and the control of the cell type-specific expression of glycosyltransferases. The structure and tissue-specific expression of the β-galactoside α2,6-sialyltransferase (7–10) and β,1,4-galactosyltransferase (11–13) genes have been analyzed. However, the enzymes encoded by these two genes catalyze the synthesis of carbohydrate chains on various glycoproteins, and the glycosyltransferase genes that are mainly responsible for the synthesis of glycosphingolipids have not been studied.

The level of β1,4-N-acetylgalactosaminyltransferase (β1,4GalNAc-T) (Gm/Gd2 synthase; EC 2.4.1.92) activity is generally high in neuroectoderm-derived tumor cells, such as neuroblastomas and malignant melanomas (14). Northern blots and reverse transcription-polymerase chain reactions have revealed that the mRNA levels of the gene almost correlate with enzyme activities and Gm/Gd2 expression in individual cell lines (15). Moreover, the specific expression of Gd2 and the possible transactivation of the β1,4GalNAc-T gene by human T cell lymphotrophic virus type I p40tax protein in adult T cell leukemia cells has been reported (16). In addition, we demonstrated that the mRNA level of β1,4GalNAc-T in mouse brain gradually increases during development until birth (17), which corresponds with the reported changes in enzyme activity (18). These results indicate that the synthesis and expression of Gm/Gd2 are mainly controlled at the level of transcription of the β1,4GalNAc-T gene, with additional modification by many other epigenetic factors, and they suggest that there are complex regulatory systems of gene expression depending on the state of individual cells, such as neural development, malignant transformation or progression, and viral infection.

1 The abbreviations used are: β1,4GalNAc-T, β1,4-N-acetylgalactosaminyltransferase; FISH, fluorescence in situ hybridization; RACE, rapid amplification of cDNA ends; PCR, polymerase chain reaction; CAT, chloramphenicol acetyltransferase; bp, base pair(s).

2 Ganglioside nomenclature is based on that of Svennerholm (49).

*** Present address: Dept. of Biochemistry, Mie University School of Medicine.
To understand the basis for the cell type-specific expression of the β1,4GalNAc-T gene, we isolated genomic clones of the gene, analyzed the genomic organization, and defined multiple transcription initiation sites and their promoters. We also determined the chromosomal location of the β1,4GalNAc-T gene by fluorescence in situ hybridization (FISH).

EXPERIMENTAL PROCEDURES

Materials—Restriction enzymes used for the cloning and sequencing of DNA were purchased from Takara Shuzo Corp. (Kyoto, Japan) and Nippon Gene Corp. (Toyama, Japan). A library of a Sau3AI partial digest of human placental DNA cloned into EMBL-3 SP6/T7 (CLONTECH, Palo Alto, CA) was screened using probes derived from a human β1,4GalNAc-T cDNA clone (M2T1-1) (19). The human melanoma lines SK-MEL-31 and MeWo as well as the astrocytoma line AS were obtained from Dr. L. J. Old (Sloan-Kettering Cancer Center, New York).

Insertion and Characterization of Genomic Sequences—The amplified genomic library containing ~3 × 10^6 plaques was screened by hybridization with a radiolabeled XbaI fragment of human β1,4GalNAc-T cDNA (pM2T1-1) (19). Insert DNA fragments were initially characterized by restriction enzyme digestion and Southern blotting using oligonucleotide probes based on the cDNA sequences. Human genomic DNA fragments that hybridized with β1,4GalNAc-T cDNA probes were subcloned into pBluescript II SK+ (Stratagene). All genomic exon regions were completely sequenced to determine intron interruptions and to confirm that the genomic sequences were in complete agreement with known β1,4GalNAc-T cDNA (pM2T1-1) data (19). Double-stranded sequencing was performed using the Taq Dye Primer Cycle sequencing kit and a Mod270 DNA sequencing kit (both from Applied Biosystems, Foster City, CA).

5′-Rapid Amplification of cDNA Ends (RACE) Analysis—A modified RACE analysis was performed to clone gene-specific 5′-ends using the 5′-AmpliFINDER® RACE kit (CLONTECH) according to the manufacturer’s instructions. Firststrand cDNAs were synthesized by avian myeloblastosis virus reverse transcriptase from 2 μg of poly(A) mRNA of SK-MEL-31 with the gene-specific antisense reverse transcription primer corresponding to nucleotides +235 to +254 (5′-CTCGAGCTCAACAGCTGAGT-3′; position +1 = A of ATG of β1,4GalNAc-T DNA) (19) downstream of the ATG codon. RNA was digested with NaOH, and the remaining oligonucleotide primers were removed. The purified cDNA was ligated with the AmpliFINDER® anchor using T4 RNA ligase and then amplified by PCR using an anchor primer complementary to the anchor sequences combined with a nested antisense primer corresponding to nucleotides +98 to +118 (5′-ACGGCGGCAAAGG-TAGCGCGGA-3′) of β1,4GalNAc-T cDNA (see Fig. 2). The amplified product was cloned into pTBlueR (Novagen) and sequenced as described above.

RACE Protection Analysis—The cRNA probes were synthesized from subcloned templates. To construct the plasmid used to prepare RNA probe as (see Fig. 3A), a 5′-AvaiI-3′/BamHI fragment (nucleotides −1141 to −948 of β1,4GalNAc-T genomic DNA) in which the AvaiI fragment was blunt-ended with Klenow enzyme and ligated with a HindIII linker was cloned into the HindII and BamHI sites of pBluescript II SK−. For RACE protection analysis (see Fig. 3A), two genomic DNA fragments spanning the regions from nucleotides −823 to −638 (probe b) and from nucleotides −651 to −441 (probe c) were amplified by PCR. PCR products b and c were subcloned into the BamHI and XbaI sites of pBluescript II SK− and then into pTBlueR (Novagen), respectively. The PCR products were confirmed by sequencing. Total RNA was isolated from cell lines as described. RACE protection assay was performed using the ribonuclease protection assay kit RPA II (Ambion INC., Austin, TX) according to the manufacturer’s instructions. Total RNA (50 μg) was hybridized with 4 × 10^4 to 10^5 copies each of the [α-32P]UTP-labeled riboprobes at 55 °C overnight, followed by digestion with RNase A and RNase T1. The protected products were compared on a 6% sequencing gel with a known DNA sequencing ladder. The protected bands were detected by exposing the gel to an imaging plate (Bas III, Fuji Photo Film Co.) for 10–12 h.

Primer Extension—Primer extension analysis was performed using the avian myeloblastosis virus reverse transcriptase primer extension system (Promega). An oligonucleotide primer with a sequence complementary to exon 1c of the β1,4GalNAc-T gene (nucleotides −491 to −459) was radiolabeled at the 5′-end with T4 polynucleotide kinase and [γ-32P]ATP (3000 Ci/mmol). The labeled primer was added to 50 μg of total RNA isolated from SK-MEL-31, A5, or NALM6 in avian myeloblastosis virus primer extension buffer. The reaction mixtures were heated at 68 °C for 5 min, annealed at 58 °C for 90 min, and then left at room temperature to cool for 10 min. Avian myeloblastosis virus reverse transcriptase was added, and primer extension was performed at 42 °C for 1 h. The reactions were terminated, and the extension products were precipitated and then analyzed on an 8% DNA sequencing gel. The size was determined based on a DNA sequencing ladder.

Chloramphenicol Acetyltransferase (CAT) Assay—Each fragment of the β1,4GalNAc-T gene, nucleotides −2228 (SalI) to −692 (BanII), −2228 to −948 (BanIII), −947 to −156 (BglI), and −947 to −156, were ligated to the CAT reporter gene (see Figs. 5A and 6A). To construct pGT-CAT−2228 (SalI)−692 (BanII) and pGT-CAT−2228 (SalI)−156 (BglI), the BanII and BglII sites were blunt-ended and ligated with a HindIII linker; then, the SalI/BanII (Δ) means a modified restriction enzyme site) or SalI/BglII fragment of the genomic DNA was ligated into the SalI and HindIII sites of pSV000CAT (Nippon Gene Corp.). For pGT-CAT−947/−692 and pGT-CAT−947/−156, fragment −2228 to −948 was deleted from pGT-CAT−2228/−692 and pGT-CAT−2228/−156, respectively. Plasmid pGT-CAT−2228/−948 was constructed by deleting fragment −947 to −692 from pGT-CAT−2228/−692. For constructs pGT-CAT−662/−441 and pGT-CAT−441/−662 (see Fig. 6B), the region between nucleotides −662 and −441 of the β1,4GalNAc-T gene was amplified by PCR and sequenced for confirmation. This PCR product (fragment −662 to −441) was ligated with the promoterless CAT gene in the sense or antisense direction. To prepare 5′-deletion constructs, fragment −2228 to −156 of the β1,4GalNAc-T gene linked to the CAT gene was ligated into the XbaI and SalI sites of pUC19 to make convenient restriction sites for deletion and was then deleted from the 5′-end (see Fig. 6B) using a deletion kit (Takara Shuzo Corp.). Cell lines were seeded at 1.5–2.0 × 10^5/60-mm dish and were transfected using the calcium phosphate procedure as described (21). Briefly, the cells were transfected with 7 μg of DNA mixed with 3 μg of the pSV-β-galactosidase plasmid (Promega) or a luciferase expression vector (pCEV/Luc) presented by Dr. J. Takeda (Osaka University) to quantify transfection efficiencies and incubated at 37 °C for 4–6 h. Thereafter, the growth medium was changed, and the cells were incubated at 37 °C for an additional 48 h, washed with phosphate-buffered saline, harvested, and assayed for CAT reporter gene activity.

| Table 1 | Exons of the human β1,4GalNAc-T gene | For each exon, the cDNA residues and encoded amino acids are given as numbered by Nagata et al. (19). |
|---|---|---|
| Exon No. | Exon length | Corresponding cDNA residues | Amino acids encoded |
| 1 | 59 | −60 to −2 | Noncoding |
| 2 | 219 | −1 to 218 | 1-72 |
| 3 | 165 | 219–318 | 73–127 |
| 4 | 107 | 384–490 | 129–167 |
| 5 | 41 | 491–531 | 165–172 |
| 6 | 181 | 532–712 | 178–237 |
| 7 | 99 | 713–811 | 238–270 |
| 8 | 191 | 812–1002 | 271–334 |
| 9 | 141 | 1003–1143 | 335–391 |
| 10 | 239 | 1144–1382 | 382–460 |
| 11 | >1000 | 1383 to >2450 | 461–561 |

Fig. 1. Genomic organization of human β1,4GalNAc-T. E1-E11 are exons defined based on the cloned cDNA (pM2T1-1). Restriction enzyme sites were mapped and are indicated as follows: B, BamHI; E, EcoRI; H, HindIII; K, KpnI; S, SalI. The translation initiation site is also shown (ATG). Open boxes indicate the noncoding region, and dotted boxes indicate the coding region. kb, kilobase.
gene activity by thin layer chromatography as described (21). β-Galactosidase activities were measured according to the manufacturer’s instructions (Promega). The luciferase assay was performed using Pica Gene (Toyo Ink Corp., Tokyo) according to the manufacturer’s instructions.

FISH—One of the phage clones (C4-1), in which the \( \beta_{1,4} \text{GalNAc-T} \) gene was inserted, was labeled with biotin-16-dUTP using a nick translation labeling kit (Boehringer Mannheim). FISH was performed on non- and R-banded normal human metaphase chromosomes using human Cot-1 DNA (Life Technologies, Inc.) as a competitor as described (22). FISH signals in 50 mitotic cells were detected with fluorescein isothiocyanate-conjugated avidin, and chromosomes were counterstained with propidium iodide. Photomicroscopy was performed under a fluorescence microscope using Nikon B-2A and B-2E filters.

### RESULTS

#### Genomic Analysis and Mapping of the \( \beta_{1,4} \text{GalNAc-T} \) Gene

We screened a human genomic library in the \( \lambda \) vector EMBL3 and isolated six clones (C2-1, C3-1, C4-1, C7-2, C8-1, and C10-1) based on restriction mapping. These clones hybridized with 5'-cDNA (nucleotides 60 to +327 of cDNA) and 3'-cDNA (nucleotides +2025 to +2450 of cDNA) probes after digestion by restriction enzymes. Fragments of clones C3-1 and C8-1 hybridized with both probes, so the C3-1 clone, which was longer than C8-1, was subcloned into pBluescript II SK and sequenced. The insert of clone C3-1 was found to span almost the entire \( \beta_{1,4} \text{GalNAc-T} \) gene, containing all exons corresponding to the 2.5-kilobase \( \beta_{1,4} \text{GalNAc-T} \) cDNA clone (pM2T1-1).

### Table II

Upper-case letters represent exon sequences, and boldface letters represent agreement with the consensus sequence reported by Breathnach and Chambon (23).

| Intron No. | Intron length | Splice junction sequence | Donor Acceptor |
|-----------|---------------|--------------------------|----------------|
| 1         | 650 bp        | A\text{CATG}:gtgagt      | \text{cccccccctag:GATG} |
| 2         | 570 bp        | TGGG:gtgagttcctcgag      | \text{GCGG:GCTG} |
| 3         | 112* bp       | C\text{GGAG}:gtgagtt      | \text{cccccccctag:GATG} |
| 4         | 430 bp        | C\text{GGAG}:gtgagtt      | \text{cccccccctag:GATG} |
| 5         | 145* bp       | C\text{GGAG}:gtgagtt      | \text{cccccccctag:GATG} |
| 6         | 1000 bp       | C\text{GGAG}:gtgagtt      | \text{cccccccctag:GATG} |
| 7         | 144* bp       | C\text{GGAG}:gtgagtt      | \text{cccccccctag:GATG} |
| 8         | 560* bp       | C\text{GGAG}:gtgagtt      | \text{cccccccctag:GATG} |
| 9         | 200 bp        | C\text{GGAG}:gtgagtt      | \text{cccccccctag:GATG} |
| 10        | 680 bp        | C\text{GGAG}:gtgagtt      | \text{cccccccctag:GATG} |

*The length of introns was determined by sequencing.

### Identification of the Transcription Initiation Sites

To define the transcription initiation sites for the \( \beta_{1,4} \text{GalNAc-T} \) gene, 5'-RACE analysis and RNase protection assays...
were performed (Figs. 2 and 3).

5'-RACE Assay—After PCR amplification of the 5'-ends of cDNAs prepared from SK-MEL-31 mRNA, two bands at −550 and 280 bp were obtained (data not shown). These extended cDNAs were cloned and Southern-blotted using oligonucleotide probes generated from exons 1 (nucleotides 1073 to −1036 of the genomic sequence as shown in Fig. 2B) and 2 (nucleotides 66 to +87 of cDNA) as shown in Fig. 1 to discriminate aberrant clones. The clones that hybridized with these probes were sequenced and compared with the genomic sequence of the β1,4GalNAc-T gene. The sequences of the 5'-end of 550-bp cDNA clones started at nucleotides −1073 to −1036 of the genomic sequence as shown in Fig. 2B. The sequences of the 5'-site of 280-bp cDNA clones matched either one of two portions of the genomic sequence. About 10% (5/48) of the 280-bp RACE products corresponded to the original exon 1 and started at nucleotides −814 to −809 or at −720 as shown in Fig. 2B. The remaining RACE clones (90%, 43/48) did not hybridize with an oligonucleotide probe derived from exon 1. The region of the cDNA sequences corresponding to exon 1 was replaced by a rather different sequence, which was found in intron 1 of the original genomic sequence corresponding to nucleotides −520 to −412 and termed exon 1c as shown in Fig. 2 (B and C). The sequence of the 3'-end (nucleotide −412) of exon 1c was adjacent to the original exon 2 junction and matched the consensus splice site sequences shown in Fig. 2B. Consequently, the results of 5'-RACE suggested the presence of three sites of transcription initiation as shown in Fig. 2C. The transcripts starting from exons 1a, 1b, and 1c were tentatively named a, b, and c, respectively. In transcripts a and b, exon 1b was completely covered by exon 1a, sharing −170 bp, and it might consist of two different initiation sites. Transcript c started from new exon 1c alternatively with the two former transcripts. Although three possible transcription initiation sites were detected by RACE analysis, all the transcripts contained suitable Kozak sequences (24–26) prior to the ATG codon of exon 2, resulting in identical single peptide sequences. Although other ATG codons are also found in exons 1a and 1c, they are not in a favorable context for initiating translation.

RNase Protection Assay—To confirm the results from 5'-RACE analysis, RNase protection assay using three cRNA probes surrounding exons 1a, 1b, and 1c was performed (Fig. 3A). Total RNA was prepared from the melanoma cell line SK-MEL-31, the astrocytoma cell line AS, and the pre-B cell lymphocytic leukemia cell line NALM6. SK-MEL-31 and AS cells express high levels of β1,4GalNAc-T mRNA, and NALM6 cells do not express this mRNA (15). As shown in Fig. 3B, the protected fragments were detected in SK-MEL-31 and AS cells and not in NALM6 cells or yeast tRNA using all three cRNA probes. The quality of RNA preparations used for RNase protection experiments was confirmed by Northern blotting using a β-actin probe (Fig. 3B, panel d). A 139-base segment in cRNA probe a was protected. A 204-base segment in cRNA probe b and a 130-base segment in cRNA probe c were also protected.

Since RNA has less mobility than DNA of the same size on urea-polyacrylamide gels (27), we compared the size of the cRNA prepared using a known cDNA template with that estimated based on the sequencing ladder. The correct sizes were likely to be −20% smaller than those obtained based on the sequencing ladder. According to these results, the sizes of the protected bands for exons 1a, 1b, and 1c were −111, 163, and 104 bp, respectively. The sizes of the protected bands for exons 1a and 1b were almost in accordance with the results of 5'-RACE analysis, although the size for exon 1c was slightly higher than that calculated from 5'-RACE analysis.

Primer Extension Analysis

Since the protected band for exon 1c (Fig. 3B, panel c) was faint and the transcription initiation site estimated by RNase...
5'-RACE analysis, primer extension analysis was also performed to determine the transcription initiation site of exon 1c. As shown in Fig. 3C, the extension product obtained with the total RNA from SK-MEL-31 and AS cells was 66 nucleotides in length, corresponding to a position 2524 relative to the ATG codon. The results from primer extension analysis were almost in accordance with the results from 5'-RACE analysis.

Sequence Analysis of the 5'-Flanking Region

The sequence of the 5'-flanking region of the β1,4GalNAc-T gene is shown in Fig. 4. The three transcription initiation sites that were determined by 5'-RACE analysis and RNase protection assay were marked by arrows. In the 5'-flanking region of exon 1a, there are consensus binding sites for the transcription factors EGR-1, HNF-5, and Sp-1. In the 5'-flanking region of exon 1b, there are three binding sites for Sp-1 and one for AP-2. In the 5'-flanking region of exon 1c, one AP-2 site and two S1 HS sites are present. There is a TATA box at nucleotide 21730, although we did not detect another exon near the TATA box in the transcripts of SK-MEL-31.

Promoter/Enhancer Activity of the 5'-Flanking Region of the β1,4GalNAc-T Gene

To roughly test whether the 5'-flanking regions of these initiation sites of the gene had promoter activity, we prepared constructs in which the 5'-flanking and intron 1 regions were inserted before a promoterless CAT gene in plasmid pSV00CAT (Fig. 5A). Construct pGT-CAT—2228/692 pro-

**Fig. 4. Nucleotide sequences of the 5'-flanking region of the β1,4GalNAc-T gene.** The sequence of nucleotides -2228 to +20 is shown. The negative numbers at the left are relative to ATG (position +1). The putative transcription factor-binding sites based on a sequence comparison with known motifs are indicated at underlined (sense) and overlined (antisense) sequences. The arrows indicate the transcription initiation sites determined by 5'-RACE, RNase protection, and/or primer extension analysis. Boldface sequences indicate exons, and lightface sequences represent introns. GMCSF, granulocyte/macrophage colony-stimulation factor; SIF, sis-inducing factor; FBG, fibrinogen; IL, interleukin; MRE, metal-responsive element; XRE, xenobiotic-responsive element.
MeWo cells, another melanoma line that does not express the scarcely detectable when these constructs were transfected into AS cells (Fig. 5). Results were obtained when these pGTCAT constructs were examined for promoter activity of these three transcripts, the detected by 5'-flanking regions ligated upstream of the promoterless CAT gene. The results were obtained in at least three separate experiments. B, results of the CAT assay in SK-MEL-31 (panel a), AS (panel b), and MeWo (panel c) cells transfected with each CAT construct. The numbers indicate the constructs and correspond to those in A.

**FIG. 5.** Functional assay of CAT activity in human cell lines transiently transfected with the β1,4GalNAc-T/CAT constructs. A, structure and CAT activity of β1,4GalNAc-T/CAT constructs. The numbers at the left indicate the β1,4GalNAc-T gene 5'-flanking sequences ligated upstream of the promoterless CAT gene. The bars at the right indicate relative CAT activity in SK-MEL-31 (solid bars), AS (hatched bars), and MeWo (shaded bars) cells transfected with each CAT construct. CAT activity was normalized for transfection efficiency by the β-galactosidase activity from cotransfected pSV-β-galactosidase. Absorbance (405 nm) was 0.4–0.6 as determined by the β-galactosidase kit from Promega. An arbitrary value of 100 was given to the transcriptional activity resulting from the pGTCAT-2228/692 construct in SK-MEL-31 cells. Similar results were obtained in at least three separate experiments. B, results of the CAT assay in SK-MEL-31 (panel a), AS (panel b), and MeWo (panel c) cells transfected with each CAT construct. The numbers indicate the constructs and correspond to those in A.

Promoter/Enhancer Activity of Transcripts a, b, and c

As shown in Figs. 2 and 3, at least three initiation sites were detected by 5'-RACE analysis and RNase protection assay. To examine the promoter activity of these three transcripts, the CAT constructs pGTpCAT-2228/–948, pGTpCAT-947/–692, and pGTpCAT-662/–441 were prepared and examined by transfection into SK-MEL-31 cells. pGTpCAT-2228/–948 showed almost equivalent activity compared with pGTpCAT-2228/–692, and pGTpCAT-662/441 showed much higher activity. On the other hand, only background levels of CAT activity were detected in pGTpCAT-947/–692 (Fig. 6A).

To examine the region essential for the promoter/ enhancer activity of the β1,4GalNAc-T gene, several deletion constructs were prepared using pGTpCAT-2228/–156, which showed high levels of CAT activity as shown in Fig. 5. CAT activity was measured following transfection of the constructs into SK-MEL-31, AS, and MeWo cells. As shown in Fig. 6B, plasmids pGTpCAT-2228/–156 and pGTpCAT-947/–156 promoted CAT activity quite considerably in SK-MEL-31 and AS cells, although the CAT activity decreased with the increased shortening of the 5'-side. These constructs did not show any CAT activity in MeWo cells. The more deleted constructs showed very low levels of CAT activity even though they contained exon 1. However, pGTpCAT-688/–156 or more deleted constructs showed definite CAT activity in not only SK-MEL-31 and AS cells, but also in MeWo cells. pGTpCAT-662/441, in which the CAT gene was ligated to exon 1c, also promoted CAT activity in all three cell lines. The reverse construct, pGTpCAT-441/–662, scarcely promoted CAT activity, suggesting the presence of promoter/enhancer activity upstream of exon 1c (data not shown).

**DISCUSSION**

Chromosomal Mapping of the Human β1,4GalNAc-T Gene

FISH signals were constantly emitted by chromosome band 12q13.3 among the 50 cells analyzed (Fig. 7).

**DISCUSSION**

Since cDNAs of a number of glycosyltransferase genes have been isolated, the genomic organizations of several glycosyl-
transferases have been characterized so far. These results revealed that there are two types of genomic structure. The first group is represented by human \( b1,2-N\)-acetylglucosaminyltransferase I (28) and human \( a1,3\)-fucosyltransferase (29, 30) genes, in which the coding sequences are completely contained within a single large exon. The second group is represented by \( b1,4\)-galactosyltransferase (31), \( a2,6\)-sialyltransferase (9, 32), and \( a1,3\)-galactosyltransferase (33) genes, in which the coding sequences are scattered over several exons. The \( b1,6\)-N-acetylglucosaminyltransferase gene forming the core 2 O-glycan branch belongs to the former group (34), and the blood group A synthase gene is of the latter type (35). The \( b1,4\)GalNAc-T gene analyzed here is also of the latter type. Like many other glycosyltransferase genes, the presence of an untranslated 5'-exon and of one unusually long coding exon has also been recognized in this gene (5).

The \( b1,4\)GalNAc-T gene is expressed under various biological conditions. The gene is abundantly expressed in the brain tissues of vertebrates. We demonstrated that this gene is expressed at high levels in mouse brain at the late stage of development (17), when differentiation of neuronal cells proceeds and synapses are formed. Among malignant tumor cells, human cancer cells derived from the neural crest characteristically express the products of this gene. Almost all neuroblastoma cells ubiquitously express the \( b1,4\)GalNAc-T gene and \( G02\) (36), while malignant melanoma cells usually express this gene at the progressed and "vertical" phase (37). Furthermore, the expression of \( G02\) appears to be due to the transactivation of the \( b1,4\)GalNAc-T gene by human T cell lymphotropic virus type I p40\(^{tax}\) protein (16). These results suggest that this gene is regulated by several units of promoters and transcripts, provided there are no other glycosyltransferases catalyzing a similar reaction. This regulatory system of gene expression appears common among glycosyltransferase genes, as shown in \( b1,4\)-galactosyltransferase (13). The promoter units defined here should be differentially involved for the appropriate expression of this gene under the control of cell type-specific transcription factors. Among consensus binding sites for the transcription factors in the 5'-noncoding region of individual transcription initiation sites, EGR-1, known as NGFI-A, Krox-24, and zif268, is specially abundant in mouse brain, thymus, lung, and heart (38-41). HNF-5 is a liver-specific DNA-binding protein, and HNF-binding sites are present in multiple regulatory sequences of other genes with liver-specific expression.
signals were identified at chromosome region 12q13.3 on the R-banded metaphase chromosomes. Which consensus sequences among these elements are really significant remains to be elucidated.

Alternative splicing between exons 1 (1a or 1b) and 1c should also be involved in the differential expression of this gene. As has been shown, the α2,6-sialyltransferase gene is regulated by complicated alternative splicing mechanisms (47, 48). A few exons located in the 5′-non-coding region are alternatively used in a lineage- or differentiation-specific manner in B lymphoblastoid cell lines (48). The alternative usage of several exons coding the enzyme protein was also identified as a basis of tissue-specific molecular form and gene expression (47). The significance of the alternative usage of three exons in this gene also remains to be investigated.

The regulatory mechanisms for each transcription of this gene appear to be very complicated. Promoter/enhancer activity in intron 1 remains to be analyzed. Promoter/enhancer activity in the 5′-flanking region of exon 1c (nucleotides −688 to −634) was observed even in MeWo cells, if less than in the two other cell lines, although no expression of the gene was detectable in this cell line. These results suggest the possibility that nucleotides −810 to −688 have a regulatory activity, repressing downstream promoter activity. Whether the repressor is a major mechanism for the negative expression of the gene is a very interesting issue and is now under investigation in our laboratory.

Acknowledgments—We thank Dr. M. Sakai (Hokkaido University) and Dr. K. Yashikawa (Aichi Medical School) for valuable discussion. We also thank T. Shimomura and Y. Honda for excellent assistance.

REFERENCES
1. Wiegandt, H. (1985) in Glycolipids (Wiegandt, H., ed) pp. 199–260, Elsevier Science Publishing Co., Inc., New York
2. Hakomori, S. (1990) J. Biol. Chem. 265, 18713–18716
3. Hakomori, S. (1981) Annu. Rev. Biochem. 50, 733–764
4. Paulson, J. C., and Coley, K. J. (1989) J. Biol. Chem. 264, 17615–17618
5. Jorissen, D. H. (1992) Glycobiology 2, 271–277
6. Nakanishi, Y. (1994) Curr. Opin. Strut. Biol. 4, 683–691
7. Paulson, J. C., Weinstein, J., and Schauer, A. (1989) J. Biol. Chem. 264, 10931–10934
8. O’Hanlon, T. P., Lau, K. M., Wang, X., and Lau, J. Y. T. (1989) J. Biol. Chem. 264, 17389–17394
9. Svensson, E. C., Soreghan, B., and Paulson J. C. (1990) J. Biol. Chem. 265, 20863–20868
10. Svensson, E. C., Conley, P. B., and Paulson, J. C. (1992) J. Biol. Chem. 267, 3466–3472
11. Shaper, N. L., Wright, W. W., and Shaper, J. H. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 791–795
12. Russo, R. N., Shaper, N. L., and Shaper, J. H. (1990) J. Biol. Chem. 265, 10324–10331
13. Harduin-Lepers, A., Shaper, J. H., and Shaper, N. L. (1993) J. Biol. Chem. 268, 14348–14359
14. Ruan, S., and Lloyd, K. O. (1992) Cancer Res. 52, 5725–5731
15. Yamashiro, S., Ruan, S., Furukawa, K., Tai, T., Lloyd, K. O., Shiku, H., and Furukawa, K. (1993) Cancer Res. 53, 5395–5400
16. Furukawa, K., Akag, T., Nagata, Y., Yamada, Y., Shimotohno, K., Cheung, N.-K., Shiku, H., and Furukawa, K. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 1972–1976
17. Yamamoto, A., Haraguchi, M., Yamashiro, S., Fukumoto, S., Furukawa, K., Takamiya, K., Atsuta, M., Shiku, H., and Furukawa, K. (1996) J. Neurochem. 66, 26–34
18. Yu, R. K., Macala, L. J., Taki, T., Weinfeld, H. M., and Yu, F. S. (1988) J. Neurochem. 50, 1825–1832
19. Nagata, Y., Yamashiro, S., Yodoi, J., Lloyd, K. O., Shiku, H., and Furukawa, K. (1992) J. Biol. Chem. 267, 12082–12089
20. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 294–297
21. Davis, L. G., Dibner, M. D., and Battey, J. F. (1986) Basic Methods in Molecular Biology, Elsevier Science Publishing Co., Inc., New York
22. Takahashi, E., Yamauchi, T., Tsuji, H., Hitomi, A., Meuth, M., and Hori, T. (1991) Hum. Genet. 88, 119–126
23. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383
24. Kozak, M. (1984) Nucleic Acids Res. 12, 3873–3893
25. Kozak, M. (1986) Cell 44, 281–292
26. Kozak, M. (1989) J. Cell Biol. 108, 229–241
27. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
28. Hull, E., Sarkar, M., Spruitj, M. P. N., Hopkins, J. W. M., Dunn, R., and Schachter, H. (1991) Biochem. Biophys. Res. Commun. 176, 608–615
Genomic Organization of the $G_{M1}/G_{D2}$ Synthase Gene

29. Lowe, J. B., Kukowska-Latallo, J. F., Nair, R. P., Larsen, R. D., Marks, R. M., Macher, B. A., Kelly, R. J., and Ernst, L. K. (1991) J. Biol. Chem. 266, 17467–17477

30. Weston, B. W., Nair, R. P., Larsen, R. D., and Lowe, J. B. (1992) J. Biol. Chem. 267, 4152–4160

31. Hollis, G. F., Douglas, J. G., Shaper, N. L., Shaper, J. H., Stafford-Hollis, J. M., Evans, R. J., and Kirsch, I. R. (1989) Biochem. Biophys. Res. Commun. 162, 1069–1075

32. Wang, X.-C., O’Hanlon, T. P., Young, R. F., and Lau, J. T. Y. (1990) Glycobiology 1, 25–31

33. Joziasse, D. H., Shaper, N. L., Kim, D., Van den Eijnden, D. H., and Shaper, J. H. (1992) J. Biol. Chem. 267, 5534–5541

34. Bierhuizen, M. F. A., Maemura, K., Kudo, S., and Fukuda, M. (1995) Glycobiology 5, 417–425

35. Yamamoto, F., McNeill, P. D., and Hakomori, S. (1995) Glycobiology 5, 51–58

36. Cheung, N.-K. V., Sarrinen, U. M., Neely, J. E., Landmeier, B., Donovan, D., and Cozza, P. F. (1985) Cancer Res. 45, 2642–2649

37. Thurin, J., Thurin, M., Herlyn, M., Elder, D. E., Staplewski, Z., Clark, W. H., J., and Koprowski, H. (1986) FEBS Lett. 208, 17–22

38. Sukhatme, V. P., Cao, X., Chang, L. C., Tsai-Morris, C.-H., Stamenkovich, D., Ferreira, P. C. P., Cohen, D. R., Edwards, S. A., Shows, T. B., Curran, T., Le Beau, M. M., and Adamson, E. D. (1988) Cell 53, 37–43

39. Milbrandt, J. (1987) Science 238, 797–799

40. Lemaire, P., Revelant, O., Bravo, R., and Charnay, P. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 4691–4695

41. Christy, B. A., Lau, L. F., and Nathans, D. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 7857–7861

42. Grange, T., Roux, J., Rigaud, G., and Pictet, R. (1991) Nucleic Acids Res. 19, 131–139

43. Rigaud, G., Roux, J., Pictet, R., and Grange, T. (1991) Cell 67, 977–986

44. Mitchell, P. J., Timmons, P. M., Hebert, J. M., Rigby, P. W., and Tjian, R. (1991) Genes Dev. 5, 105–119

45. Williams, T., and Tjian, R. (1995) Genes Dev. 5, 4691–4695

46. Johnson, A. C., Jinno, Y., and Merlino, G. T. (1988) Molec. Cell. Biol. 8, 4174–4184

47. Wen, D. X., Svensson, E. C., and Paulson, J. C. (1992) J. Biol. Chem. 267, 2512–2518

48. Wang, X., Vertino, A., Edy, R. L., Byers, M. G., Jani-Sait, S. N., Shows, T. B., and Lau, J. T. Y. (1993) J. Biol. Chem. 268, 4355–4361

49. Svennerholm, L. (1963) J. Neurochem. 10, 613–623