

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/25518>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

A Family of Human β 4-Galactosyltransferases

CLONING AND EXPRESSION OF TWO NOVEL UDP-GALACTOSE: β -N-ACETYLGLUCOSAMINE β 1,4-GALACTOSYLTRANSFERASES, β 4Gal-T2 AND β 4Gal-T3*

(Received for publication, June 27, 1997, and in revised form, September 3, 1997)

Raquel Almeida^{‡§}, Margarida Amado^{‡§}, Leonor David[§], Steven B. Levery[¶], Eric H. Holmes^{||}, Gerard Merckx^{**}, Ad Geurts van Kessel^{**}, Eske Rygaard[‡], Helle Hassan[‡], Eric Bennett[‡], and Henrik Clausen[‡] ^{‡‡}

From the [‡]School of Dentistry, University of Copenhagen, Nørre Allé 20, DK-2200 Copenhagen N, Denmark, the [§]Institute of Molecular Pathology and Immunology of University of Porto, IPATIMUP, Rua Dr. R. Frias s/n, 4200 Porto, Portugal, the [¶]Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, the ^{||}Department of Cell Surface Biochemistry, Northwest Hospital, Seattle, Washington 98125, and the ^{**}Department of Human Genetics, University Hospital, Nijmegen 6050 HB, The Netherlands

BLAST analysis of expressed sequence tags (ESTs) using the coding sequence of the human UDP-galactose: β -N-acetylglucosamine β 1,4-galactosyltransferase, designated β 4Gal-T1, revealed a large number of ESTs with identical as well as similar sequences. ESTs with sequences similar to that of β 4Gal-T1 could be grouped into at least two non-identical sequence sets. Analysis of the predicted amino acid sequence of the novel ESTs with β 4Gal-T1 revealed conservation of short sequence motifs as well as cysteine residues previously shown to be important for the function of β 4Gal-T1. The likelihood that the identified ESTs represented novel galactosyltransferase genes was tested by cloning and sequencing of the full coding region of two distinct genes, followed by expression. Expression of soluble secreted constructs in the baculovirus system showed that these genes represented genuine UDP-galactose: β -N-acetylglucosamine β 1,4-galactosyltransferases, thus designated β 4Gal-T2 and β 4Gal-T3. Genomic cloning of the genes revealed that they have identical genomic organizations compared with β 4Gal-T1. The two novel genes were located on 1p32-33 and 1q23. The results demonstrate the existence of a family of homologous galactosyltransferases with related functions. The existence of multiple β 4-galactosyltransferases with the same or overlapping functions may be relevant for interpretation of biological functions previously assigned to β 4Gal-T1.

During the last decade, more than 40 mammalian glycosyltransferases have been cloned and characterized (1, 2). The initial strategy for cloning glycosyltransferases was cumbersome purification of labile enzyme proteins followed by screening of cDNA libraries with antibodies or DNA probes based on amino acid sequence information (3–11). The introduction of

transfection cloning by Lowe and co-workers (12) resulted in a marked increase in the cloning of novel glycosyltransferase genes (13–17). A third successful approach has taken advantage of conserved sequences in glycosyltransferases that share donor and/or acceptor substrates. Thus, searches for novel members of homologous glycosyltransferase gene families utilizing conserved sequence motifs for RT-PCR¹ cloning with degenerate primers have resulted in the identification and cloning of a number of novel genes (18, 19).

One part of the human genome project is the establishment of a data base of expressed sequence tags (ESTs), which currently has over 700,000 unique sequences. ESTs represent short 5'- and 3'-sequences (200–500 bp) of cDNA clones from a large variety of human and animal organs (20). The EST data base is now estimated to contain sequence information from more than half the human genes; it therefore provides a unique source for identifying novel members of homologous gene families (21). The EST data base has recently been successfully utilized in searches for novel glycosyltransferase genes of the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase family, where several novel members of this homologous gene family have been isolated by identification of ESTs with sequence similarities (22, 23).

A number of homologous families of glycosyltransferase genes exist. The largest family identified so far is the sialyltransferase family, for which at least 11 distinct members have been identified (24). Sequence motifs shared between sialyltransferases are part of the catalytic domain, suggesting that conservation of these sequences is related to functional requirements (25). Additional homologous glycosyltransferase gene families include the α 2- and α 3/4-fucosyltransferase families (26), the α 3-GalNAc-transferase family (27), the β 6-GlcNAc-transferase family (28), the ceramide galactosyltransferase family (29), and the polypeptide GalNAc-transferase family (23). Analysis of the sequence similarities within these glycosyltransferase gene families reveals that conserved sequences are generally limited to short sequence stretches in the putative catalytic domains, which are located in the central or C-terminal portions of enzymes that are type II transmembrane proteins. An additional characteristic is that cysteine residues in these areas are conserved in spacing (23, 30). Cysteine residues are important for intramolecular disulfide bonding as well as for the catalytic activity of glycosyltransferases (31–34).

* This work was supported by the Danish Cancer Society, the Ingeborg Roikjer Foundation, the Velux Foundation, the Danish Medical Research Council, the Lundbeck Foundation, Praxis XXI 2/2.1/BIA/276/94, and Grant 5 P41 RR05351 from the National Institutes of Health Resource Center for Biomedical Complex Carbohydrates. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EBI Data Bank with accession number(s) Y12509 and Y12510.

^{‡‡} To whom the correspondence should be addressed. Tel.: 45-35326835; Fax: 45-35326505; E-mail: henrik.clausen@odont.ku.dk.

¹ The abbreviations used are: RT-PCR, reverse transcription-polymerase chain reaction; EST, expressed sequence tags; bp, base pair(s); Cer, ceramide.

The UDP-galactose: β -N-acetylglucosamine β 1,4-galactosyltransferase (designated β 4Gal-T1; GenBank™/EMBL Data Bank accession number X14085) was the first glycosyltransferase to be isolated and cloned (3–5, 7, 35), and early searches for homologous genes by low stringency Southern hybridization suggested that this gene was unique. Characterization of β 4Gal-transferase activities from different sources, however, indicated that distinct activities exist (36, 37). Emerging evidence now reveals that several β 4-galactosyltransferase genes may exist. Shaper *et al.* (38) have identified two different chick cDNA sequences, which have 65 and 48% sequence similarity to human β 4Gal-T1. Both chick cDNAs were shown to encode catalytically active β 4Gal-transferases. Thus, the β 4Gal-T1 gene is likely to be part of a homologous gene family with recognizable sequence motifs, and this is supported by a large number of human ESTs with sequence similarities to β 4Gal-T1 in EST data bases.² Two independent groups have analyzed β 4Gal-transferase activities in mice homozygously deficient in β 4Gal-T1 (39, 40). Both studies showed residual β 4Gal-transferase activity, providing clear evidence for the existence of additional β 4Gal-transferases. Uehara and Muramatsu (41) have identified and cloned a mouse cDNA that was shown to encode a protein with low β 4-galactosyltransferase activity when expressed in *Escherichia coli*. This putative β 4-galactosyltransferase exhibited little or no sequence similarity to β 4Gal-T1, indicating that the gene is unlikely to represent an evolutionarily related member of a galactosyltransferase family. Since the gene was expressed in *E. coli*, it makes evaluation of the kinetic parameters uncertain as this expression system has been shown to be poor for glycosyltransferases (42).

In this study, we used available human EST sequence information to identify novel genes homologous to β 4Gal-T1. The β 4Gal-T1 gene was found to be highly represented in the EST data base by ESTs with nearly identical sequences, but in addition, a number of ESTs with similar sequences were identified, and these had shared motifs with conserved cysteine residues previously shown to be functionally important for β 4Gal-T1. The full coding sequences of two of these genes were established, and expression demonstrated that these genes represented active UDP-Gal: β -GlcNAc β 1,4-Gal-transferase genes.

EXPERIMENTAL PROCEDURES

Identification of Genes Homologous to β 4Gal-T1—Data base searches were performed with the coding sequence of the human β 4Gal-T1 sequence (43) using the BLASTn and tBLASTn algorithms against the dbEST data base at the NCBI. The BLASTn algorithm was used to identify ESTs representing the query gene (identities of $\geq 95\%$), whereas tBLASTn was used to identify non-identical, but similar EST sequences. ESTs with 50–90% nucleotide sequence identity were regarded as different from the query sequence. The results of tBLASTn searches were evaluated by visual inspection after elimination of ESTs regarded as identical to the query sequence ($>95\%$ nucleotide sequence identity). ESTs with several apparent short sequence motifs and cysteine residues arranged with similar spacing were selected for further sequence analysis. Initially, the identified ESTs (5'-sequence) were used in BLASTn searches of the dbEST data base to search for overlapping ESTs (95–100% identity in at least 30 bp) (see Fig. 1). If new ESTs were identified, the procedure was repeated, and sequences were merged. In addition, all identified ESTs were analyzed in the Unigene data base to confirm that they were from the same gene transcript and to select cDNA clones with the longest inserts as well as to identify additional ESTs with a non-overlapping 5'-sequence. Composites of all the sequence information for each set of ESTs were compiled and analyzed for sequence similarity to human β 4Gal-T1. EST cDNA clones with the longest inserts (see Fig. 1) were obtained from Genome Systems Inc.

Cloning and Sequencing of the Full Coding Sequence of β 4Gal-T2—

Two partly overlapping ESTs were identified (see Fig. 1). Sequencing of the inserts revealed an open reading frame that potentially encoded a sequence similar to that of β 4Gal-T1, but the 5'-sequence was shorter and without an initiation codon. Further 5'-sequence was obtained by 5'-rapid amplification of cDNA ends using human fetal brain Marathon-Ready cDNA (CLONTECH) in combination with the antisense primers EBER102 (5'-GAAACTGAGCCTTACTCAGGC) and EBER104 (5'-TCCACATCGCTGAAGATGAAGC) for 35 cycles at 95 °C for 45 s, 55 °C for 15 s, and 68 °C for 3 min using the Expand kit enzyme (Boehringer Mannheim). The products from the 5'-rapid amplification of cDNA ends were cloned into the *Bam*HI/*Not*I site of pT7T3U19, and multiple clones were sequenced. The entire sequence was confirmed by sequencing genomic P1 clones. The composite sequence contained an open reading frame of 1191 bp potentially encoding a protein with a type II domain structure (see Fig. 2), with an overall sequence identity of $\sim 63\%$ to β 4Gal-T1.

Cloning and Sequencing of the Full Coding Sequence of β 4Gal-T3—One EST clone (184081) with a 1980-bp insert was identified by its 3'-EST sequence in Unigene (National Center for Biotechnology Information) (see Fig. 1). Sequencing of the insert revealed an intact open reading frame of 1179 bp potentially encoding a protein with a type II domain structure (see Fig. 3), with an overall sequence identity of $\sim 54\%$ to β 4Gal-T1.

Expression of β 4Gal-T2 and β 4Gal-T3 in Sf9 Cells—An expression construct designed to encode amino acid residues 56–397 of β 4Gal-T2 was prepared by RT-PCR with mRNA from the Colo205 cell line using the primer pair EBER100FOR (5'-TACTTTGACGTCTACGCCAG) and EBER114 (5'-GAAAACAGAGCCAGTCCAG) with *Bam*HI restriction sites (see Fig. 2). An expression construct designed to encode amino acid residues 23–393 of β 4Gal-T3 was prepared by RT-PCR with RNA from the MKN45 gastric carcinoma cell line using the primer pair EBER200FOR (5'-CATGATGTACCTGTCACTGGGG) and EBER214 (5'-TAGCACGGCACCAGAGTTCAG) with *Bam*HI restriction sites (see Fig. 3). The PCR products were cloned into the *Bam*HI site of pAcGP67 (Pharmingen), and the construct was sequenced to verify correct insertion and sequence. A full-length coding expression construct of β 4Gal-T3 was prepared by RT-PCR with MKN45 RNA using the primer pair EBER200FUL (5'-CCAGGATGTTGCGGAGGCTGC) and EBER214 (see Fig. 3), and the product was cloned into the *Bam*HI site of pVL1193 (Pharmingen). Plasmids pAcGP67- β 4Gal-T2-sol, pAcGP67- β 4Gal-T3-sol, and pVL- β 4Gal-T3-full were cotransfected with BaculoGold™ DNA (Pharmingen) as described previously (19). Recombinant baculovirus was obtained after two successive amplifications in Sf9 cells grown in serum-containing medium, and titers of virus were estimated by titration in 24-well plates with monitoring of enzyme activities. Controls included pAcGP67-GalNAc-T3-sol (19). Standard assays were performed in 50 μ l of total reaction mixture containing 25 mM Tris (pH 7.5), 10 mM MnCl₂, 0.25% Triton X-100, 100 μ M UDP-[¹⁴C]Gal (2300 cpm/nmol; Amersham Corp.), and varying concentrations of acceptor substrates (Sigma) (see Table I for structures). The soluble constructs were assayed with 5–20 μ l of culture supernatant from infected cells, whereas the full-length construct was assayed with 1% Triton X-100 homogenates of washed cells. Bovine milk β 1,4-Gal-transferase (Sigma) was used as control. Assays used for determination of K_m of acceptor substrates were modified to include 200 μ M UDP-[¹⁴C]Gal, and assays for donor substrate K_m were performed with 2 mM (for β 4Gal-T3 and bovine milk Gal-transferase) or 0.25 mM (for β 4Gal-T2) benzyl- β -GlcNAc.

Characterization of the Product Formed with β 4Gal-T3—Soluble β 4Gal-T3 was partially purified by sequential DEAE and S-Sepharose chromatographies from serum-free medium as described previously (44). Terminal glycosylation of Lc₃Cer (see Table III for definitions of Svennerholm nomenclature) was performed in a reaction mixture consisting of 3 milliunits of β 4Gal-T3 (specific activity determined with benzyl- β -GlcNAc), 250 μ g of Lc₃, 25 mM Tris (pH 7.4), 10 mM MnCl₂, 0.25% Triton CF-54, and 3.5 μ mol of UDP-Gal in a final volume of 100 μ l. The Lc₃Cer substrate was prepared from sialosyllactoneotetraosylceramide of bovine erythrocytes (45) by desialylation with 5% HAc in 2-propanol/hexane/H₂O (55:25:20, v/v/v; lower phase) at 80 °C for 4.5 h, isolation of the nLc₄Cer product by passage through DEAE-Sephadex A-25 (Na⁺ form) in CHCl₃/MeOH/H₂O (30:60:8, v/v/v), and treatment of the product with jack bean β -galactosidase as described previously (46, 47). The Lc₃Cer product was evaluated by high performance TLC and ¹H NMR (47). The glycosylation of Lc₃Cer with β 4Gal-T3 was monitored by high performance TLC and run for 24 h until completed. The crude reaction mixture was taken up in ~ 3 ml of 0.1 M NaCl, sonicated thoroughly, applied to a 1-ml disposable octadecylsilica cartridge (Bakerbond, J. T. Baker Inc.), washed through with 4 ml each of 0.1 M

² R. Almeida and H. Clausen, unpublished observation.

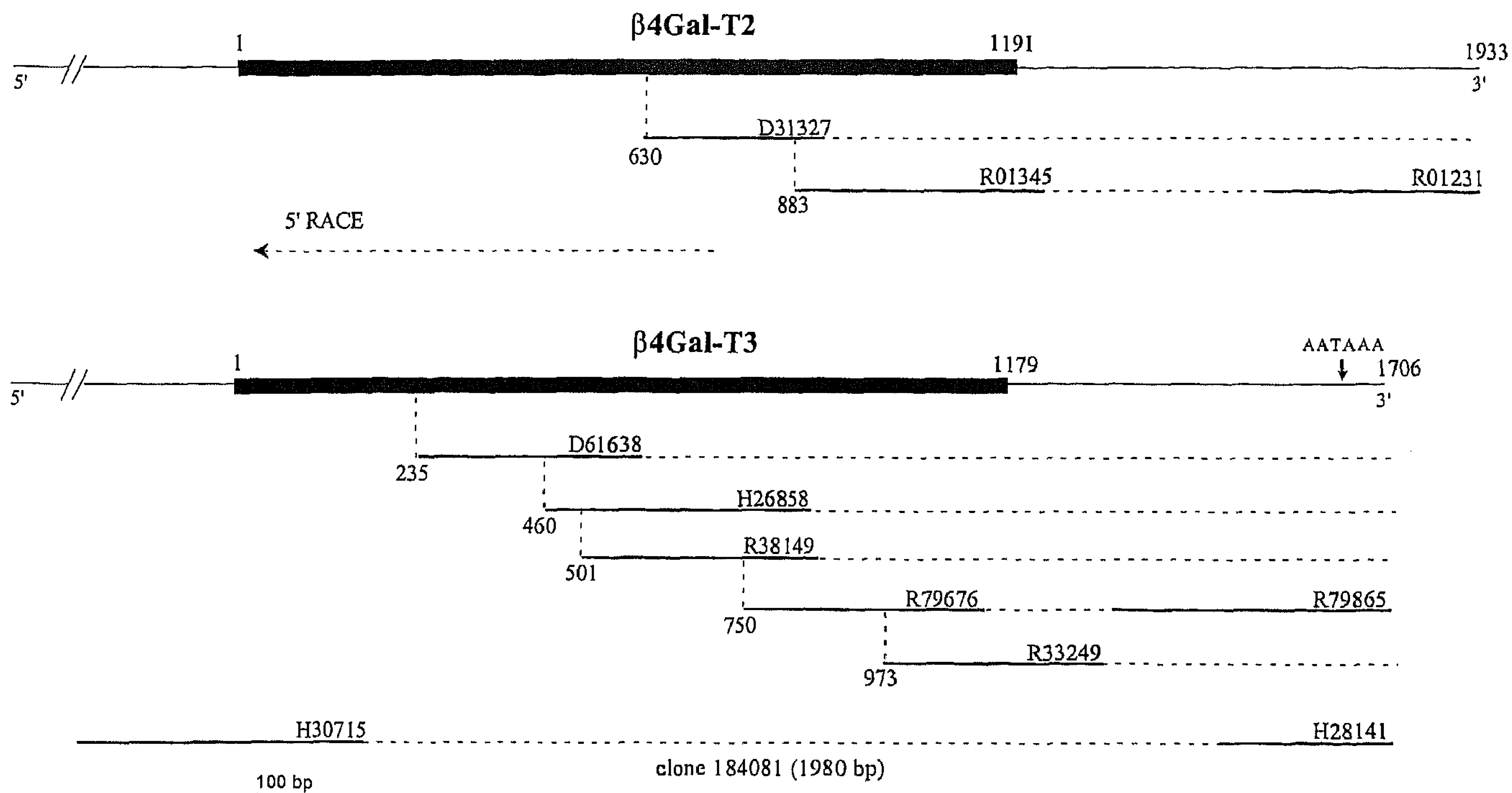


FIG. 1. Strategy for identification and cloning of β 4Gal-T2 and β 4Gal-T3. Identified ESTs are indicated by their GenBank™/EMBL Data Bank accession numbers with available sequence lengths in parentheses. Vertical dashed lines labeled with numbers indicate 5'-positions of EST clones compared with the coding sequence of the gene. RACE, rapid amplification of cDNA ends.

NaCl and H₂O, and then eluted with 4-ml portions of H₂O containing increasing amounts of MeOH from 10 to 90%, followed by 100% MeOH. Following concentration of each fraction, an assessment by high performance TLC analysis showed that the glycosphingolipid product was eluted as a single band with 100% MeOH, whereas the detergent eluted at 80% MeOH. The purified product was deuterium-exchanged by dissolving in CDCl₃/CD₃OD (2:1), evaporating thoroughly under dry nitrogen (repeating twice), and then dissolving in 0.5 ml of Me₂SO-*d*₆ and 2% D₂O (48). One-dimensional ¹H NMR spectroscopy was performed on a Bruker AMX-500 spectrometer (temperature, 308 K; spectral width, 5000 Hz acquired over 16,000 data points; relaxation delay, 2 s; and solvent suppression by presaturation pulse). NMR spectra of both the substrate and product were interpreted by reference to spectra of relevant glycosphingolipid standards acquired previously under identical conditions (46, 47, 49).

Northern Analysis—Human multiple tissue Northern blots were obtained from CLONTECH. The soluble expression construct of β 4Gal-T2 and the full coding construct of β 4Gal-T3 were used as probes. Probes were random prime-labeled using [α -³²P]dCTP (Amersham Corp.) and an oligonucleotide labeling kit (Pharmacia Biotech Inc.). The blots were probed overnight at 42 °C as described previously (19) and washed twice for 10 min each at room temperature with 2 × SSC and 1% Na₄P₂O₇; twice for 20 min each at 65 °C with 0.2 × SSC, 1% SDS, and 1% Na₄P₂O₇; and once for 10 min with 0.2 × SSC at room temperature.

Genomic Cloning and Characterization of the Organization of β 4Gal-T2 and β 4Gal-T3—A human foreskin genomic P1 library (DuPont Merck Pharmaceutical Co. Human Foreskin Fibroblast P1 Library) was screened using the primer pairs EBER100 (5'-TGAAGGAGGATGC-CGCCTATGAC) and EBER102 for β 4Gal-T2 and EBER200 (5'-CCCT-GCGTGATGAAGAGTGGG) and EBER202 (5'-ATCTTCAGGTAAGT-GTCAGG) for β 4Gal-T3. Three clones for β 4Gal-T2 (DPMC-HFF10638:515:G9, DPMC-HFF10639:516:G4, and DPMC-HFF10640:924:A11) and three clones for β 4Gal-T3 (DPMC-HFF1-0288-D2, DPMC-HFF1-0838-A8, and DPMC-HFF1-1486-H6) were obtained from Genome Systems Inc. DNA from P1 phages was prepared as recommended by Genome Systems Inc. The entire coding sequence of each gene was sequenced in full using automated sequencing (ABI377, Perkin-Elmer) with dye terminator chemistry. Intron/exon boundaries were determined by comparison with the cDNA sequences optimizing for the *gt/ag* rule (50).

In Situ Hybridization to Metaphase Chromosomes—Fluorescence *in situ* hybridization was performed on normal human lymphocyte metaphase chromosomes using essentially the procedures as described previously (51). Briefly, P1 DNA was labeled with biotin-14-dATP using

the bioNICK labeling system (Life Technologies, Inc.). The labeled DNA was precipitated with ethanol in the presence of herring sperm DNA. A total of 300 ng of P1 DNA was precipitated with 50 × human Cot1 DNA (Life Technologies, Inc.) and dissolved in 12 ml of hybridization solution (2 × SSC, 10% dextran sulfate, 1% Tween 20, and 50% formamide (pH 7.0)). Prior to hybridization, the probe was heat-denatured at 80 °C for 10 min, chilled on ice, and incubated at 37 °C to allow re-annealing of highly repetitive sequences. After denaturation of the slides, probe incubations were carried out under a 18 × 18-mm coverslip in a moist chamber for 45 h. Immunocytochemical detection of the probe was achieved using fluorescein isothiocyanate-labeled avidin (Vector Laboratories, Inc.) and several successive steps with rabbit anti-fluorescein isothiocyanate- and mouse anti-rabbit fluorescein isothiocyanate-conjugated antibodies. For evaluation of the chromosomal slides, a Zeiss epifluorescence microscope equipped with appropriate filters for visualization of fluorescein isothiocyanate was used. Hybridization signals and 4,6-diamidino-2-phenylindole-counterstained chromosomes were transformed into pseudo-colored images using image analysis software. For precise localization and chromosome identification, 4,6-diamidino-2-phenylindole-converted banding patterns were generated using the BDS-image™ software package (ONCOR).

RESULTS

Identification and Cloning of Human β 4Gal-T2 and β 4Gal-T3—The strategy outlined in Fig. 1 produced two novel genes with significant sequence similarity to β 4Gal-T1 (Figs. 2 and 3). Additionally, two genes with significant similarities in the 3'-region were identified (data not shown). Multiple sequence alignment analysis of β 4Gal-T1, the two novel human β 4Gal-T enzymes, two homologous sequences from chick (38), and a snail β 4GlcNAc-transferase (52) is shown in Fig. 4. The β 4Gal-T1 gene shows high sequence similarity (65%) throughout the coding region to one of the chick genes (GenBank™/EMBL Data Bank accession number U19890), and β 4Gal-T2 shows high similarity (72%) to the second chick gene (GenBank™/EMBL Data Bank accession number U19889), suggesting that these may represent species-associated members of related gene families. The amino acid sequence similarities among the three human β 4-Gal-transferases and the snail β 4GlcNAc-transferase are limited to the central regions; there were no significant similarities in the NH₂-terminal regions. Human

β 4Gal-T2

-93AGCCTGGTCCCAGTTGGCCTGCCCTGCTGTGCTGGGATCTGAATGACCAAACCACTTCCCACCATGGCTCCTGGAAGGACTAAATGAAGTC	-1
1	M S I K C S C M A S S R M P G P T G R A S G R L R D E Q T A G G T L E R V C K A	40
1	ATGAGTATAAAGTGTCTCCTGCATGGCCAGCAGCCGGATGCCCGGGCCACTGGGCGGGCCAGTGGCCGCTTGGGGATGAGCAGACTGCTGGGGGACGCTGGAGCGGCTCTGCAAGGCT	120
41	<u>V L L L C L L H F L V A V I L Y F D V Y A Q H L A F F S R F S A R G P A H A L H</u>	80
121	GTGCRCTTCTCTGCCTGCTGCACTTCCCTCGTGGCCGTCATCCTCTACTTTGACGCTCTACGCCAGCACCTGGCCTTCTTCAGCCGCTTCAGTCCCCAGGCCCTGCCATGCCCTCCAC	240
81	P A A S S S S S S S S N' C S R P N' A T A S S S S G L P E V P S A L P G P T A P T L P	120
241	CCAGCTGCTAGCAGCAGCAGCAGCAGCAACTGCTCCCGGCCAACGCCACCGCCTTAGCTCCGGGCTCCCTGAGGTCCCAGTGCCTGCCCGGTCCCAGGCTCCCAGCTGCCA	360
121	P C P D S P P G L V G R L L I E F T S P M P L E R V Q R E N P G V L M G G R Y T	160
361	CCCTGTCTGACTCGCCACCTGGTCTTGTGGGCAGACTGTGATCGAGTTCACCTCACCCATGCCCTGGAGCGGGTGCAGAGGGAGAACCAGGCGTGCATGGGGCCGATACACA	480
161	P P D C T P A Q T V A V I I P F R H R E H H L R Y W L H Y L H P I L R R Q R L R	200
481	CCGCCGACTGCACCCAGCCAGACGGTGGCGGTTCATCATCCCCTTTAGACACCGGGAACACCACCTGGCCTACTGGCTCCACTATCTACACCCCATCTTGGGGCCGAGCGGCTGCGC	600
201	Y G V Y V I N Q H G E D T F N R A K L L N V G F L E A L K E D A A Y D C F I F S	240
601	TACGGCGTCTATGTCAACCCAGCATGGTGGAGACACCTTCAACCGGGCCAGCTGCTTAAAGTGGGCTTCTAGAGGCGTGAAGGAGGATGCCGCTATGACTGCTTCACTTCCAGC	720
241	D V D L V P M D D R N L Y R C G D Q P R H F A I A M D K F G F R L P Y A G Y F G	280
721	GATGTGGACTGGTCCCCTAGGATGACCGCAACCTATACCGTGGCGGACCAACCCCGCCACTTTGCCATGGCATGGACAAGTTTGGCTTCCGGCTTCCCTATGCTGGCTACTTTGGA	840
281	G V S G L S K A Q F L R I N G F P N E Y W G W G G E D D D I F N R I S L T G M K	320
841	GGTGTCTCAGCCCTGAGTAAGGCTCAGTTTCTGAGAATCAATGGCTTCCCAATGAGTACTGGGGCTGGGGTGGCGAGGATGATGACATCTTCAACCGGATCTCCCTGACTGGGATGAAG	960
321	I S R P D I R I G R Y R M I K H D R D K H N E P N P Q R F T K I Q N T K L T M K	360
961	ATCTCACGCCAGACATCCGAATCGGCCGCTACCGCATGATCAAGCAGCAGCCGACAAGCATAACGAACCTAACCTCAGAGGTTTACCAAGATTCAAACACGAAGCTGACCATGAAG	1080
361	R D G I G S V R Y Q V L E V S R Q P L F T N' I T V D I G R P P S W P P R G *	397
1081	CGGGACGGCATTGGGTGAGTGGTACCAGGTCTTGGAGGTGTCTCGGCAACCACTCTTACCAATATCACAGTGGACATGGGCGGCCTCCGTGGTGGCCCTCGGGGCTGACACTAA	1200
1201	TGGACAGAGGCTCTCGGTGCCGAAGATTGCCCTGCCAGAGGACTGACCACAGCCTGGCTGGCAGCTGCTCTGTGGAGGACCTCCAGGACTGAGACTGGGCTCTGTTTTCCAAGGCTCTTCA	1320
1321	CTAGGCCCTTAGCTACACCTGGAAGTTTTCAGAACCCTTGGGGGGCTTCTGCTGGCAGGCTCTTCAAGTGTGGCCCTTCTTGGAGTCAACCCCTCTCCCGACCCCTCCCT	1440
1441	AGCCAGCCCTAGCTACTGTGAGGTCGGGCGCCCTGCACTGCTCGCAGAGTGGCTGGGCTAGGCTCACTCCACTCTCTGTGCCTCAGTTCCCCCTTGGAGTCCCTAGGGCC	1560
1561	TGGAAGGGTGGGAGGTATGTCTAGGGGCGAGTGTCTTCCAGGGGAATTCAGCTCTTGGGAACCCCTTGTCCCAGGGGAGGGGAAACCTTTTTTCAATCAACATTGTAGGGGCA	1680
1681	AGCTTTGGTGGCCCTGCTGAGGAGCAGCCAGGAGGGGACAGAGGGGATGCTGTGTGCTGCTGGGATCTTGGGGTGGCCTTTGCAATGGGAGGCAGGTGGGGCTTGGATCAGT	1800
1801	AAGTTTGGTCCCGCCTCCCTGTTTGGAGAGGAGGAGCCAGGAGCCCGGGCGGCTTGTGTTGTACATTGCACAGAACTTGTGTGGGTGCTTTAGTAAAAACGTAATGGAAAAAA	1920
1921	AAAAAAAAAAAAA	1933

FIG. 2. Nucleotide sequence and predicted amino acid sequence of human β 4Gal-T2. The amino acid sequence is shown in single-letter code. The hydrophobic segment representing the putative transmembrane domain is *double-underlined* (Kyte and Doolittle (85), window of 8). Three potential *N*-glycosylation sites are indicated by *asterisks*. The locations of the primers used for RT-PCR preparation of the expression construct are *single-underlined*.

 β 4Gal-T3

-94ATCTCTGTCCCCCGGACCCCTGGACCCCATGATCCCTAAGAAGACAGCTTGAACCTAGAATCCCCTCTCCCTCCCTAGATCTCACCCCCAGG	-1
1	M L R R L L E R P C T L A L L V G S O L A V M M Y L S L G G F R S L S A L F G R	40
1	ATGTTGCGGAGGCTGCTGGAGCGGCTTGCACGCTGGCCCTGCTTGTGGGCTCCAGCTGGCTGTTCATGATGTACCTGTCACTGGGGGCTTCCGAAGTCTCAGTGCCTATTGGCCGA	120
41	D Q G P T F D Y S H P R D V Y S N' L S H L P G A P G G P P A P Q G L P Y C P E R	80
121	GATCAGGGACCGACATTTGACTATTCTACCCCTCGTATGTCTACAGTAACCTCAGTCACTGCTGGGGCCCGAGGGGCTCTCCAGCTCCTCAAGTCTGCCCTACTGTCCAGAACGA	240
81	S P L L V G P V S V S F S P V P S L A E I V E R N P R V E P G G R Y R P A G C E	120
241	TCTCTCTCTTAGTGGGCTCTGTGCTCGGTGCTTGTAGCCAGTGCATCACTGGCAGAGATTGTGGAGCGGAATCCCCGGGTAGAACCAGGGGGCCGTACCGCCCTGCAGGTTGTGAG	360
121	P R S R T A I I V P H R A R E H H L R L L L Y H L H P F L Q R Q Q L A Y G I Y V	160
361	CCCCGCTCCCGAACAGCCATCATTGTGCTCATCGTCCCGGGAGCACCACTGCGCCTGCTGCTTACCACCTGCACCCCTTCTTGCAGCGCCAGCAGCTTGTATGGCATCTATGTC	480
161	I H Q A G N' G T F N R A K L L N V G V R E A L R D E E W D C L F L H D V D L L P	200
481	ATCCACCAGGCTGGAATGGAACATTTAACAGGGCAAACCTGTTGAACGTTGGGGTGCAGAGGGCCCTGCGTGTAGAGAGTGGGACTGCCTGTTCTTGCACGATGTGGACCTCTTGCCA	600
201	E N D H N L Y V C D P R G P R H V A V A M N K F G Y S L P Y P Q Y F G G V S A L	240
601	GAAATGACCACAATCTGTATGTGTGACCCCGGGGACCCCGCCATGTTGCCGTTGTCTATGAACAAGTTTGGATACAGCCTCCCGTACCCCGAGTACTTCCGAGGAGTCTCAGCACTT	720
241	T P D Q Y L K M N G F P N E Y W G W G G E D D D I A T R V R L A G M K I S R P P	280
721	ACTCTGACCAGTACCTGAAGATGAATGGCTTCCCAATGAATACTGGGGCTGGGGTGGTGGAGGATGACGACATGTCTACCAGGGTGCCTGGCTGGGATGAAGATCTCTCGGCCCCC	840
281	T S V G H Y K M V K H R G D K G N E E N P H R F D L L V R T Q N S W T Q D G M N	320
841	ACATCTGTAGGACTATAAGATGGTGAAGCACCGAGGAGATAAGGCAATGAGGAAAATCCCCACAGATTTGACCTCCTGGTCCGTACCAGAAATCTCGACGCAAGATGGGATGAAC	960
321	S L T Y Q L L A R E L G P L Y T N' I T A D I G T D P R G P R A P S G P R Y P P G	360
961	TCACTGACATACCAGTGTGCTGGCTCGAGAGCTGGGGCCCTTTATACCAACATCACAGCAGACATTTGGGACTGACCCCTCGGGGCTCCTCGGGCTCCTTCTGGGCCAGTTACCCACCTGGT	1080
361	S S Q A F R Q E M L Q R R P P A R P G P L S T A N' H T A L R G S H *	393
1081	TCCTCCCAAGCCTTCCGTCAAGAGATGTGCAACGCCGGCCCGCCAGCCAGGCTGGGCTCTATCTACTGCCAACACACAGCCCTCCGAGGTTCACTGACTCCTCCTTCTGTCTAC	1200
1201	CTTAATCATGAACCGAATTCATGGGGTGTATTCTCCCCACCTCAGCTCCTCACTGTTCTCAGAGGATGTGAGGGAACTGAACTCTGGTGGCGTCTAGGGGGTAGGGGCTCTCC	1320
1321	TCACTGCTGGACTGGAGCTGGGCTCCTGTAGACCTGAGGGGTCCCTCTCTTAGGCTCTCTGTAGGGCTTATGACTGTGAATCCTTGATGTATGATTTTATGTGACGATCTTAGGAG	1440
1441	TCCCTGCCCTAGAGTAGGAGCAGGGCTGGACCCCAAGCCCTCCCTCTCCATGGAGAGAAGAGTGTCTGGCTTCTCTCGGACCTCTGTGAATATTTATTCTATTATGGTTCCCGG	1560
1561	GAAGTTGTTTGGTGAAGGAAGCCCTCCCTGGGCATTTCTGCCTATGCTGGAATAGCTCCCTCTTCTGGTCTGGCTCAGGGGGCTGGGATTTTGATATATTTCTAATAAAGGACTT	1680
1681	TGTCTCGCAAAAAAAAAAAAAA	1706

FIG. 3. Nucleotide sequence and predicted amino acid sequence of human β 4Gal-T3. Designations are as described in the legend to Fig. 2. A potential polyadenylation signal is indicated in *boldface underlined type*.

β 4Gal-T1 is closest in sequence to human β 4Gal-T2 (52% identity), more distant from human β 4Gal-T3 (44% identity), and most distant from snail β 4-GlcNAc-transferase (34% identity).

Several sequence motifs in the putative catalytic domains are conserved among all the sequences. More important, four cysteine residues are conserved in all coding sequences (Fig. 4).

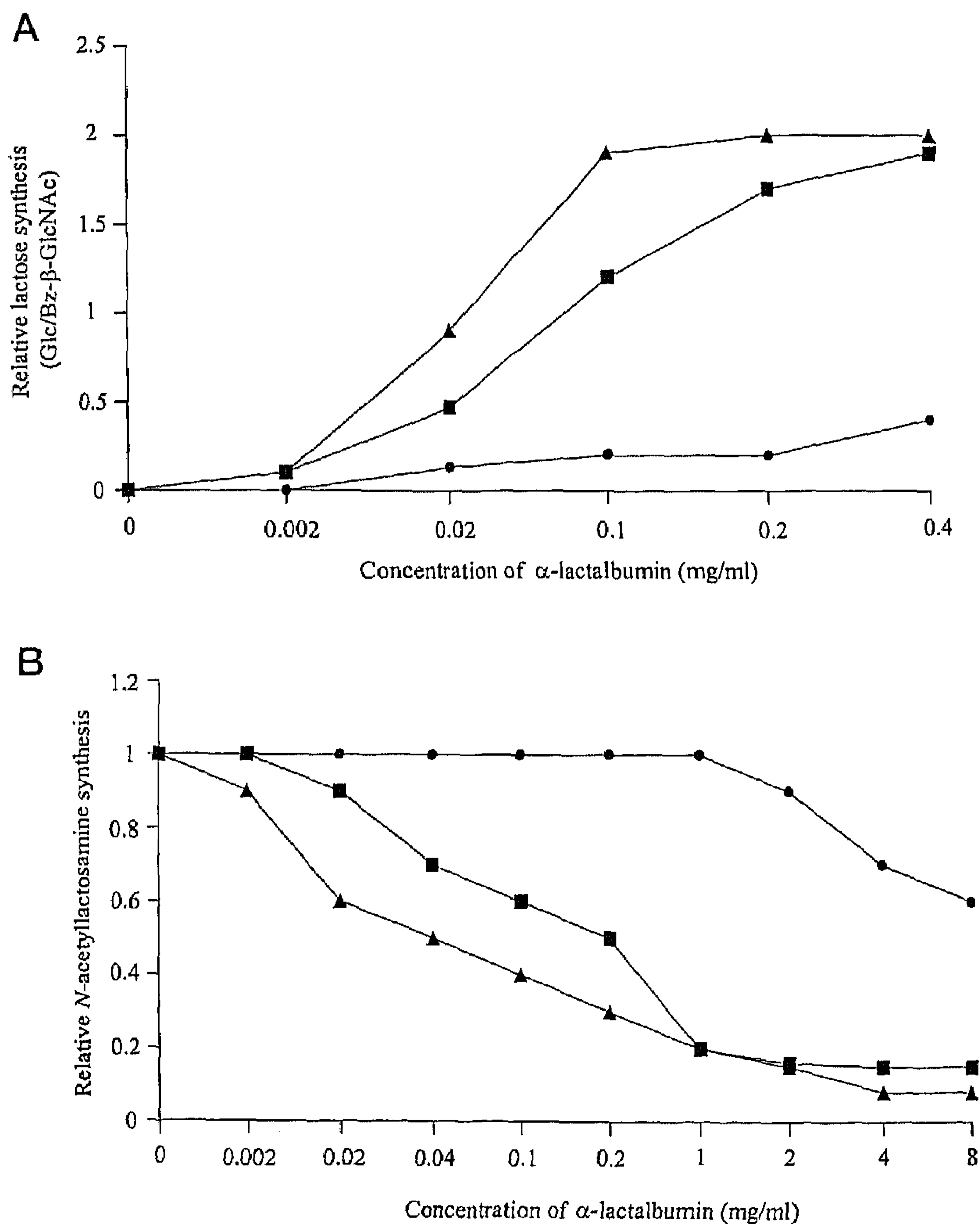
TABLE I
 Substrate specificity of β 4-Gal-transferases

Substrate concentration	β 4Gal-T2 ^a			β 4Gal-T3 ^a			Bovine β 4-Gal-transferase ^a		
	1 mM	3 mM	9 mM	1 mM	3 mM	9 mM	1 mM	3 mM	9 mM
	nmol/min/ml			nmol/min/ml			nmol/min/ml		
D-GlcNAc	1.4	3.2	4.8	2.3	4.5	9.4	63	150	286
Bzl- β -D-GlcNAc ^b	6.8	3.6	1.5	16.1	13.9	5.6	411	407	232
Bzl- α -D-GlcNAc	0.4	1.1	1.7	0.4	1.3	1.9	0.0	15	30
<i>o</i> -Nph- α -D-GlcNAc	0.4	0.8	1.5	0.3	1.2	1.0	0.0	0.0	48
<i>p</i> -Nph- β -D-GlcNAc	3.0	2.3	0.9	11.8	12.5	6.5	188	380	226
<i>p</i> -Nph-1-thio- β -D-GlcNAc	1.2	1.6	0.2	9.9	6.9	2.1	140	236	160
4-Me-lumb- β -D-GlcNAc	0.8	0.6	0.4	7.9	3.8	3.3	213	198	159
β -D-GlcNAc-(1-3)- β -D-Gal-1-OMe	5.8	7.7	ND	7.8	15.1	ND	246	350	ND
β -D-GlcNAc-(1-6)- α -D-Man-1-OMe	8.5	11.3	ND	12.7	20.4	ND	441	523	ND
Bzl-2-(2- β -D-GlcNAc)- α -D-GlcNAc	9.9	2.6	1.3	13.0	7.0	1.4	535	288	140
4-Me-lumb- β -D-GalNAc	ND	0.0	ND	ND	0.0	ND	ND	0.0	ND
<i>o</i> -Nph- β -D-GalNAc	ND	0.0	ND	ND	0.0	ND	ND	0.0	ND
Bzl- α -D-GalNAc	ND	0.0	ND	ND	0.0	ND	ND	0.0	ND
4-Me-lumb- β -D-Gal	ND	0.0	ND	ND	0.0	ND	ND	0.0	ND
<i>o</i> -Nph- β -D-Gal	ND	0.0	ND	ND	0.0	ND	ND	0.0	ND

^a Enzyme sources were media of infected Sf9 cells. Background values obtained with uninfected cells or cells infected with an irrelevant construct were subtracted. The background rates were not higher than 0.5 nmol/min/ml.

^b Bzl, benzyl; Nph, nitrophenyl; lumb, lumbelyl; ND, not determined.

FIG. 5. α -Lactalbumin modulation of β 4-galactosyltransferase activities. *A*, activities with glucose in the presence of increasing amounts of α -lactalbumin. The results are presented relative to the activities obtained with 40 mM glucose. *B*, activities with GlcNAc in the presence of increasing amounts of α -lactalbumin. The results are presented relative to the activities obtained with 2 mM (for the bovine milk enzyme and β 4Gal-T3) or 0.25 mM (for β 4Gal-T2) benzyl- β -GlcNAc (*Bz*- β -GlcNAc). Purified bovine milk enzyme or medium from Sf9 cells expressing secreted forms of either β 4Gal-T2 or β 4Gal-T3 was used as the enzyme source. \blacktriangle , bovine milk Gal-transferase; \blacksquare , β 4Gal-T2; \bullet , β 4Gal-T3.



for UDP-Gal using benzyl- β -GlcNAc was 11 μ M (Table II). Bovine milk β 4-galactosyltransferase showed higher a K_m for UDP-Gal, in agreement with previous studies (37, 55-58), and the measured K_m for GlcNAc was similar to that determined in some studies (59, 60), but 5-10-fold higher compared with other studies (55-58). As shown in Fig. 6, β 4Gal-T2 was inhibited at high concentrations of both benzyl- β -GlcNAc and free *N*-acetylglucosamine to a higher degree than bovine milk β 4-

Gal-transferase and β 4Gal-T3 (61). β 4Gal-T2 showed strict donor substrate specificity for UDP-Gal and did not utilize UDP-GalNAc or UDP-GlcNAc with the acceptor substrates tested (data not shown).

Expression of β 4Gal-T3—Expression of the soluble construct of β 4Gal-T3 in Sf9 cells produced a marked increase in Gal-transferase activities with the benzyl- β -GlcNAc acceptor substrate compared with uninfected cells or cells infected with the

control construct for GalNAc-T3 (Table I). A similar increase in activity was found in Sf9 cell homogenates after expression of the full-length construct of β 4Gal-T3 (data not shown). The substrate specificity was similar to that of β 4Gal-T2. β 4Gal-T3 was largely insensitive to α -lactalbumin and did not efficiently utilize glucose (Fig. 5). The apparent K_m for benzyl- β -GlcNAc was 580 μ M, and the K_m for UDP-Gal using benzyl- β -GlcNAc was 84 μ M (Table II). β 4Gal-T3 showed strict donor substrate specificity for UDP-Gal. Analysis of the specificity of β 4Gal-T3 with glycolipid substrates revealed that Lc₃Cer was efficiently utilized, whereas Glc β 1-Cer and nLc₅Cer were poor substrates (Table III). This was in contrast to the activities found with the bovine milk β 4-Gal-transferase that efficiently utilized both Lc₃Cer and nLc₅Cer.

Structural characterization of the product formed with Lc₃Cer by ¹H NMR showed that β 4Gal-T3 forms the same linkage as β 4Gal-T1. One-dimensional ¹H NMR spectroscopy showed that the product was a single glycosphingolipid with a

spectrum virtually identical to that of nLc₄Cer acquired previously under identical conditions (49) and distinct from that of Lc₄Cer (46). In the downfield region of the spectrum (Fig. 7), four distinct β -anomeric resonances ($^3J_{1,2} \cong 7-9$ Hz) were observed at chemical shifts of 4.169 ppm ($^3J_{1,2} = 7.4$ Hz), 4.212 ppm ($^3J_{1,2} = 7.7$ Hz), 4.263 ppm ($^3J_{1,2} = 6.8$ Hz), and 4.663 ppm ($^3J_{1,2} = 8.3$ Hz). These were all within ± 0.004 ppm of those previously observed (71) for I-1, IV-1, II-1, and III-1, respectively, of nLc₄Cer and thus clearly consistent with a β 1-4 linkage of the terminal Gal residue. Under these conditions, the resonance for H-1 of the terminal Gal β 1-3 of Lc₄Cer is found at 4.140 ppm ($^3J_{1,2} = 7.3$ Hz), whereas that of the β -GlcNAc III-1 to which it is linked is found at 4.780 ppm ($^3J_{1,2} = 7.9$ Hz) (46).

Northern Analysis of β 4Gal-T2 and β 4Gal-T3—Northern analysis with mRNA from 16 human adult organs showed a single transcript of both genes of ~ 2.2 kb (Fig. 8). β 4Gal-T2 was expressed weakly in several adult organs, with the highest expression in prostate, testis, ovary, intestine, and muscle. β 4Gal-T3 appeared to be more strongly expressed than β 4Gal-T2, but with a similar pattern, with the exception of placenta, where strong expression of T3 was also found.

Genomic Organization and Chromosomal Localizations—The coding regions of β 4Gal-T2 and β 4Gal-T3 were found in seven and six exons, respectively (Fig. 9). Human β 4Gal-T1 and mouse β 4Gal-T1 are encoded in six exons (62, 63). The first putative coding exon of β 4Gal-T2 encodes the first potential initiation codon and only eight amino acid residues of the N-terminal sequence of the longest form (Fig. 9), and it is

TABLE II
Kinetic properties of β 4-Gal-transferases

Substrate	β 4Gal-T2		β 4Gal-T3		Bovine β 4-Gal-transferase	
	K_m	V_{max}	K_m	V_{max}	K_m	V_{max}
	mM	pmol/min	mM	pmol/min	mM	pmol/min
UDP-Gal	0.011	51.9	0.084	128.7	0.034	146.0
Bzl- β -D-GlcNAc ^a	0.16	88.0	0.58	112.5	0.53	88.2
D-GlcNAc	2.65	59.4	13.1	121.7	9.6	113.6

^a Bzl, benzyl.

FIG. 6. Differential inhibition of β 4-Gal-transferase activities by high acceptor substrate concentrations. A, benzyl- β -GlcNAc; B, GlcNAc. \blacktriangle , bovine milk Gal-transferase; \blacksquare , β 4Gal-T2; \bullet , β 4Gal-T3. Bz β -GlcNAc, benzyl- β -GlcNAc.

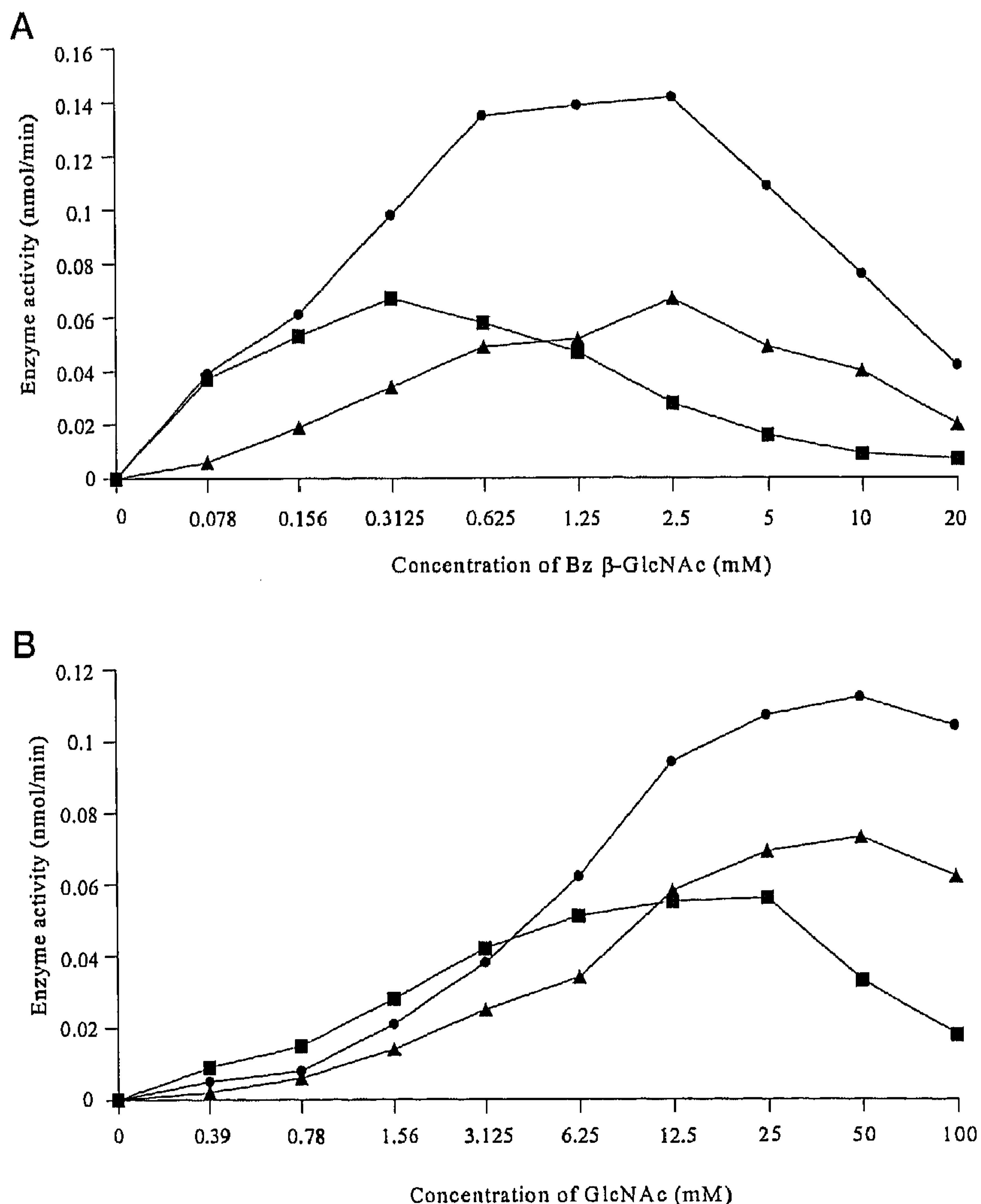


TABLE III
 Substrate specificities with glycolipid acceptors

Acceptor substrate	β 4Gal-T3 ^a nmol/h/ml	Bovine milk β 4-Gal-transferase ^b nmol/h/mg
GlcCer (Glc β 1-Cer)	0.38	40.3(ND) ^c
LacCer (Gal β 1-4Glc β 1-Cer)	ND ^d	ND
Gb ₃ (Gal α 1-4Gal β 1-4Glc β 1-Cer)	ND	ND
Gb ₄ (GalNAc β 1-3Gal α 1-4Gal β 1-4Glc β 1-Cer)	ND	ND
Gg ₃ (GalNAc β 1-4Gal β 1-4Glc β 1-Cer)	ND	ND
G _{M2} (GalNAc β 1-4(NeuAc α 2-3)Gal β 1-4Glc β 1-Cer)	ND	ND
G _{M1} (Gal β 1-3GalNAc β 1-4(NeuAc α 2-3)Gal β 1-4Glc β 1-Cer)	ND	ND
Lc ₃ (GlcNAc β 1-3Gal β 1-4Glc β 1-Cer)	9.7	297 (386) ^c
nLc ₄ (Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc β 1-Cer)	ND	ND
nLc ₅ (GlcNAc β 1-3Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc β 1-Cer)	0.93	333 (58) ^c

^a Assayed using 100 μ g of Triton CF-54/100- μ l reaction mixture (enzyme source medium from infected Sf9 cells).

^b Assayed using 100 μ g of taurodeoxycholate/100- μ l of reaction mixture (enzyme source, Sigma).

^c Assayed using 100 μ g of Triton CF-54/100 μ l of reaction mixture.

^d ND, not detectable.

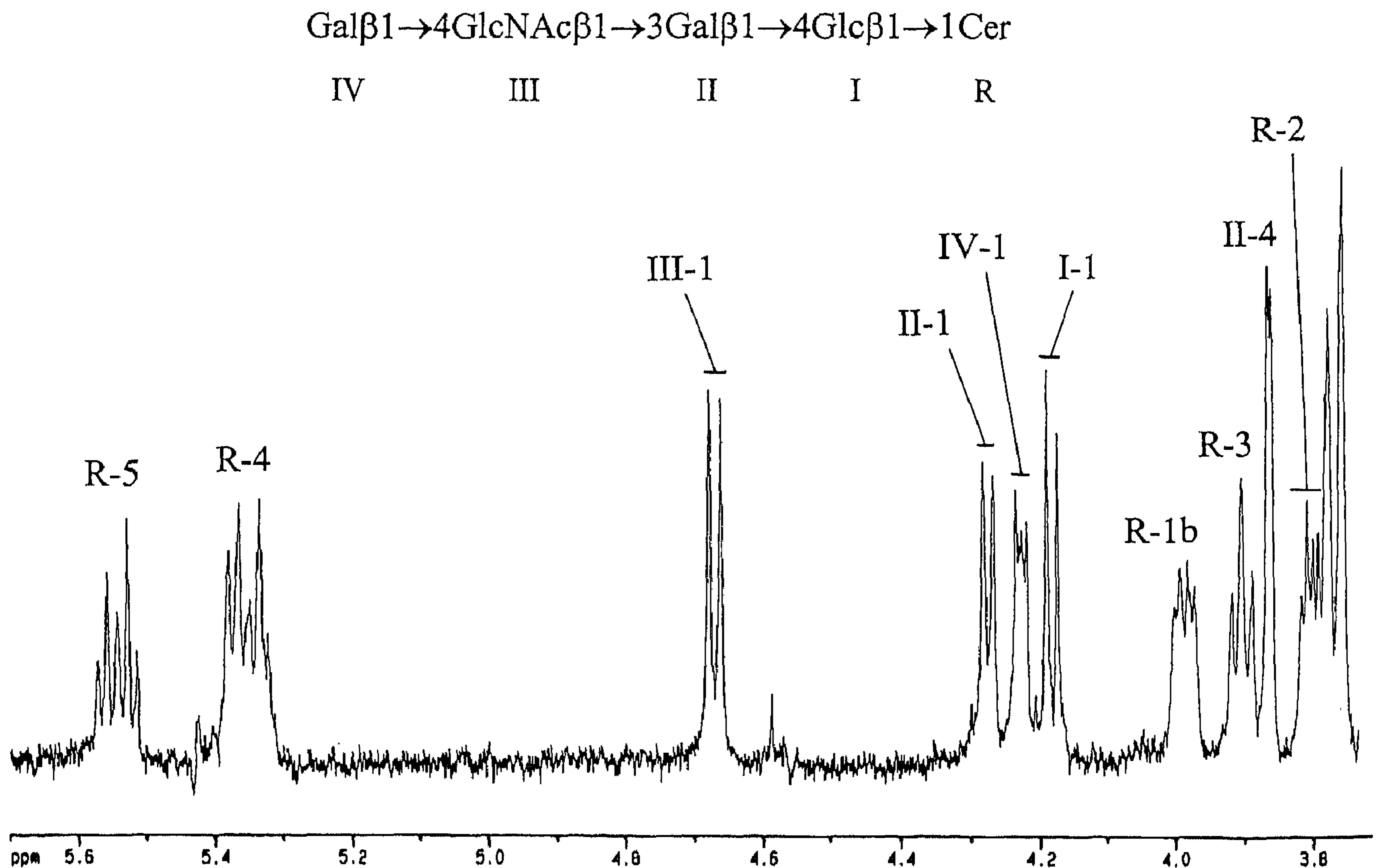


FIG. 7. Downfield region of 500-MHz ^1H NMR spectrum of glycosphingolipid produced by enzymatic glycosylation of Lc₃Cer with β 4Gal-T3. The product isolated by Sep-Pack C₁₈ chromatography was deuterium-exchanged and dissolved in Me₂SO-*d*₆ and 2% D₂O. 2048 free induction decays were accumulated at 308 K. Arabic numerals refer to ring protons of residues designated by Roman numerals or upper-case letters in the corresponding structure. R refers to protons of the sphingosine backbone.

possible that the most 5'-initiation codon is not used. An intron in the 5'-untranslated region of β 4Gal-T3 was found in a similar position as the most 5'-intron in β 4Gal-T2 (Fig. 10). Comparison of the intron/exon boundaries of β 4Gal-T1, β 4Gal-T2, and β 4Gal-T3 revealed that the five introns in the coding regions of the three genes are placed identically (Fig. 10). The central coding exons of all three genes are of nearly identical length, and variation in lengths of coding regions may be attributable to different initiation and stop codons in the first and last coding exons.

Human β 4Gal-T1 was previously localized to chromosome 9p13 (64). The β 4Gal-T2 and β 4Gal-T3 genes were localized to chromosomes 1p32-33 and 1q23 by fluorescence *in situ* hybridization (Fig. 11). Nearly identical localizations were obtained from the EST mapping. No specific hybridization signals were

observed at other chromosomal sites. For each gene, a total of 20 cells in metaphase were analyzed.

DISCUSSION

The data presented here demonstrate that the human β 4Gal-T1 gene is a member of a large family of homologous glycosyltransferase genes and that at least three of the members of this family, β 4Gal-T1, β 4Gal-T2, and β 4Gal-T3, encode β 1,4-galactosyltransferases with similar kinetic parameters. The EST strategy used in this study has successfully been applied to other homologous gene families (22, 23, 65). The short sequence information of \sim 200-400 bp obtained from a single EST, which may contain sequencing artifacts, may not be considered strong evidence for a particular gene sequence of interest. However, in our experience, the reliability of EST

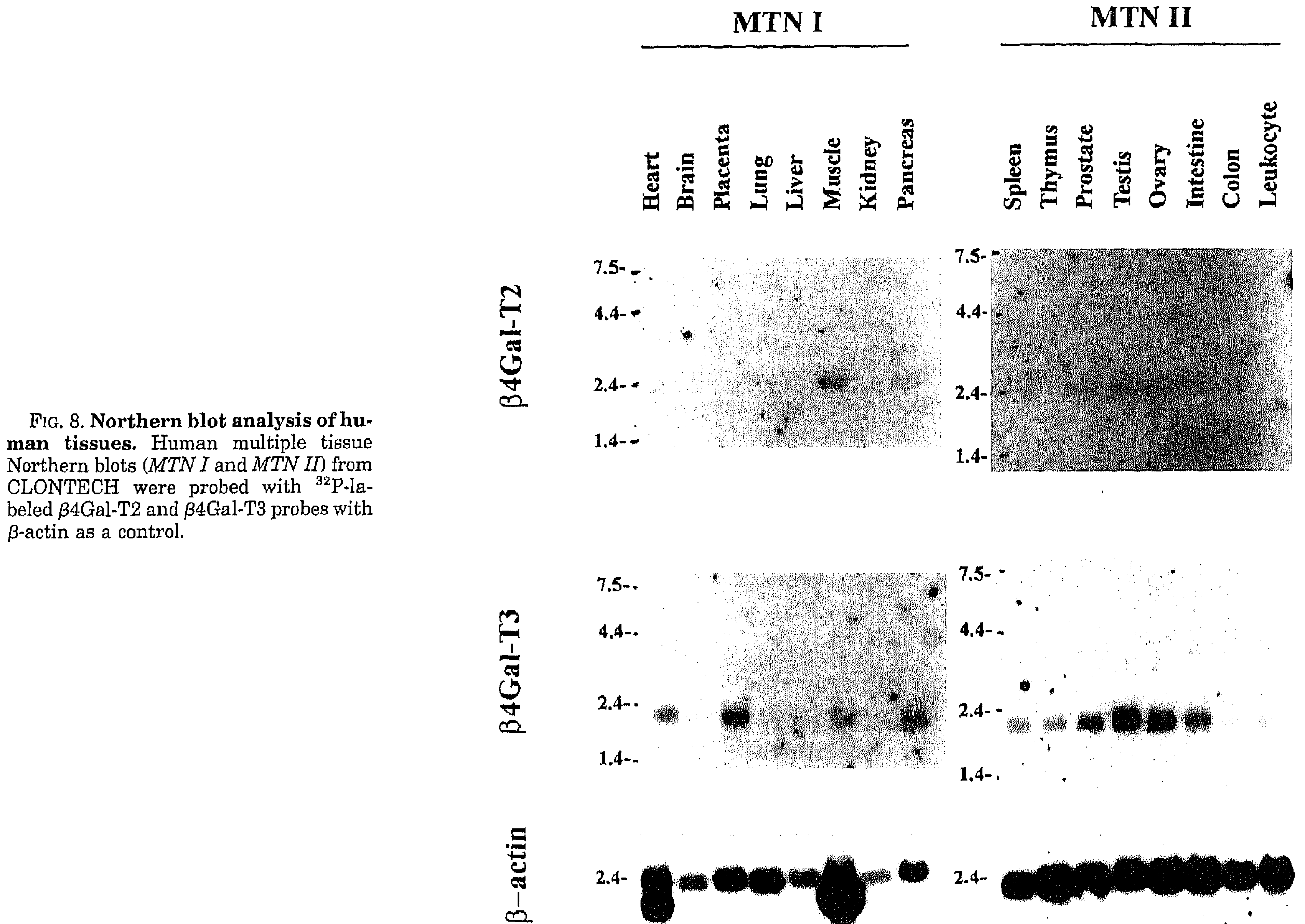


FIG. 8. Northern blot analysis of human tissues. Human multiple tissue Northern blots (*MTN I* and *MTN II*) from CLONTECH were probed with 32 P-labeled β 4Gal-T2 and β 4Gal-T3 probes with β -actin as a control.

sequences is very high, and the finding of similar protein sequence motifs and conserved cysteine residues similar to those of the query sequence strongly indicates the identification of a novel homologous gene. The establishment of the Unigene data base further confirms this by identifying other ESTs from the same gene from which more sequence can be compiled by direct comparison of sequences or with available insert sizes of ESTs. Thus, with the two genes identified in this study, most of the coding region can be identified from merged EST sequences, and a more detailed assessment of the authenticity of the putative homologous gene can be made. One drawback of the EST cloning strategy is that genes with long 3'-untranslated sequences may not be represented, and this has been a feature associated with some glycosyltransferase genes (8, 10, 66). A strong point of the EST strategy is that it offers access to expressed sequences from many different adult and fetal tissues, leading to the possibility of identifying genes with very restricted expression patterns.

Some evidence suggested that β 4Gal-T1 was a member of a homologous gene family. One distant member of this family was found to use UDP-GlcNAc and not UDP-Gal (52), indicating that members may have different donor substrate specificities and functions. Interestingly, UDP-Gal: β -GlcNAc β 1,3-galactosyltransferases have no sequence similarity with β 4Gal-T1 (GenBankTM/EMBL Data Bank Accession number E07739).³ Several features of the two identified genes suggested that they represented genuine glycosyltransferase genes. The genes had higher sequence similarities to β 4Gal-T1

than to a homologous β 4-GlcNAc-transferase from snail (52). Four cysteine residues were conserved in all sequences, and these have been found to be involved in intramolecular disulfide bonding and catalysis for β 4Gal-T1 (31, 32). Finally, the genomic organizations of both genes were shown to be identical to that of β 4Gal-T1, with the positions of five intron/exon boundaries conserved (Fig. 10). The last finding strongly suggests that this gene family arose through complete gene duplication late in evolution. Furthermore, the genomic organization of the homologous snail β 4-GlcNAc-transferase gene was recently shown to be organized similarly, with conservation of the same five intron positions, but additionally, two exons were found as well (Fig. 10) (67). More important, the two chick β 4-Gal-transferases identified by Shaper *et al.* (38) were also found to have genomic organizations identical to that of the β 4Gal-T1 gene. The three β 4-Gal-transferase genes have different chromosomal locations (64). Several large glycosyltransferase gene families have now been found to have different chromosomal locations (23, 24), but some are also clustered in one region (26, 28, 66).

The β 4Gal-T1 gene is different from most other known glycosyltransferase genes in having two translational initiation sites controlled by two different promoters (35, 68). The two isoforms of β 4Gal-T1 differ by having a short (11 residues) or long (24 residues) cytoplasmic sequence, and studies indicate that the short form primarily localizes to the Golgi apparatus and the long form primarily to the cell membrane (69). A number of biological roles have been attributed to a cell membrane form of β 4Gal-T1 (70). The β 4Gal-T2 gene, showing the highest similarity to β 4Gal-T1, also resembles β 4Gal-T1 in

³ M. Amado and H. Clausen, manuscript in preparation.

				β 4Gal-T2					
	S	C	M	Intron 1		A	S	S	
C	TCC	TGC	ATG ₂₄	gtaggtgac.....ccgtgctcag	GCC	AGC	AGC	AGC	C
	P	G	L	Intron 2		V	G	R	L
CCT	GGT	CTT	G ₃₈₈	gtgagcctgg.....ctgtctgcag	TG	GGC	AGA	CT	
	I	N	Q	Intron 3		H	G	E	
C	ATC	AAC	CAG ₆₂₄	gtgcccattgc.....ccctaccag	CAT	GGT	GAG	G	
	F	G	R	Intron 4		L	P	Y	
TT	GGC	TTC	CG ₈₁₅	gtgagggctc.....catgccacag	G	CTT	CCC	TAT	
	I	F	R	Intron 5		I	S	L	
TC	TTC	AAC	CG ₉₃₈	gtgagtaagc.....ccatcctcag	G	ATC	TCC	CTG	
	N	P	R	Intron 6		F	T	K	
AC	CCT	CAG	AG ₁₀₄₃	gtgaccccag.....tcctatctag	G	TTT	ACC	AAG	

FIG. 9. Intron junctions in the coding regions of the β 4Gal-T2 and β 4Gal-T3 genes. Exon sequences are shown in upper-case letters, with the nucleotide positions from the initiation codon in subscript and the predicted amino acid sequence in single-letter code above the sequence. Flanking intron sequences are shown in lower-case letters, and introns are labeled according to Fig. 1. Sequences were aligned to the best fit of the gt/ag consensus rule.

				β 4Gal-T3					
				Intron 1					
TTG	AAC	CTA	G ₋₃₄	gtgagtctat.....ctaccccag	AA	TTC	CCC	TC	
	P	L	L	Intron 2		V	L	P	V
CCT	CTC	TTA	G ₂₅₃	gtaagtgcag.....ttcccaacag	TG	GGT	CCT	GT	
	I	H	Q	Intron 3		A	G	N	
C	ATC	CAC	CAG ₄₈₉	gtaccccacaa.....tatcttccag	GCT	GGA	AAT	G	
	F	G	S	Intron 4		L	P	Y	
TT	GGA	TAC	AG ₆₈₀	gtacagggca.....tctccctcag	C	CTC	CCG	TAC	
	I	A	R	Intron 5		V	R	L	
TT	GCT	ACC	AG ₈₀₃	gtcagacttc.....tttcccatag	G	GTG	CGC	CTG	
	N	P	R	Intron 6		F	D	L	
AT	CCC	CAC	AG ₉₀₈	gtaggaagaa.....taaccctcag	A	TTT	GAC	CTC	

having three initiation codons, potentially yielding isoforms with 25-, 30-, or 37-residue cytoplasmic sequences. A sequence motif (L(E/Q)RXC) is found in the cytoplasmic sequences of the human and chick genes immediately preceding the hydrophobic putative transmembrane signal sequence (Fig. 4). Since the major parts of the N-terminal regions of these genes show little or no sequence similarities, this conserved motif may be important. Bendiak *et al.* (71) have shown that the β 4-Gal-transferase appears to exist as a high molecular weight aggregate, and this motif could potentially play a role in such aggregation. Aoki *et al.* (34) showed that a cysteine residue in the hydrophobic transmembrane sequence was important for Golgi retention, and this cysteine is conserved in β 4Gal-T2, but not β 4Gal-T3 or the snail β 4-GlcNAc-transferase (Fig. 4). The implications of this are not known at present.

The fine specificity of the three β 4-Gal-transferases for different types of glycoconjugates and branch points of oligosaccharide structures needs to be determined in comparative assays with recombinant forms of the enzymes. Here, we have correlated the activities of recombinant soluble β 4Gal-T2 and β 4Gal-T3 with purified bovine milk enzyme, which potentially contains other β -Gal-transferase activities than β 4Gal-T1. The substrate specificities with monosaccharides and simple derivatives were similar (Table I). β 4Gal-T2 had a lower K_m for UDP-Gal than both bovine milk β 4-Gal-transferase and β 4Gal-T3 (Table II) and also a lower K_m for the best acceptor substrates (benzyl- β -GlcNAc and GlcNAc) used in this study. At high concentrations of acceptor substrates (benzyl- β -GlcNAc and GlcNAc), all three enzymes showed inhibition of activities, but β 4Gal-T2 was the most sensitive (Fig. 6). Inhibition of β 4Gal-T1 was previously noted at concentrations in excess of 20–30 mM (61), but in the present study, this was seen at higher concentrations.

The substrate specificity of β 4Gal-T1 is modulated by α -lactalbumin (72, 73). A similar modulation of the substrate specificity was found for β 4Gal-T2, whereas β 4Gal-T3 was largely insensitive (Fig. 5). Interestingly, mice rendered deficient in β 4Gal-T1 were shown not to produce lactose (39, 40). This suggests that β 4Gal-T2 is not expressed in lactating mammary glands. Differences in sensitivity to α -lactalbumin modulation of putative members of the β 4-Gal-transferase family have previously been observed. Two distinct porcine trachea β 4-Gal-transferases have been separated, and one of these was shown to be insensitive to α -lactalbumin, similar to β 4Gal-T3 (36). The homologous β 4-GlcNAc-transferase identified in snail is apparently not sensitive to α -lactalbumin (74); however, another putative member of this enzyme family, a β 4-GalNAc-transferase, was found to be modulated by α -lactalbumin (75).

As β 4Gal-T1 and β 4Gal-T2 were most similar in sequence and seemed to have similar properties, the specificity of β 4Gal-T3 was studied in more detail. Analysis of β 4Gal-T3 and bovine milk β -Gal-transferase with glycolipid substrates showed that both enzymes utilize Lc₃Cer efficiently, and characterization of the product formed by β 4Gal-T3 using ¹H NMR clearly established the expected product as being nLc₄ with a Gal β 1-4GlcNAc linkage. Interestingly, marked differences in activities were found with the extended substrate nLc₅Cer (Table III). In contrast to the bovine milk enzyme, β 4Gal-T3 poorly utilized nLc₅Cer, which may suggest that β 4Gal-T1 is involved in all steps of the synthesis of poly-*N*-acetyllactosamine chain extension, whereas β 4Gal-T3 may be mainly involved in the synthesis of the first *N*-acetyllactosamine unit. Differential expression of β 4-Gal-transferases may be part of the regulation of poly-*N*-acetyllactosamine chain extension, which appears to be developmentally regulated (76, 77).

The differences found in the kinetic properties of the three

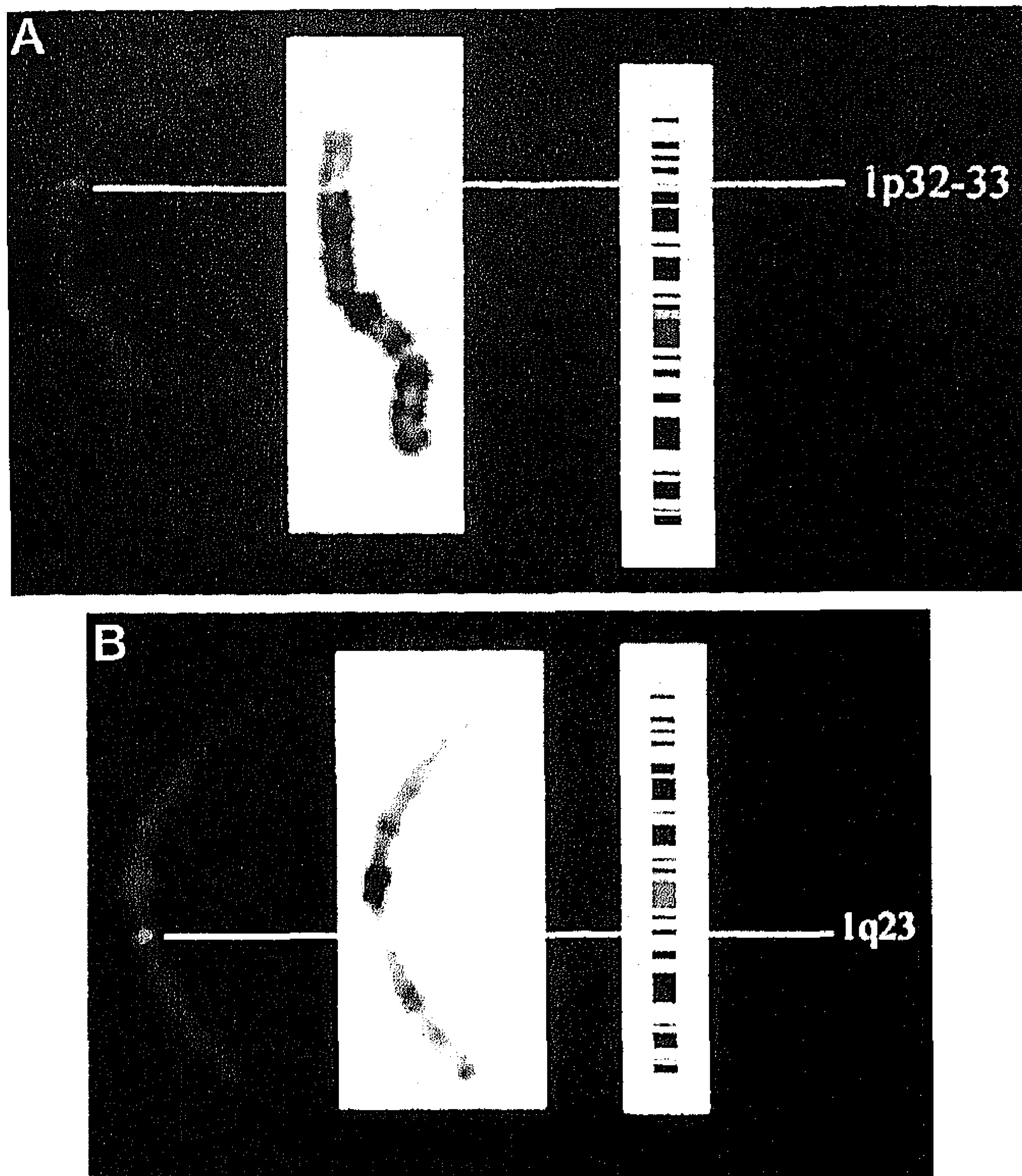


FIG. 11. Fluorescence *in situ* hybridization of β 4Gal-T2 and β 4Gal-T3 to metaphase chromosomes. A, the β 4Gal-T2 probe (P1 DNA from clone DPMC-HFF10638:515:G9) labeled band 1p32-33. B, the β 4Gal-T3 probe (P1 DNA from clone DPMC-HFF1-0288-D2) labeled 1q23.

concentrations may be misleading for the actual activity present. Since mice deficient in β 4Gal-T1 developed normally before birth and only exhibited reduced growth and lethality after birth, it is likely that other β 4-Gal-transferases besides β 4Gal-T1 are functioning in fetal life. Analysis of serum glycoproteins of 10-week-old mice deficient in β 4Gal-T1 showed a marked decrease in galactosylation (40), suggesting that the major β 4-Gal-transferase involved in the synthesis of these glycoproteins is β 4Gal-T1. In contrast, Lu *et al.* (39) found that despite a lack of detectable β 4-Gal-transferase activity in salivary glands, a normal level of Gal β 1-4GlcNAc terminating chains in salivary gland extracts was found by lectin analysis and α 2,6-sialylation. It is therefore likely that another β 4-Gal-transferase besides β 4Gal-T1 is active in salivary glands.

It is clear that the single step of forming the Gal β 1-4GlcNAc β linkage is performed by a family of β 4-Gal-transferases and that each member of this family may play a distinct role in different cells. A similar finding was made for the initiation step of GalNAc and Man *O*-glycosylation in animals and yeast. Large families of homologous enzymes with distinct but overlapping specificities and different expression patterns perform a reaction previously considered to be maintained by a single enzyme (23, 54). The α 2-fucosyltransferases (82), α 3/4-fucosyltransferases (83), and some of the sialyltransferases (24) are other examples of this phenomenon of partial redundancy. The findings that the genes in most glycosyltransferase families have diverged considerably and show only 30–50% sequence similarities combined with differential expression patterns suggest that each member of these gene families plays a distinct role. The seemingly similar enzyme activities determined in *in vitro* assays may not fully reveal the *in vivo* functions of these enzymes.

From the present status of cloning of glycosyltransferases, one may predict that there is no requirement for primary sequence similarities between glycosyltransferases utilizing

the same donor or acceptor substrates *per se*. Analysis of the known glycosyltransferase gene families shows that these have common specificities for the nucleotide of the donor substrate and the anomeric configuration of the linkage formed, whereas the acceptor substrate and the linkage positions formed seem to be more variable. The sialyltransferases all utilize CMP-NeuAc and form α -anomeric linkages, but they form linkages to several different acceptors (Gal, GalNAc, GlcNAc, and NeuAc) and at different positions (C-3, C-6, and C-8) (24). Although all members of this gene family share some sequence similarity, it is apparent that members forming the same linkage type have higher sequence similarity. The α 1,3/4-fucosyltransferase family utilizes GDP-Fuc and forms either α 1-3 or α 1-4 linkages to GlcNAc (84). The α 3-GalNAc-transferase family utilizes UDP-Gal or UDP-GalNAc and forms α 1-3 linkages to Gal (27). The β 6-GlcNAc-transferases use different acceptor substrates (28). The polypeptide GalNAc-transferase family generally utilizes UDP-GalNAc, but one member (GalNAc-T2) was recently found also to utilize UDP-Gal (23, 44). Finally, the snail β 4-GlcNAc-transferase homologues to the β 4-Gal-transferase family utilize UDP-GlcNAc, but form the same linkage. Therefore, specific functions of a newly identified homologous gene in a glycosyltransferase family are not easily predicted even though homologous glycosyltransferases recognize the same nucleotide of the donor substrate and form the same anomeric configuration of linkages.

The existence of multiple β 4-Gal-transferases has implications for interpretation of past studies of the expression pattern and biological functions previously assigned to β 4Gal-T1; this is what was learned from analysis of mice deficient in β 4Gal-T1. Detection of gene expression of β 4Gal-T1 by Northern or *in situ* hybridization is likely to be representative of genuine β 4Gal-T1 expression, but analysis of enzyme activities in extracts or body fluids may represent all or any of the β 4-Gal-transferases, and knowledge of the kinetic parameters of each of the enzymes is required to ascertain that all activities are measured. Immunodetection with antibodies prepared against purified rather than recombinant β 4-Gal-transferases may likewise be directed toward any or several of the enzymes in the family.

Acknowledgments—We thank Drs. M. A. Hollingsworth and E. G. Berger for many helpful suggestions and critical reading of the manuscript and Dr. M. Sobrinho-Simoes for continuous support throughout the project.

Note Added in Proof—A fourth human β 4-galactosyltransferase was recently cloned and expressed. (Sato, T., Furukawa, K., Bakker, H., Van den Eijnden, D. H. (1997) *Glycoconj. J.* 14, (suppl.) 46). This gene showed lower sequence similarity to β 4 Gal-T1 than the two genes reported here, and the expressed enzyme was not alactalbumin-sensitive.

REFERENCES

- Kleene, R., and Berger, E. G. (1993) *Biochim. Biophys. Acta* 1154, 283–325
- Schachter, H. (1997) in *Molecular Glycobiology* (Fukuda, M., and Hindsgaul, O., eds) pp. 88–149, IRL Press Ltd., Oxford
- Shaper, N. L., Shaper, J. H., Meuth, J. L., Fox, J. L., Chang, H., Kirsch, I. R., Hollis, and G. F. (1986) *Proc. Natl. Acad. Sci. U. S. A.* 83, 1573–1577
- Narimatsu, H., Sinha, S., Brew, K., Okayama, H., and Qasba, P. K. (1986) *Proc. Natl. Acad. Sci. U. S. A.* 83, 4720–4724
- D'Agostaro, G., Bendiak, B., and Tropak, M. (1989) *Eur. J. Biochem.* 183, 211–217
- Shur, B. D., and Neely, C. A. (1988) *J. Biol. Chem.* 263, 17706–17714
- Nakazawa, K., Ando, T., Kimura, T., and Narimatsu, H. (1988) *J. Biochem. (Tokyo)* 104, 165–168
- Weinstein, J., Lee, E. U., McEntee, K., Lai, P. H., and Paulson, J. C. (1987) *J. Biol. Chem.* 262, 17735–17743
- Joziassse, D. H., Shaper, J. H., Van den Eijnden, D. H., Van Tunen, A. J., and Shaper, N. L. (1989) *J. Biol. Chem.* 264, 14290–14297
- Yamamoto, F., Marken, J., Tsuji, T., White, T., Clausen, H., and Hakomori, S. (1990) *J. Biol. Chem.* 265, 1146–1151
- Sarkar, M., Hull, E., Nishikawa, Y., Simpson, R. J., Moritz, R. L., Dunn, R., and Schachter, H. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 234–238
- Larsen, R. D., Rajan, V. P., Ruff, M. M., Kukowska-Latallo, J., Cummings, R. D., and Lowe, J. B. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 8227–8231
- Rajan, V. P., Larsen, R. D., Ajmera, S., Ernst, L. K., and Lowe, J. B. (1989)

- J. Biol. Chem.* **264**, 11158–11167
14. Larsen, R. D., Ernst, L. K., Nair, R. P., and Lowe, J. B. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6674–6678
 15. Kukowska-Latallo, J. F., Larsen, R. D., Nair, R. P., and Lowe, J. B. (1990) *Genes Dev.* **4**, 1288–1303
 16. Kumar, R., Yang, J., Larsen, R. D., and Stanley, P. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 9948–9952
 17. Fukuda, M., Bierhuizen, M. F., and Nakayama, J. (1996) *Glycobiology* **6**, 683–689
 18. Livingston, B. D., and Paulson, J. C. (1993) *J. Biol. Chem.* **268**, 11504–11507
 19. Bennett, E. P., Hassan, H., and Clausen, H. (1996) *J. Biol. Chem.* **271**, 17006–17012
 20. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., and Kerlavage, A. R., McCombie, W. R., and Venter, J. C. (1991) *Science* **252**, 1651–1656
 21. Gerhold, D., and Caskey, C. T. (1996) *Bioessays* **18**, 973–981
 22. Hagen, F. K., Gregoire, C. A., and Tabak, L. A. (1995) *Glycoconjugate J.* **12**, 901–909
 23. Clausen, H., and Bennett, E. P. (1996) *Glycobiology* **6**, 635–646
 24. Tsuji, S., Datta, A. K., and Paulson, J. C. (1996) *Glycobiology* **6**, v–vii
 25. Datta, A. K., and Paulson, J. C. (1995) *J. Biol. Chem.* **270**, 1497–1500
 26. Lowe, J. B. (1993) *Bailliere's Clin. Haematol.* **6**, 465–492
 27. Haslam, D. B., and Baenziger, J. U. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10697–10702
 28. Bierhuizen, M. F., Maemura, K., Kudo, S., and Fukuda, M. (1995) *Glycobiology* **5**, 417–425
 29. Schulte, S., and Stoffel, W. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10265–10269
 30. Drickamer, K. (1993) *Glycobiology* **3**, 2–3
 31. Wang, Y., Wong, S. S., Fukuda, M. N., Zu, H., Liu, Z., Tang, Q., and Appert, H. E. (1994) *Biochem. Biophys. Res. Commun.* **204**, 701–709
 32. Boeggeman, E. E., Balaji, P. V., and Qasba, P. K. (1995) *Glycoconjugate J.* **12**, 865–878
 33. Aoki, D., Appert, H. E., Johnson, D., Wong, S. S., and Fukuda, M. N. (1990) *EMBO J.* **9**, 3171–3178
 34. Aoki, D., Lee, N., Yamaguchi, N., Dubois, C., and Fukuda, M. N. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4319–4323
 35. Shaper, N. L., Hollis, G. F., Douglas, J. G., Kirsch, I. R., and Shaper, J. H. (1988) *J. Biol. Chem.* **263**, 10420–10428
 36. Sheares, B. T., and Carlson, D. M. (1984) *J. Biol. Chem.* **259**, 8045–8047
 37. Furukawa, K., Matsuta, K., Takeuchi, F., Kosuge, E., Miyamoto, T., and Kobata, A. (1990) *Int. Immunol.* **2**, 105–112
 38. Shaper, J. H., Joziassse, D. H., Meurer, J. A., Chou, T.-D. D., Schnaar, R. A., and Shaper, N. L. (1995) *Glycoconjugate J.* **12**, 477
 39. Lu, Q., Hasty, P., and Shur, B. D. (1997) *Dev. Biol.* **181**, 257–267
 40. Asano, M., Furukawa, K., Kido, M., Matsumoto, S., Umesaki, Y., Kochibe, N., and Iwakura, Y. (1997) *EMBO J.* **16**, 1850–1857
 41. Uehara, K., and Muramatsu, T. (1997) *Eur. J. Biochem.* **244**, 706–712
 42. Nakazawa, K., Furukawa, K., Narimatsu, H., and Kobata, A. (1993) *J. Biochem. (Tokyo)* **113**, 747–753
 43. Masri, K. A., Appert, H. E., and Fukuda, M. N. (1988) *Biochem. Biophys. Res. Commun.* **157**, 657–663
 44. Wandall, H. H., Hassan, H., Mirgorodskaya, E., Kristensen, A. K., Roepstorff, P., Bennett, E. P., Nielsen, P. A., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (1997) *J. Biol. Chem.* **272**, 23503–23514
 45. Chien, J. L., Li, S. C., Laine, R. A., and Li, Y. T. (1978) *J. Biol. Chem.* **253**, 4031–4035
 46. Holmes, E. H., and Lavery, S. B. (1989) *Arch. Biochem. Biophys.* **274**, 633–647
 47. Lavery, S. B., Nudelman, E. D., and Hakomori, S. (1992) *Biochemistry* **31**, 5335–5340
 48. Dabrowski, J., Hanfland, P., and Egge, H. (1980) *Biochemistry* **19**, 5652–5658
 49. Lavery, S. B., Nudelman, E., Kannagi, R., Symington, F. W., Andersen, N. H., Clausen, H., Baldwin, M., and Hakomori, S. (1988) *Carbohydr. Res.* **178**, 121–144
 50. Breathnach, R., and Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383
 51. Suijkerbuijk, R. F., van de Veen, A. Y., van Echten, J., Buys, C. H., de Jong, B., Oosterhuis, J. W., Warburton, D. A., Cassiman, J. J., Schonk, D., and Geurts van Kessel, A. (1991) *Am. J. Hum. Genet.* **48**, 269–273
 52. Bakker, H., Agterberg, M., Van Tetering, A., Koeleman, C. A., Van den Eijnden, D. H., and Van Die, I. (1994) *J. Biol. Chem.* **269**, 30326–30333
 53. Kozak, M. (1992) *Annu. Rev. Cell Biol.* **8**, 197–225
 54. Gentsch, M., and Tanner, W. (1996) *EMBO J.* **15**, 5752–5759
 55. Nakazawa, K., Furukawa, K., Kobata, A., and Narimatsu, H. (1991) *Eur. J. Biochem.* **196**, 363–368
 56. Paquet, M. R., and Moscarello, M. A. (1984) *Biochem. J.* **218**, 745–751
 57. Fujita-Yamaguchi, Y., and Yoshida, A. (1981) *J. Biol. Chem.* **256**, 2701–2706
 58. Malissard, M., Borsig, L., Di Marco, S., Grutter, M. G., Kragl, U., Wandrey, C., and Berger, E. G. (1996) *Eur. J. Biochem.* **239**, 340–348
 59. Moscarello, M. A., Mitranic, M. M., and Vella, G. (1985) *Biochim. Biophys. Acta* **831**, 192–200
 60. Powell, J. T., and Brew, K. (1974) *Eur. J. Biochem.* **48**, 217–228
 61. Shur, B. D. (1982) *J. Biol. Chem.* **257**, 6871–6878
 62. 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
 63. Mengle-Gaw, L., McCoy-Haman, M. F., and Tiemeier, D. C. (1991) *Biochem. Biophys. Res. Commun.* **176**, 1269–1276
 64. Shaper, N. L., Shaper, J. H., Bertness, V., Chang, H., Kirsch, I. R., and Hollis, G. F. (1986) *Somatic Cell Mol. Genet.* **12**, 633–636
 65. Koch-Nolte, F., Haag, F., Braren, R., Kuhl, M., Hoovers, J., Balasubramanian, S., Bazan, F., and Thiele, H. G. (1997) *Genomics* **39**, 370–376
 66. White, T., Bennett, E. P., Takio, K., Sorensen, T., Bonding, N., and Clausen, H. (1995) *J. Biol. Chem.* **270**, 24156–24165
 67. Bakker, H., Van Tetering, A., Agterberg, M., Smit, A. B., Van den Eijnden, D. H., and Van Die, I. (1997) *J. Biol. Chem.* **272**, 18580–18585
 68. Russo, R. N., Shaper, N. L., and Shaper, J. H. (1990) *J. Biol. Chem.* **265**, 3324–3331
 69. Youakim, A., Dubois, D. H., and Shur, B. D. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10913–10917
 70. Dubois, D. H., and Shur, B. D. (1995) *Adv. Exp. Med. Biol.* **376**, 105–114
 71. Bendiak, B., Ward, L. D., and Simpson, R. J. (1993) *Eur. J. Biochem.* **216**, 405–417
 72. Brodbeck, U., Denton, W. L., Tanahashi, N., and Ebner, K. E. (1967) *J. Biol. Chem.* **242**, 1391–1397
 73. Brew, K., Vanaman, T. C., and Hill, R. L. (1968) *Proc. Natl. Acad. Sci. U. S. A.* **59**, 491–497
 74. Bakker, H., Schoenmakers, P. S., Koeleman, C. A., Joziassse, D. H., Van Die, I., and Van den Eijnden, D. H. (1997) *Glycobiology* **7**, 539–548
 75. Neeleman, A. P., van der Knaap, W. P., and Van den Eijnden, D. H. (1994) *Glycobiology* **4**, 641–651
 76. Kobata, A. (1992) *Eur. J. Biochem.* **209**, 483–501
 77. Dennis, J. W. (1991) *Semin. Cancer Biol.* **2**, 411–420
 78. Keusch, J., Lydyard, P. M., Isenberg, D. A., and Delves, P. J. (1995) *Glycobiology* **5**, 365–700
 79. Axford, J. S., Alavi, A., Bond, A., and Hay, F. C. (1994) *Autoimmunity* **17**, 157–163
 80. Wilson, I. B., Platt, F. M., Isenberg, D. A., and Rademacher, T. W. (1993) *J. Rheumatol.* **20**, 1282–1287
 81. Shaper, J. H., Harduin-Lepers, A., Rajput, B., and Shaper, N. L. (1995) *Adv. Exp. Med. Biol.* **376**, 95–104
 82. Kelly, R. J., Rouquier, S., Giorgi, D., Lennon, G. G., and Lowe, J. B. (1995) *J. Biol. Chem.* **270**, 4640–4649
 83. Smith, P. L., Gersten, K. M., Petryniak, B., Kelly, R. J., Rogers, C., Natsuka, Y., Alford, J. A., III, Scheidegger, E. P., Natsuka, S., and Lowe, J. B. (1996) *J. Biol. Chem.* **271**, 8250–8259
 84. Mollicone, R., Cailleau, A., and Oriol, R. (1995) *Transf. Clin. Biol.* **2**, 235–242
 85. Kyte, J., Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132