GENOMICS 34, 69–75 (1996)
ARTICLE NO. 0242

The Human Gene CGT Encoding the UDP-Galactose Ceramide Galactosyl Transferase (Cerebroside Synthase): Cloning, Characterization, and Assignment to Human Chromosome 4, Band q26

ANDREAS BOSIO,* ERIKA BINCZEK,* MICHELLE M. LE BEAU,†
ANTHONY A. FERNALD,† AND WILHELM STOFFEL*,[1]

*Institut für Biochemie, Medizinische Fakultät, Universität Köln, Joseph-Stelzmann-Strasse 52, D-50931 Cologne, Germany; and †Department of Medicine, Section of Hematology/Oncology, University of Chicago, Chicago, Illinois 60637

Received July 26, 1995; accepted November 6, 1995

We have previously cloned the human UDP-galactose ceramide galactosyltransferase (CGT, E.C. 2.4.1.45) cDNA. Its open reading frame encodes the key enzyme in the biosynthesis of the glycosphingolipids, cerebrosides and sulfatides, essential constituents of the myelin membrane of the central nervous system (CNS) and PNS. Expression of the CGT gene and of the myelin-specific proteins in the terminal differentiated oligodendrocyte of CNS and in Schwann cells of PNS is cell-specific and highly time-regulated. The CGT gene therefore is important in the differentiation program of the oligodendrocyte lineage. Here we report the structural organization and the chromosomal localization of the human CGT gene.The coding sequence is separated into five exons, which are distributed over >40 kb. The CGT locus was mapped to the distal region of human chromosome 4, band q26. The organization of the CGT gene and of the UGT (uridylglucuronosyltransferases) gene family suggests a correlation to functional domains of the encoded proteins.

INTRODUCTION

Oligodendrocytes are the myelinating glia cells of the central nervous system (CNS). The genes of a number of myelin membrane proteins, e.g., proteolipid protein (PLP), myelin basic protein (MBP), myelin-associated glycoprotein (MAG), and myelin oligodendrocyte glycoprotein (MOG), and enzymes responsible for the biosynthesis of myelin lipids are expressed specifically in oligodendrocytes in a strictly time-regulated manner and provide the structural components for the formation of the highly ordered multilayer myelin membranes (for review see Morell *et al.*, 1994).

The myelin membrane of CNS contains two major proteins, the peripheral membrane protein MBP and the integral membrane protein PLP, which contribute more than 90% of the total myelin membrane proteins, approximately 20–25% of the dry weight of myelin. The molecular biology of these, minor constituents such as MAG and MOG of CNS, and the protein constituents of the myelin sheath of peripheral nerves produced by Schwann cells has been unravelled in recent years (Lemke, 1993).

The lipid bilayer of the myelin membrane contains cholesterol and complex phospho- and sphingolipids as the main lipid constituents. They contribute more than 75% of the dry weight of CNS and PNS myelin. The most abundant myelin membrane sphingolipids are cerebrosides (gal C, ceramide (β (1-1) galactopyranoside) and their 3-sulfate esters, the sulfatides. Their occurrence is mostly confined to the myelin membrane of CNS and PNS, but they are also present in small amounts in kidney.

The key enzymatic step in the biosynthesis of the galactocerebrosides consists of the transfer of galactose to ceramide catalyzed by the UDP-galactose ceramide galactosyltransferase (CGT, E.C. 2.4.1.45) (Morell and Radin, 1969).

We have recently isolated from myelinating rat brain and purified to homogeneity the UDP-galactose:ceramide galactosyltransferase, a 64-kDa glycoprotein residing in the endoplasmic reticulum (Schulte and Stoffel, 1993). Protein analytical data facilitated the isolation of CGT-specific cDNA clones with an open reading frame of 1623 bp encoding the 541-aa residue core protein.

The N-terminal 20-aa residue signal sequence, three

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. U30930, U31353, U31461, U31658, U31861, and U32370.

[1] To whom correspondence and reprint requests should be addressed. Telephone: +49-221-478-6980. Fax: +49-221-478-6979. E-mail: akc03@rs1.rrz.uni-koeln.de.

0888-7543/96 $18.00

putative N-glycosylation sites, a C-terminal 23-aa residue transmembrane domain, and the KKXK endoplasmic reticulum retention signal are consistent with physical and topological data derived from the purified enzyme (Schulte and Stoffel, 1993). These data were later confirmed by Stahl *et al.* (1994).

The gene encoding CGT is the first of the enzymes involved in complex lipid biosynthesis in the myelinating oligodendrocyte. The product of the CGT reaction gal C has been a valuable tool for studying oligodendrocyte lineage (Bansal and Pfeiffer, 1989).

We report here the isolation of the human CGT gene, present as a single copy in the human genome. The sequence of the open reading frame of CGT-specific cDNA is distributed over five exons in a gene with an approximate length of >40 kb. The CGT locus was assigned to chromosome 4 by primer-specific PCR screening of a somatic cell hybrid panel and more specifically with higher resolution to 4q26 by fluorescence *in situ* hybridization (FISH). So far no linkage to a specific genetic disease has been established.

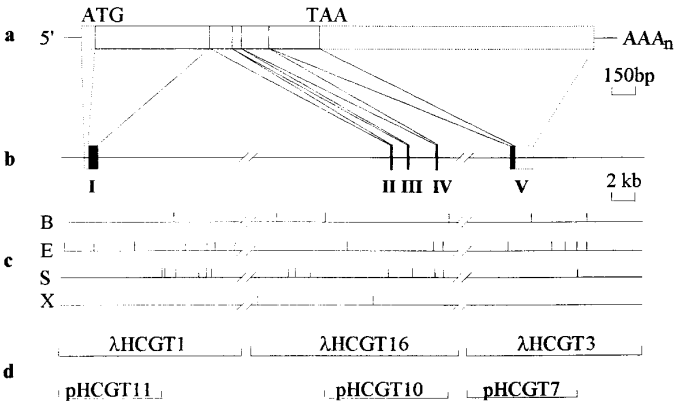


FIG. 1. Structure of the human CGT gene. (a) Human CGT mRNA. (b) Arrangement of the five exons (depicted by numbers and boxes). Solid boxes represent coding regions, 5′- and 3′-untranslated regions are represented by open boxes. (c) Result of the restriction mapping using the enzymes *Bam*HI (B), *Eco*RI (E), *Sac*I (S), and *Xho*I (X). (d) The positions of the three genomic phage clones λHCGT1, λHCGT16, and λHCGT3 and the subclones pHCGT11, pHCGT10, and pHCGT7 used for sequencing.

MATERIALS AND METHODS

Genomic cloning. Three independent genomic clones were obtained by screening a human placenta genomic DNA library (HL 1067j; Clontech). An *Eco*RI–*Nco*I fragment (1190 bp) and a *Nco*I fragment (690 bp) isolated from the rat CGT cDNA clone RNCGT1 (Schulte and Stoffel, 1993) and a mouse genomic DNA fragment derived from pMCGT2 (unpublished data) served as hybridization probes, yielding the genomic clones λHCGT1, λHCGT16, and λHCGT3. Phage DNA of the three clones was prepared by established procedures (Maniatis *et al.*, 1982) and partially mapped using the enzymes *Bam*HI, *Eco*RI, *Sac*I, and *Xho*I. For further characterization the entire phage inserts were subcloned as *Bam*HI, *Eco*RI, *Sac*I, or *Xho*I fragments into the vector pBluescript II SK(+) (Stratagene), denoted pHCGTx (x = 1–30), and if necessary submapped with *Pst*I, *Eco*RV, *Cla*I, *Hin*dIII, *Acc*I, *Dra*II, and *Xba*I. Genomic fragments containing coding sequences were identified by Southern blot analysis using cDNA fragments and oligonucleotides derived from rat CGT cDNA (Fig. 1). Sequencing was carried out by the dideoxy chain termination method (Sanger *et al.*, 1977) using the T7 sequencing kit (Pharmacia) with synthetic oligonucleotides derived from rat cDNA, mouse CGT cDNA, or established sequences.

Genomic Southern blot and PCR amplifications. For estimating the size of the gaps between the three genomic clones λHCGT1, λHCGT16, and λHCGT3, human leukocyte-derived genomic DNA from blood of a Caucasian healthy volunteer was digested using *Bam*HI, *Eco*RI, *Sac*I, and *Xho*I and separated on a 0.7% agarose gel. Fragments were transferred to a nylon membrane by capillary blotting and hybridized with either a 1.4-kb *Eco*RI/*Sac*I fragment of λHCGT1 (containing the 3′ end of the genomic insert) or a 0.6-kb *Sac*I/*Xho*I fragment of λHCGT16 (containing the 5′ end of the genomic insert). The gap in intron IV could not be determined with the available probes. The size of intron I is >29 kb, and intron IV is >6 kb. Size determination of introns II and III was checked by polymerase chain reaction (PCR) using appropriate primer pairs derived from exons II, III, and IV in the subclone pHCGT10.

Chromosomal assignment of human CGT. The sense primer CP81 5′-TCATCATTCTGTGTTTTGTCCCCTCTCC-3′ complementary to a sequence of intron IV (322 bp upstream of exon V) and the antisense primer CP82R 5′-ATAATAGCAATAAAAATTCAATGAGTGAAC-3′ complementary to a sequence of the 3′ UTR (33–63 bp downstream of the translation termination signal) of the human CGT gene were used for PCR of human/rodent somatic cell hybrid DNA with individual chromosomes and of human, mouse, and Chinese hamster total DNA (1 μg DNA template) as controls (NIGHS human/rodent somatic cell hybrid panel 2). Standard Mg^{2+}, dNTP, primer concentrations, and the following conditions were applied: denaturing step at 94°C for 3 min; 25 cycles each at 93°C for 1 min, 60°C for 1 min, and 72°C for 1.5 min; and a final amplification step at 72°C for 7 min. A Perkin–Elmer 9600 PCR machine was used.

Fluorescence in situ hybridization. Human metaphase cells were prepared from phytohemagglutinin-stimulated peripheral blood lymphocytes. The hCGT probe was the human genomic clone λHCGT1. Fluorescence *in situ* hybridization was performed as described previously (Rowley *et al.*, 1990). Biotin-labeled probes were prepared by nick-translation using biotin-11–dUTP (Enzo Diagnostics). Hybridization was detected with fluorescein-conjugated avidin (Vector Laboratories), and chromosomes were identified by staining with 4,6-diamidino-2-phenylindole-dihydrochloride (DAPI).

Computational sequence analysis. Database searching and protein comparisons were performed using the software package of the Wisconsin Genetics Group (Devereux *et al.*, 1984).

RESULTS AND DISCUSSION

Cloning of the Human CGT Gene

A 1190-bp *Eco*RI–*Nco*I fragment of the full-length rat CGT clone has been used to screen a human placenta genomic library for the 5′ end of the human CGT gene and a mouse genomic fragment containing exon V to isolate the 3′ end of the human gene locus. Three nonoverlapping genomic clones that encode the complete cDNA sequence were isolated, and their restriction enzyme map was established. Five exons were localized, and their lengths and the lengths of introns II and III were determined. A gap in introns I and IV could not be closed by overlapping genomic clones (Fig. 1). The complete CGT gene locus extends over >40 kb. DNA sequencing established the exon sequences and the flanking intron nucleotide sequences (Fig. 2). The

ExonI:

gttggttccaagtctttgctattgtgaatagtgccgcaataaacatacgtgtgcatgtgtctttatagcagcatgatttatagtcctttg
ggtatatacccagtaatgggatggctgggttctagatccctgaggaatcgccacattgacttgcacaatggttgaactagtttacagtcc
caccaacagtgtaaaagtgttcctatttctccacatcctctccggcacctgttgtttcctgacttttttaatgattgccattctaactggt
gtgagatgatatctcattgtggttttgatttgcatttctctgatggccagtgatggtgagcatttttttcatgtgttttttggctgcataa
atgtcttcttttgagaagtgtctgttcatgatttttttttttagacaaagtattagatcagatatttttggggaaaaatgctttgtgatt
gcttgttttgaatggtgagcattgtatttttgttttaagttgttttctggttgttattacagctATGAAGTCTTACACTCCATATTTCATT

MetLysSerTyrThrProTyrPheIle

CTCCTGTGGAGTGCTGTTGGGATAGCGAAGGCTGCCAAAATCATCATCGTGCCGCCAATTATGTTTGAAAGCCATATGTACATTTTCAAG
LeuLeuTrpSerAlaValGlyIleAlaLysAlaAlaLysIleIleIleValProProIleMetPheGluSerHisMetTyrIlePheLys
ACGCTAGCCTCAGCCTTGCACGAGAGAGGCCACCATACAGTGTTCCTCCTCTCTGAAGGCAGAGACATCGCCCCATCTAATCATTACAGC
ThrLeuAlaSerAlaLeuHisGluArgGlyHisHisThrValPheLeuLeuSerGluGlyArgAspIleAlaProSerAsnHisTyrSer
CTCCAGCGCTACCCAGGGATCTTTAACAGTACCACCTCAGATGCTTTCCTACAGTCCAAGATGCGGAATATTTTCTCTGGGAGATTGACA
LeuGlnArgTyrProGlyIlePheAsnSerThrThrSerAspAlaPheLeuGlnSerLysMetArgAsnIlePheSerGlyArgLeuThr
GCAATCGAACTGTTTGACATACTGGATCACTATACTAAGAACTGTGACCTGATGGTTGGCAACCATGCCCTGATCCAGGGTCTGAAGAAA
AlaIleGluLeuPheAspIleLeuAspHisTyrThrLysAsnCysAspLeuMetValGlyAsnHisAlaLeuIleGlnGlyLeuLysLys
GAAAAATTTGACCTGCTGCTGGTGGACCCTAATGATATGTGTGGATTTGTGATAGCTCATCTTTTAGGGGTTAAATATGCTGTATTTTCA
GluLysPheAspLeuLeuLeuValAspProAsnAspMetCysGlyPheValIleAlaHisLeuLeuGlyValLysTyrAlaValPheSer
ACTGGCCTTTGGTATCCTGCTGAAGTGGGTGCTCCTGCTCCATTAGCATACGTCCCAGAGTTTAACTCACTCCTCACAGACCGCATGAAC
ThrGlyLeuTrpTyrProAlaGluValGlyAlaProAlaProLeuAlaTyrValProGluPheAsnSerLeuLeuThrAspArgMetAsn
TTGCTGCAAAGGATGAAAAATACCGGTGTTTACCTCATTTCCAGATTAGGGGTCAGCTTTCTGGTTCTTCCCAAATATGAAAGGATAATG
LeuLeuGlnArgMetLysAsnThrGlyValTyrLeuIleSerArgLeuGlyValSerPheLeuValLeuProLysTyrGluArgIleMet
CAGAAGTACAACCTGCTGCCGGAGAAGTCCATGTATGATTTGGTTCATGGGTCCAGCCTGTGGATGCTGTGTACTGACGTAGCACTGGAA
GlnLysTyrAsnLeuLeuProGluLysSerMetTyrAspLeuValHisGlySerSerLeuTrpMetLeuCysThrAspValAlaLeuGlu
TTCCCAAGACCCACTCTGCCTAATGTTGTTTATGTAGGAGGAATCCTAACCAAACCAGCCAGCCCACTACCAGAAgtaaggtttgcagaa
PheProArgProThrLeuProAsnValValTyrValGlyGlyIleLeuThrLysProAlaSerProLeuProGlu

.

≥ 29 kb (IntronI)

ExonII:

.

ctgccatgttttgatttacagGATCTCCAAAGATGGGTAAATGGTGCTAATGAACATGGCTTTGTCTTGGTGTCTTTTGGAGCTGGTGTC
AspLeuGlnArgTrpValAsnGlyAlaAsnGluHisGlyPheValLeuValSerPheGlyAlaGlyVal
AAGTATCTGTCAGAAGACATTGCTAACAAACTGGCAGGAGCTCTGGGGAGATTGCCTCAAAAAGTGATTTGGAGgtaaggtaataatgta
LysTyrLeuSerGluAspIleAlaAsnLysLeuAlaGlyAlaLeuGlyArgLeuProGlnLysValIleTrpAr(g)

.

≅ 1,5 kb (IntronII)

ExonIII:

.

ttgtctggtgtacattttagGTTTTCTGGACCCAAACCAAAGAATCTAGGAAACAACACTAAACTCATAGAATGGTTACCACAAAATGAC
(Ar)gPheSerGlyProLysProLysAsnLeuGlyAsnAsnThrLysLeuIleGluTrpLeuProGlnAsnAsp
CTGCTTGgtaagtcaatgatgtgtggttacttacttggctgttaggttt
LeuLeuG(ly)

.

≅ 2,5 kb (IntronIII)

ExonIV:

.

tgttttctttttcatctttcttcagGGCATTCAAAGATTAAAGCCTTCCTGAGCCATGGTGGTTTGAACAGTATTTTTGAAACTATGTAT
(G)lyHisSerLysIleLysAlaPheLeuSerHisGlyGlyLeuAsnSerIlePheGluThrMetTyr
CATGGTGTGCCTGTAGTGGGAATTCCACTCTTTGGAGACCATTATGATACTATGACCAGAGTACAGGCAAAAGGCATGGGGATATTGCTA
HisGlyValProValValGlyIleProLeuPheGlyAspHisTyrAspThrMetThrArgValGlnAlaLysGlyMetGlyIleLeuLeu
GAATGGAAGACAGTTACTGAAAAAGAGCTCTATGAAGCACTAGTGAAGGTTATCAATAATCCCAGgtaaggtttcaattaacattaaagc
GluTrpLysThrValThrGluLysGluLeuTyrGluAlaLeuValLysValIleAsnAsnProSe(r)

.

≥ 6 kb (Intron IV)

ExonV:

.

tcatcattctgtgttttgtcccctctccatagCTACCGTCAGAGGGCTCAGAAGCTTTCGGAAATTCACAAGGATCAACCTGGTCACCCT
(Se)rTyrArgGlnArgAlaGlnLysLeuSerGluIleHisLysAspGlnProGlyHisPro
GTCAATCGAACTATCTATTGGATAGATTATATTATTCGTCACAATGGAGCCCATCACCTACGTGCCGCTGTCCATCAGATCTCCTTTTGT
ValAsnArgThrIleTyrTrpIleAspTyrIleIleArgHisAsnGlyAlaHisHisLeuArgAlaAlaValHisGlnIleSerPheCys
CAGTATTTTTTACTGGATATTGCCTTTGTGCTTTTGCTTGGTGCTGCCTTGTTATACTTTCTCTTGTCTTGGGTGACAAAATTTATCTAC
GlnTyrPheLeuLeuAspIleAlaPheValLeuLeuLeuGlyAlaAlaLeuLeuTyrPheLeuLeuSerTrpValThrLysPheIleTyr
AGAAAAATCAAAAGTCTGTGGTCTAGAAATAAGCATAGCACAGTTAATGGACATTACCACAATGGAATCCTCAATGGCAAGTACAAAAGA
ArgLysIleLysSerLeuTrpSerArgAsnLysHisSerThrValAsnGlyHisTyrHisAsnGlyIleLeuAsnGlyLysTyrLysArg
AATGGCCATATTAAACATGAAAAGAAAGTGAAAtgagccaacagcccaggtgatagaaataaattggttcactcattgaatttttattgc
AsnGlyHisIleLysHisGluLysLysValLys
tattatttagtctaacagctactaaaagtaaaacatcagtaaacaattctaacatgcccttatgagactactaatgaaattctgtggaat
taagatggctgtaaaaagcacaaacctaaaatgcagaaatgtattttattcaaatactgatgtagagagttttggcactgaacctttag
aagccttaattatttaaatcaattcagtgactgtgtcagaccttagttttaaatcttgatatgtgcgtgtccc

FIG. 2. Nucleotide and deduced amino acid sequence of the human CGT gene. The nucleotide sequences of the five exons, the corresponding amino acid sequences, and parts of the intron sequences are shown (lowercase letters). Additional intron sequences are deposited with GenBank under Accession No. U31353. The sizes of introns I (>29 kb) and IV (>6 kb) are estimated.

TABLE 1

Exon/Intron Sizes and Sequences at Splice Junctions of the Human CGT Gene

Exon	Position[a]	5′-donor site	3′-acceptor site	Intron	Size (kb)	Amino acids interrupted
1	1–822	CCAGAA gtaagg	ttacag GATCTC	1	⩾29	Glu^{274}/Asp^{275}
2	823–965	TTGGAG gtaagg	ttttag GTTTTC	2	~0.5	Arg^{322}
3	966–1042	TGCTTG gtaagt	cttcag GGCATT	3	~3.5	Gly^{348}
4	1043–1262	TCCCAG gtaagg	ccatag CTACCG	4	⩾6	Ser^{421}
5	1263–1623					
	Consensus splice signal	a NAG gt agt g	cc c n ag G tt t			

[a] Exon positions in coding sequence.

sizes of the exons vary considerably between 822 bp encoding 274/275 amino acid residues for exon I and 77 bp encoding 25 amino acid residues for exon III. Intron sizes vary between 1.5 and >29 kb for intron II and intron I. The size of the gap in introns II and IV could not be determined (Table 1).

Chromosomal Localization of the Human CGT Gene

We applied two techniques for chromosome assignment of the human CGT gene locus: primer-specific PCR screening of a monochromosomal somatic cell hybrid panel (NIGHS human/rodent somatic cell hybrid panel 2) and chromosomal FISH. PCR with the two specific primers listed under Materials and Methods and the DNA of single chromosomes of the monochromosomal somatic cell hybrid panel resulted in the expected 456-bp DNA fragment. The biotinylated clone λHCGT1 spanning the 5′ end, exon I, and part of the large intron I was hybridized to elongated metaphase chromosomes from phytohemagglutinin-stimulated peripheral blood lymphocytes.

Hybridization with this probe resulted in specific labeling only of chromosome 4 (Fig. 4). Specific labeling of 4q25–q27 was observed on four (5 cells), three (5 cells), two (12 cells), or 1 (3 cells) chromatid(s) of the chromosome 4 homologues in 25 cells examined. Of 62 signals observed (62 of 100 4q chromatids from 25 metaphase cells were labeled), 9 (15%) signals were located at 4q25, 51 (82%) signals were located at 4q26, and 1 (1.5%) signal was located at 4q27. A single signal was observed at 4q23 (1.5%). No signal was observed at any other chromosomal site. We observed a specific signal at 4q26 in a second hybridization experiment using this probe. These results suggest that the human CGT locus is on chromosome 4, band q26.

Structural and Functional Relationship of CGT and the UGT Families

We have previously noticed a considerable sequence similarity between our CGT and members of the UGT family (UDP glucuronosyltransferases), which form a group of isozymes of 50 to 60 kDa located in the hepatic endoplasmic reticulum (Schulte and Stoffel, 1993). These isozymes detoxify a wide range of endogenous and xenobiotic toxic compounds by glucuronyl transfer to form a water-soluble product for excretion.

Two mammalian gene families of UDP-glucuronosyltransferases have been identified, UGT1 and UGT2 (Burchell *et al.*, 1991). Phenolic compounds and bilirubin are the substrates of the UGT1 isozymes. The gene of UGT1, also called dioxin-inducible gene, expresses a family of isozymes, UGT1A–UGT1F. It includes 10 gene products so far with 38.5 to 47.7% similarity on the nucleic acid level and 61.9 to 79.4% similarity on the deduced protein sequence level. The N-terminal domains of the gene products have little identity and similarity but their C-terminal domains are identical. Differential splicing, by which the 5′ regions are apparently spliced to shared 3′ portions of the UGT1 gene, leads to the isoforms with different substrate specificities (Ritter *et al.*, 1992).

The biochemical interpretation of the nucleic acid sequence analysis of the spliced products came from expression studies in cell cultures with chimeric cDNAs of UDP-glucuronosyltransferases for the definition of regions bearing specific acceptor and substrate sites and the binding site for the activated sugar (Mackenzie, 1990).

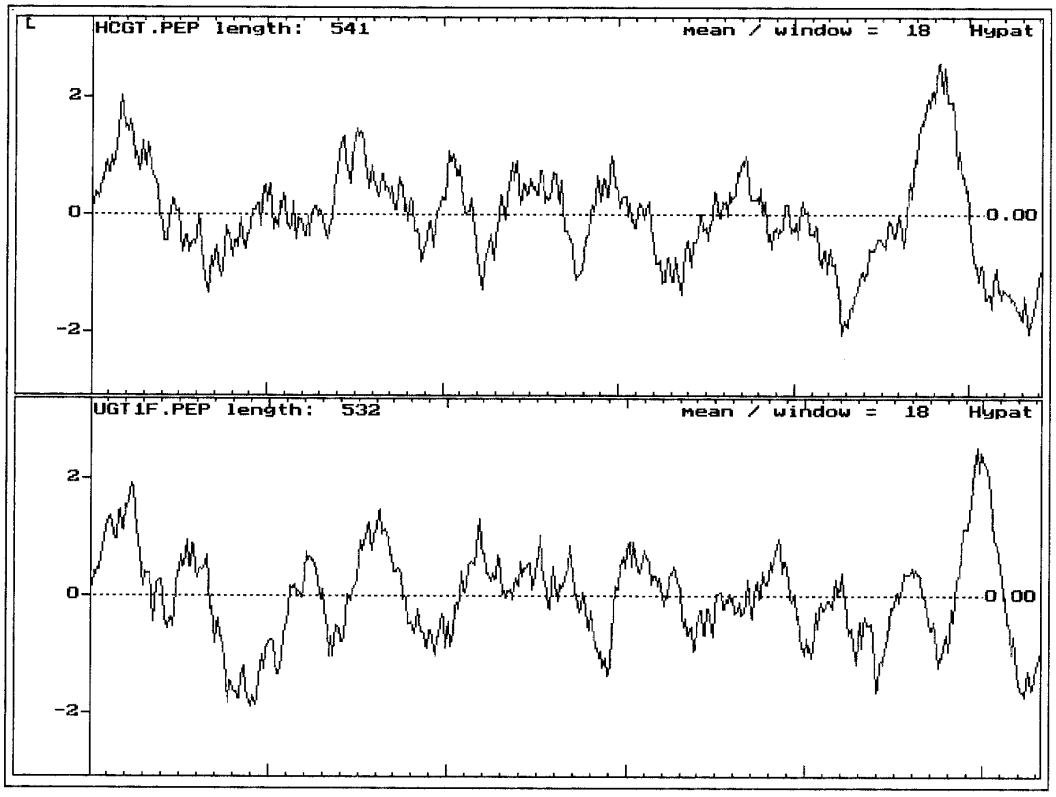
Two closely related isozymes of the UGT2 family in the rat, UDPG-Tr-3, which specifically glucuronidates testosterone, and UDPG-Tr-4, which transfers the glu-

FIG. 3. Amino acid sequence alignment of CGT and UGT1F, a member of the uridyl-glucuronide transferase family. (a) The two enzymes human CGT and UGT1F show a total similarity of 35 and 28% in exon I and 40% in exons II to V. Identical residues, black boxes; similar amino acid residues, gray boxes. Notice the identical positions of exon borders of exons I and II, exons III and IV, and exons IV and V. The membrane spanning sequence at the C-terminus is indicated by a hatched bar. Three black inverted triangles mark putative N-glycosylation sites. (b) The comparison of the prediction of the secondary structures of hCGT and UGT1F suggests their closely related overall secondary structure.

a

```
HCGT      1 ....MKSY.....TPYFTLLWSAVGIAKAAKIIVPPIMFESHMYIFKTLASALHERGHH
UGT1F     1 MACLLRSFQRISAGVFFLALWGMVV...GDKLLVVPQD..GSHWLSMKDTVEVLSDRGHE

HCGT     52 TVFLLSEGR.DIAPSNHYSLQRYPGIFNSTTSDAFLQSKMRNIFSGR..LTAIELL.....
UGT1F    56 IVVVVPEVNLLLKESKYYTRKIYPVPYDQEELKNRYQSFGNNHFAERSFLTAPQTEYRNN

HCGT    104 FDILDHYTKNCDLMVGNHALIQGLKKEKFDLLLVDPNDMCGFVIAHLLGVKYAVFSTGLW
UGT1F   116 MIVIGLYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILAEYLGLPSVYLFRGFP

HCGT    164 YPAEVG...APAPLAYVPEFNSLLTDRMNLLQRMKNTGVYLISRLGVSELVLPKYERIMQ
UGT1F   176 CSLEHTFSRSPDPVSYIPRCYTKFSDHMTFSQRVANFLVNLLEPY.LFYCLFSKYEE.LA

                                                    Exon I  Exon II
HCGT    221 KYNLLPEKSMYDLVHGSSLWMLCTDVALEFPRPTLPNVVYVGGILTKPASPLPEDLQRWV
UGT1F   234 SAVLKRDVDIITLYQKVSVWLLRYDFVLEYPRPVMPNMVFIGGINCKKRKDLSQEFEAYI

                                                Exon III
HCGT    281 NGANEHGFVLVSFGAGVKYLSEDIANKLAGALGRLPQKVIWRFSGPKPKNLGNNTKLIEW
UGT1F   294 NASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKW

            Exon IV
HCGT    341 LPQNDLLGHSKIKAFLSHGGLNSIFETMYHGVPVVGIPLFGDHYDTMTRVQAKGMGILLE
UGT1F   354 LPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTLN

                    Exon V
HCGT    401 WKTVTEKELYEALVKVINNPSYRQRAQKLSEIHKDQPGHPVNRTIYWIDYIIRHNGAHHL
UGT1F   414 VLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHL

HCGT    461 RAAVHQISFCQYFLLDIAFVLLLGAALLYFLLSWVTKFIYRKIKSLWSRNKHSTVNGHYH
UGT1F   474 RPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLG..............

HCGT    521 NGILNGKYKRNGHIKHEKKVK..
UGT1F   519 .........KKGRVKKAHKSKTH
```

b



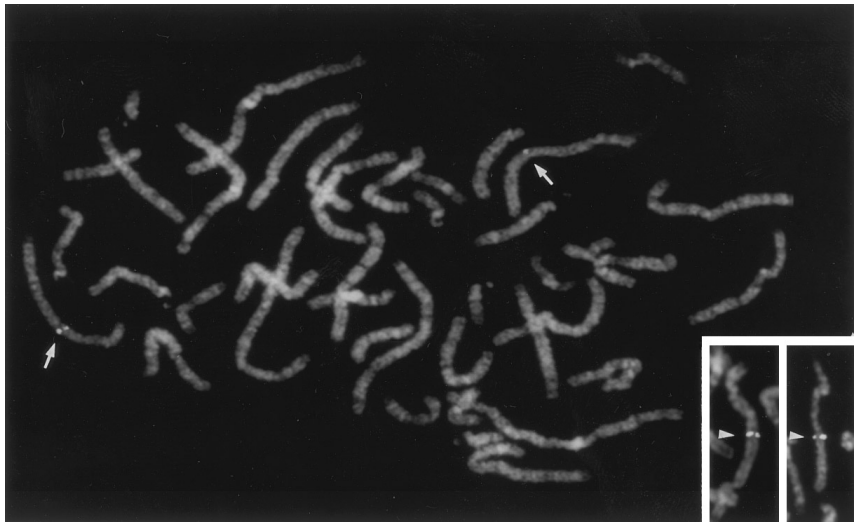


FIG. 4. Localization of human CGT to human chromosome 4q26. *In situ* hybridization of a biotin-labeld hCGT probe to human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes. The chromosome homologues are identified with arrows; specific labeling was observed at 4q26. The inset shows partial karyotypes of two chromosome 4 homologues illustrating specific labeling at 4q26 (arrowheads). Images were obtained using a Zeiss Axiophot microscope coupled to a cooled charge coupled device. Separate images of DAPI-stained chromosomes and the hybridization signal were merged using image analysis software (NU200 and Image 1.52b).

curonate to etiocholanone, have 85% similar amino acid residues in the N-terminal part, whereas they have almost identical C-terminal domains. The crosswise fusion of the 5′-terminal domains with the 3′ domains clearly demonstrates that the amino-terminal domain in the expression product is the aglycone substrate-specific part of the chimeric glucuronosyltransferases, whereas the C-terminal domain is responsible for the binding of the activated sugar moiety (Mackenzie, 1990).

The isoforms UGT1A to UGT1F are all encoded in one gene and arise from alternative splicing. The gene locus of the UGT1 family is on chromosome 2, and the loci of the subfamilies of the UGT2 genes are clustered in a region of approximately 215 kb on 4q13 (Monaghan *et al.*, 1994). No detailed assignment of the UGT1 gene to chromosome 2 is available, but the organization of the UGT1F gene has been reported (Ritter *et al.*, 1992). The cDNAs in this family show a similarity score of 57.2 to 92.5% (Burchell *et al.*, 1991).

The alignment of the amino acid sequences of human CGT and UGT1F by fusing the nucleotide-derived amino acid sequences of the individual five exons reveals not only the significant identities and similarities of CGT and UGT1F but also the identical lengths of the corresponding exons (Fig. 3a; their borders are marked by vertical bars). The lengths of the amino acid sequences encoded by the individual exons of the two genes are identical except the shift by four amino acid residues encoded by exon II of UGT1F (Fig. 3a). We also compared the hydrophobicity plots of CGT and UGT1F (Fig. 3b). A hydrophobic sequence of 20 amino acids in the C-terminal domain of CGT and the isozymes of the UGT subfamilies indicate the transmembrane helical domain. The N-terminal regions of CGT and UGT1F displayed the least homology and similarity. It is this region of the transferases that binds the different aglycone substrates (Mackenzie, 1990). Strongest identities and similarities are displayed in the C-terminal domain of the two transferases, which has been shown to harbor the binding site of the activated sugar, UDP-glucuronic acid, and likewise UDP-galactose.

We have shown that CGT is a mannose-rich glycoprotein integrated into the ER membrane. It is reasonable to assume that the enzymatically active part of CGT is oriented toward the lumen of the ER, leaving the sequence KKVK, characteristic of all ER-resident proteins, on the cytosolic side of the ER membrane. This canonical sequence is also common to the glucuronosyl transferases of the UGT1 and UGT2 families.

The hitherto known $\alpha(1-3)$ (Larsen *et al.*, 1989) and $\beta(1-4)$ (Shaper *et al.*, 1988) galactosyltransferases, on the other hand, have a membrane topology different from that of CGT. The N-terminus of the different galactosyltransferases contains the transmembrane domain (Joziasse *et al.*, 1989; Larsen *et al.*, 1989). Their protein predicts a type II transmembrane topology, with the large COOH terminal domain within the Golgi vesicular lumen.

The N-terminal halves of the deduced amino acid sequences are encoded in the large exon I and show considerably less similarity and identity than the C-terminal halves encoded in exons II to V. In view of the results obtained by fusion experiments of N- and C-terminal domains of closely related UGT1 members (Mackenzie, 1990), it is reasonable to assume that the exon I encoded amino acid sequence binds the ceramide moiety, analogous to the aglycone binding to this domain in the isozymes of the UGT subfamilies. Only

subtle changes of the rather conserved C-terminal domain appeared to be sufficient to confer the specificity of the binding site for the activated sugar donor molecule UDP-galactose.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft Sonderforschungsbereich 243, Molekulare Analyse der Entwicklung zellulärer Systeme (Project A4). One of the authors (M.M.L.) was supported by PHS Grant CA40046. We thank Dr. M. Stoffel, Laboratory of Metabolic Diseases, The Rockefeller University, New York, for carrying out the primer-specific PCR screening of a somatic cell hybrid panel.

REFERENCES

Bansal, R., and Pfeiffer, S. E. (1989). Reversible inhibition of oligodendrocyte progenitor differentiation by a monoclonal antibody against surface galactolipids. *Proc. Natl. Acad. Sci. USA* 86: 6181–6185.

Burchell, B., Nebert, D. W., Nelson, D. R., Bock, K. W., Iyanagi, T., Jansen, P. L. M., Lancet, D., Mulder, G. J., Chowdhury, J. R., Siest, G., Tephly, T. R., and Mackenzie, P. I. (1991). The UDP glucuronosyltransferase gene superfamily: Suggested nomenclature based on evolutionary divergence. *DNA Cell Biol.* 10: 487–494.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12: 387–395.

Harding, D., Heremiah, S. J., Povey, S., and Burchell, B. (1990). Chromosomal mapping of a human phenol UDP-glucuronosyltransferase, GNT1. *Ann. Hum. Genet.* 54: 17–21.

Joziasse, D. H., Shaper, J. H., Van den Eijnden, D. H., Van Tunen, A. J., and Shaper, N. L. (1989). Bovine α-1-3-galactosyltransferase: Isolation and characterization of a cDNA clone.

Larsen, R. D., Rajan, V. P., Ruff, M. M., Kukowska-Latallo, J., and Cummings, R. D. (1989). Isolation of a cDNA encoding a murine UDPgalactose:β-D-galactosyl-1,4-*N*-acetyl-D-glucosaminide α-1,3-galactosyltransferase: Expression cloning by gene transfer. *Proc. Natl. Acad. Sci. USA* 86: 8227–8231.

Lemke, G. (1993). The molecular genetics of myelination: An update. *Glia* 7: 263–271.

Mackenzie, P. I. (1990). Expression of chimeric cDNAs in cell culture defines a region of UDP glucuronosyltransferase involved in substrate selection. *J. Biol. Chem.* 265: 3432–3435.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Monaghan, G., Clarke, D. J., Povey, S., See, C. G., Boxer, M., and Burchell, B. (1994). Isolation of a human YAC contig encompassing a cluster of *UGT2* genes and its regional localization to chromosome 4q13. *Genomics* 23: 496–499.

Morell, P., Quarles, R. H., and Norton, W. T. (1994). "Myelin Formation, Structure, and Biochemistry" (G. J. Siegel, B. W. Agranoff, R. W. Albers, and P. B. Molinoff, Eds.), pp. 117–143, Raven Press, New York.

Morell, P., and Radin, N. S. (1969). Synthesis of cerebroside by brain from uridine diphosphate galactose and ceramide containing hydroxy fatty acid. *Biochemistry* 8: 506–512.

Ritter, J. K., Chen, F., Sheen, Y. Y., Tran, H. M., Kimura, S., Yeatman, M. T., and Owens, I. S. (1992). A novel complex locus *UGT1* encodes human bilirubin, phenol, and other UDP-glucuronosyltransferase isozymes with identical carboxyl termini. *J. Biol. Chem.* 267: 3257–3261.

Rowley, J. D., Diaz, M. O., Espinosa, R., Patel, Y. D., van Melle, E., Ziemin, S., Taillon-Miller, P., Lichter, P., Evans, G. A., Kersey, J. D., Ward, D. C., Domer, P. H., and Le Beau, M. M. (1990). Mapping chromosome band 11q23 in human acute leukemia with biotinylated probes: Identification of 11q23 translocation breakpoints with a yeast artificial chromosome. *Proc. Natl. Acad. Sci. USA* 819: 9358–9362.

Sanger, F., Nicklen, S., and Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463–5467.

Sato, H., Koiwai, O., Tanabe, K., and Kashiwamata, S. (1990). Isolation and sequencing of rat liver bilirubin UDP-glucuronosyltransferase CDNA: Possible alternate splicing of a common primary transcript. *Biochem. Biophys. Res.Commun.* 169: 260–264.

Schulte, S., and Stoffel, W. (1993). Ceramide UDPgalactosyltransferase from myelinating rat brain: Purification, cloning, and expression. *Proc. Natl. Acad. Sci. USA* 90: 10265–10269.

Shaper, N. L., Hollis, G. F., Douglas, J. G., Kirsch, I. R., and Shaper, J. H. (1988). Characterization of the full length cDNA for murine β-1,4-galactosyltransferase. Novel features at the 5′-end predict two translational start sites at two in-frame AUGs. *J. Biol. Chem.* 263: 10420–10428.

Stahl, N., Jurevics, H., Morell, P., Suzuki, K., and Popko, B. (1994). Isolation, characterization, and expression of cDNA clones that encode rat UDP-galactose:Ceramide galactosyltransferase. *J. Neurosci. Res.* 38: 234–242.