SHORT COMMUNICATION

Molecular Cloning and Characterization of the Mouse CGT Gene Encoding UDP-Galactose Ceramide-Galactosyltransferase (Cerebroside Synthetase)

ANDREAS BOSIO, ERIKA BINCZEK, AND WILHELM STOFFEL[1]

Institut für Biochemie, Medizinische Fakultät, Universität Köln, Joseph-Stelzmann-Strasse 52, D-50931 Cologne, Germany

Received October 27, 1995; accepted April 1, 1996

UDP-galactose ceramide galactosyltransferase, CGT, EC 2.4.1.45, is the key enzyme in the biosynthesis of cerebrosides and sulfatides, which are the most abundant glycosphingolipids in the myelin of the central nervous system and the peripheral nervous system. The cell-specific and highly time-regulated expression of the CGT gene is thought to play an important role in oligodendrocyte and Schwann cell differentiation. Three genomic clones encoding the mouse CGT gene were isolated and characterized. The gene is distributed over >42 kb, and the coding sequence is distributed over five exons ranging from 77 to 822 bp. Putative transcription start sites were determined by primer extension experiments. The CGT gene locus is highly conserved during evolution.

Myelination allows fast conduction in axons and is therefore crucial for the development of higher vertebrates. Oligodendrocytes produce myelin of the central nervous system (CNS) and Schwann cells of the peripheral nervous system (PNS). They differ from others in the expression of one of their major proteins, the proteolipid protein (PLP), and protein zero (P_0), respectively. Myelin basic protein (MBP) and minor constituents, e.g., myelin-associated glycoprotein (MAG) and myelin oligodendrocyte glycoprotein (MOG), are, like PLP and P_0, expressed in a time-specific manner but without restriction to one type of myelinating cell (for review see (11)). Lipids comprise more than 70% of the dry weight of myelin, with galactocerebrosides (GalC) being an abundant class of lipids. The occurrence of this class is restricted to CNS and PNS myelin except for kidney, which has very low amounts of it. GalC appears as an early specific marker in oligodendrocyte differentiation (13). Anti-GalC antibodies are used for the staging of the oligodendrocyte lineage during differentiation (1, 12). Numerous studies have been carried out with anti-GalC antibodies to elucidate putative functional roles of GalC and sulfatides. These studies indicate a potential role for GalC in the differentiation of oligodendrocytes and Schwann cells in transducing signals that maintain the process of myelination (1), as a receptor for viral (8) and bacterial toxins (7), and as a mediator of Ca^{2+} signal in oligodendrocytes ((5), for review see (6)).

The key enzyme in the biosynthesis of galactocerebrosides is UDP-galactose ceramide galactosyltransferase (CGT; EC 2.4.1.45), which catalyzes the transfer of galactose from UDP-galactose to ceramide (10). The enzyme has been purified in this laboratory to homogeneity from myelinating rat brain, followed by the cloning of the full-length rat CGT cDNA and its characterization (15). The CGT cDNA has an open reading frame of 1623 bp encoding a 541-aa core protein. The 64-kDa high mannose glycoprotein resides in the endoplasmic reticulum. The primary translation product carries an N-terminal 20-aa-residue signal sequence, three putative N-glycosylation sites, a C-terminal 23-aa-residue transmembrane domain, and a KKVK endoplasmic reticulum retention signal.

Recently, we have cloned and characterized the human CGT gene and assigned its chromosomal localization as a single-copy gene to 4q26 (2). The coding sequence of CGT-specific cDNA is distributed over five exons with an approximate length of >40 kb.

A 690-bp *Eco*RI/*Nco*I fragment and a 1190-bp *Nco*I fragment of the rat CGT cDNA (15) were used for screening a mouse leukocyte genomic DNA library in λEMBL3 SP6/T7 (ML 1040j; Clontech). Three clones with insert sizes of 13 kb (λMCGT7), 19 kb (λMCGT9), and 21 kb (λMCGT14) were obtained. Restriction mapping and Southern blot hybridization with specific oligonucleotides derived from rat CGT cDNA revealed that λMCGT9 and λMCGT14 overlap and that the three clones encompass the entire mouse CGT gene except intron 1 (Fig. 1c). The coding sequence of the

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. X92122–X92126 and X92177.

[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.

GENOMICS 35, 223–226 (1996)
ARTICLE NO. 0342
0888-7543/96 $18.00

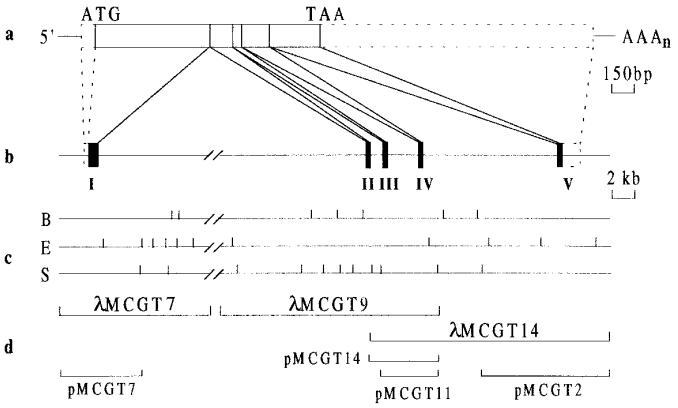

FIG. 1. Structure of the mouse CGT gene. (a) Mouse CGT mRNA. (b) Arrangement of the five exons (depicted by numbers and boxes). Solid boxes represent coding regions; 5′- and 3′-untranslated regions are represented by open boxes. (c) Restriction map using the enzymes *Bam*HI (B), *Eco*RI (E), *Sac*I (S), and *Xho*I (no *Xho*I restriction site exists in the analyzed clones). (d) Positions of the three genomic phage clones λMCGT7, λMCGT9, and λMCGT14 and the subclones pMCGT7, pMCGT14, pMCGT11, and pMCGT2, which were used for double-strand sequencing.

mouse CGT gene is distributed over five exons, the shortest comprising 77 bp (exon III) and the longest 822 bp (exon I). The four introns vary between 1.4 and >23 kb as depicted in Fig. 1b. The exact size of intron 1 is unknown because the size of the gap between λMCGT7 and λMCGT9 could not be determined by Southern blot analysis, and no phage clone harboring this region was obtained. All exons, including at least 30 bp of the flanking intron sequences and 425 bp upstream of the initiation codon, have been sequenced using the dideoxy chain-termination method (14) utilizing synthetic oligonucleotides derived from rat CGT cDNA or established sequences. These sequences as well as an assembly of exon sequences have been deposited with the EMBL nucleotide sequence database (Accession Nos. X92122, X92123, X92124, X92125, X92126, X92177). The sizes of each exon and intron, as well as the nucleotide sequences at the exon–intron junctions, are summarized in Table 1. All intron sequences are in agreement with the consensus sequences at intron–exon junctions.

The deduced protein sequence from the mouse CGT gene exhibits a high degree of homology to those from rat CGT (98.7%) and human CGT (97.6%). The coding region of the CGT gene nucleotide sequence shows 95.9% identity between mouse and rat and about 90% identity between human and mouse or human and rat. Structural features like putative N-glycosylation sites, the C-terminal hydrophobic part of the protein, the N-terminal signal peptide, and the KKVK endoplasmic reticulum retention signal are well conserved among the three species. Furthermore, comparison of the human and mouse CGT gene structures reveals a preserved five-exon composition of the coding sequence with identical positions of the introns and with only minor alterations in the size of the known intervening sequences. It is interesting to note that the major structural protein of CNS myelin, PLP, is one of the most highly conserved proteins with respect to genomic structure and amino acid and nucleotide sequences. Its amino acid sequence is identical among mouse, rat, and human with 97% identity of the coding sequence (4), which is even higher than that of the CGT gene. The lack of evolutionary divergence of the major structural protein and of the enzyme catalyzing the synthesis of the most abundant myelin class of lipids may emphasize the crucial importance of the myelin sheet and its major components for the development of vertebrates.

Primer extension experiments were carried out using mouse brain poly(A)$^+$ RNA prepared from brain of 20-day-old mice (3) and the 5′-end-labeled antisense oligonucleotide CP40r (5′-GAGCATGAAATATGGAGTGT-AAGAGTTCAT-3′) derived from the 10 N-terminal amino acid residues of the signal sequence, following the standard procedure (9). Extension products were analyzed on a 7 *M* urea 5% polyacrylamide gel adjacent to a sequence ladder generated with the same 30-nt oligonucleotide primer. As shown in Fig. 3, a major extension band is located at position −144 upstream of the translation start codon. This major band is ac-

TABLE 1

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

Exon	Position[a]	Exon/intron junction sequences: 5′-Donor site	Exon/intron junction sequences: 3′-Acceptor site	Intron	Size (kb)	Amino acids interrupted
I	1–822	CCAGAA gtaagg	ttacag GATCTC	1	⩾23	Glu^{274}/Asp^{275}
II	823–965	TTGGAG gtaagg	ttttag GTTTTC	2	≅1.4	Arg^{322}
III	966–1042	TGCTTG gtaagt	cttcag GGCATT	3	≅3.2	Gly^{348}
IV	1043–1262	TCCCAG gtaagg	ccatag CTACCG	4	≅12	Ser^{421}
V	1263–1623					
Consensus splice signal		NAG gt$\frac{a}{g}$agt	$(\frac{c}{t})_7$n$\frac{c}{t}$ag G			

[a] Exon positions in coding sequence.

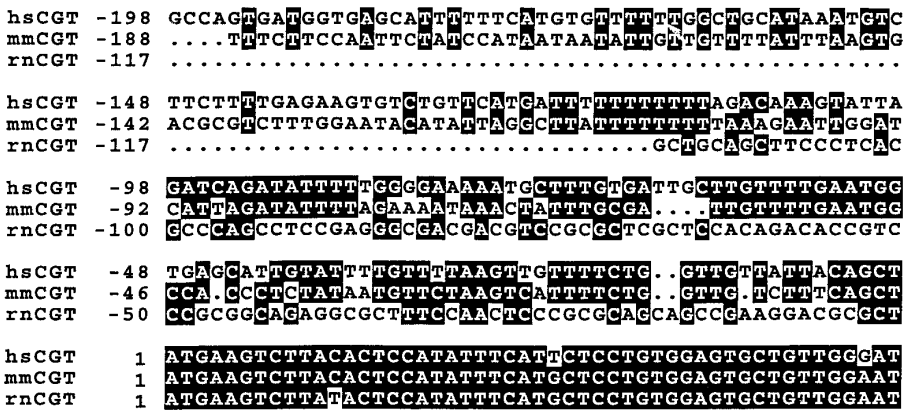

FIG. 2. Alignment of the 5′-untranslated genomic sequence and part of exon I of mouse and human CGT DNA and of rat cDNA. Sequence identities are indicated by using white on black lettering.

companied by a weaker signal at position −239 and a cluster of three bands at positions −54, −55, and −56.

As depicted in Fig. 2, the 5′-untranslated genomic sequences of mouse and human CGT show a notable homology in contrast to the aligned rat CGT cDNA sequence. This might indicate that the 5′-untranslated sequence of rat cDNA is encoded in an additional exon 0. Also, the nucleotides AG in positions −3 and −4 relative to ATG of the mouse sequence fit the consensus sequence of a splice acceptor site as well as the nucleotide C in position −5, which is conserved up to 80% at this position of the intron–exon junctions.

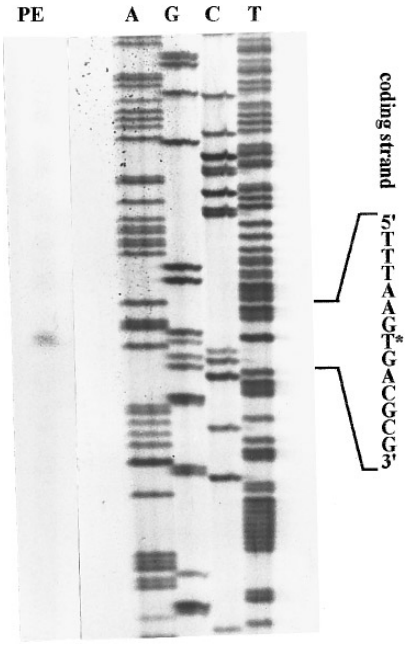

FIG. 3. Primer extension analysis for the determination of the transcription start site(s). The exact lengths of the primer extension products were determined by aligning the corresponding sequence of mouse CGT DNA generated with the same 30-nt oligonucleotide primer on the same polyacrylamide gel. Photographs were taken of autoradiograms after 3 days exposure (PE) and 1 day exposure (sequence ladder).

The mouse CGT gene structure reported here and the available genomic clones provide the molecular basis for the analysis of its regulated expression and for studies of its functional role in oligodendrocytes and Schwann cell maturation by means of transgenic and gene disruption experiments.

REFERENCES

1. 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.
2. Bosio, A., Binzcek, E., Le Beau, M. M., Fernald, A. A., and Stoffel, W. (1995). The human gene CGT encoding the UDP-galactose ceramide galactosyl transferase (cerebroside synthase): Cloning, characterization and chromosomal assignment to human 4q26. *Genomics* 28: 607–609.
3. Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* 162: 156–159.
4. Diehl, H. J., Schaich, M., Budzinski, R. M., and Stoffel, W. (1986). Individual exons encode the integral membrane domains of human myelin proteolipid protein [published erratum appears in *Proc. Natl. Acad. Sci. USA* 1991, 86(6): 617–618]. *Proc. Natl. Acad. Sci. USA* 83: 9807–9811.
5. Dyer, C. A., and Benjamins, J. A. (1991). Galactocerebroside and sulfatide independently mediate Ca^{2+} responses in oligodendrocytes. *J. Neurosci. Res.* 30: 699–711.
6. Dyer, C. A. (1993). Novel oligodendrocyte transmembrane signaling systems. Investigations utilizing antibodies as ligands. *Mol. Neurobiol.* 7: 1–22.
7. Hakomori, S. (1990). Bifunctional role of glycosphingolipids. *J. Biol. Chem.* 265: 18713–18716.
8. Harouse, J. M., Bhat, S., Spitalnik, S. L., Laughlin, M., Stefano, K., Silberberg, D. H., and Gonzales-Scarano, F. (1991). Inhibition of entry of HIV-1 in neural cell lines by antibodies against galactosyl ceramide. *Science* 253: 320–323.
9. McNight, S. L., and Kingsbury, R. (1982). Transcription control signals of a eukaryotic protein-coding gene. *Science* 217: 316–324.
10. 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.

11. Morell, P., Quarles, R. H., and Norton, W. T. (1994). *In* "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.
12. Raff, M. C., Mirsky, R., Fields, K. L., Lisak, R. P., Dorfman, S. H., Silberberg, D. H., Gregson, D. H., Leibowitz, S., and Kennedy, M. C. (1978). Galactocerebroside is a specific cell-surface antigenic marker for oligodendrocytes in culture. *Nature* 274: 813–816.
13. Raff, M. C., Fields, K. L., Hakomori, S., Mirsky, R., Pruss, R. M., and Winter, J. (1979). Cell-type-specific markers for distinguishing and studying neurons and the major classes of glial cells in culture. *Brain Res.* 174: 283–308.
14. Sanger, F., Nicklen, S., and Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463–5467.
15. Schulte, S., and Stoffel, W. (1993). Ceramide UDPgalactosyltransferase from myelinating rat brain: Purification, cloning, and expression. *Proc. Natl. Acad. Sci. USA* 90: 10265–10269.