Evidence for two cDNA clones encoding human GM2-activator protein

Shanmugam NAGARAJAN,* Hao-Chia CHEN,† Su-Chen LI,*§ Yu-Teh LI* and Jean M. LOCKYER*†

*Department of Biochemistry and the †Human Genetics Program, Tulane University School of Medicine, New Orleans, LA 70112, U.S.A. and ‡Endocrinology and Reproduction Research Branch, NICHD, NIH Bethesda, MD 20892, U.S.A.

Two cDNAs encoding GM2 activator, pGM2A (648 bp) and GAP (1093 bp), were isolated from human placenta λ gt11 libraries. The DNA sequence of pGM2A from 1 to 302 was almost identical with GAP, but diverged from 303–648. PCR was used to demonstrate the presence of both species of GM2 activator in placental RNA. Both cDNAs hybridized to mRNAs of approximately 2.3 kb and to identical single bands on genomic Southern blots.

INTRODUCTION

Glycosphingolipids in mammalian tissues are catabolized by the sequential release of monosaccharide units by specific lysosomal acid glycohydrolases [1]. It has been well documented that the catabolism of some of the glycosphingolipids by specific glycohydrolases requires the presence of activator proteins [2,3]. There are four classes of these protein cofactors: the first activates the enzymes for the hydrolyses of GalCer-SO₄ [4], GM1 [5], GbOse₃Cer [6–8] and several other glycolipids [9]; the second stimulates the hydrolyses of GlcCer [10–13] and GalCer [14]; the third is the activator for the hydrolysis of sphingomyelin [15,16]; and the fourth is GM2 activator which activates β -hexosaminidase A to hydrolyse GM2 [17–21].

The mechanism of action of GM2 activator is not well understood, although it has been postulated that the activator extracts a single GM2 molecule from the micelles and presents it to β -hexosaminidase A [22,23]. The physiological importance of GM2 activator has been demonstrated by the fact that the biochemical lesion of Type AB GM2 gangliosidosis is caused by a deficiency of this activator protein [18,24,25]. In order to perform expression studies to examine the mechanism of action of GM2 activator, we isolated a cDNA clone encoding GM2 activator. During the course of this work, a partial sequence of a GM2 activator cDNA (pGAP1) isolated from a human fibroblast library was reported [26]. Interestingly, the pGM2A cDNA we isolated from a human placenta library differed in some aspects from the fibroblast pGAP1. In this report, we present evidence for the presence of multiple GM2 activator mRNAs.

MATERIALS AND METHODS

Materials

Restriction endonucleases, T4 polynucleotide kinase and T4 DNA ligase were purchased from Promega (Madison, WI, U.S.A.) and BRL (Grand Island, NY, U.S.A.). Human placenta λ gt11 libraries were the generous gifts of Dr. B. Knoll (University of Texas Medical Center, Houston, TX, U.S.A.) and Dr. J. Millan (La Jolla Cancer Research Foundation, La Jolla, CA, U.S.A.). A chicken β -actin probe was obtained from Dr. L. Levy (Tulane Medical School). GM2 activator was isolated from

human kidney according to the method described for the isolation of the activator from human liver [21]. Polyclonal antibodies against GM2 activator were raised in rabbits as reported previously [25] and purified by a protein A column.

Peptide sequence of GM2 activator

An equal volume of 6 M-guanidium chloride in 0.1 M-*N*-methylmorpholine/acetic acid buffer, pH 8.3, containing 1 mM-EDTA and 2.3 mg of DL-dithiothreitol were added to 0.2 ml (25 μ g) of the activator solution. After a 2 h incubation at 37 °C under N₂, 10 μ l of 4-vinylpyridine was added. The reaction proceeded for 30 min and was then quenched by the addition of 10 μ mol of cysteine in 50 μ l of the same buffer solution. The reduced and *S*-pyridinylated protein was isolated by Bio-Rad P-10 gel chromatography using 50 % formic acid as the eluent. The modified protein was then freeze-dried, resuspended in 50 μ l of 0.1 M-ammonium bicarbonate and treated with 0.25 μ g (in 2.4 μ l of water) of TPCK-treated trypsin twice at 37 °C at 2 h intervals. The tryptic fragments were then separated using a Hewlett-Packard HP-1090 h.p.l.c. system: the digest was freeze-dried, redissolved in 1 % trifluoroacetic acid, applied to a Vydac protein C4 column (214TP54, particle size 5 μ m, 0.46 cm $\times$ 25 cm) and eluted with an acetonitrile gradient (solvent A = 0.1 % trifluoroacetic acid in water, solvent B = 0.1 % trifluoroacetic acid in 70 % acetonitrile) at a flow rate of 0.5 ml/min. The percentage of solvent B was programmed as follows: 5 min, 13 %; 50 min, 55 %; 65 and 70 min, 100 %; 75 min, 13 %. The tryptic peptides were sequenced on a gas-phase protein sequencer (Applied Biosystems model 470A) with an on-line phenylthiohydantoin analyser (model 120A) using a standard version 3.0 program.

Isolation of cDNA clones

A human placenta λ gt11 cDNA library (B. Knoll) was screened as described [27] using purified polyclonal anti-(GM2 activator) antibodies. Positive clones were detected using alkaline phosphatase-conjugated polyclonal goat anti-rabbit IgG as a second antibody, followed by colour development with 5-bromo-4-chloroindol-3-yl phosphate and Nitro Blue Tetrazolium [28]. Additional clones were obtained from a second human placenta λ gt11 library (J. Millan) using the pGM2A cDNA identified with the antibody (Fig. 1). The probe was labelled to a specific

Abbreviations used: TPCK, *N*-tosyl-L-phenylalanylchloromethane; 1 $\times$ SSC, 0.15 M-NaCl/0.015 M-sodium citrate.

§ To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBK Nucleotide Sequence Databases under the accession numbers X61094 and X61095.

radioactivity of 10^8 c.p.m./ μ g by random priming (BRL). Pre-hybridization and hybridization conditions were $6 \times \text{SSC}/0.1\%$ SDS/ $2 \times \text{Denhardt's}$ at 65°C . The filters were washed three times at room temperature for 15 min each and once for 20 min in $0.1 \times \text{SSC}/0.1\%$ SDS at 65°C .

DNA sequence analysis

EcoRI inserts from positive phage clones were subcloned into Bluescript. Dideoxy sequencing [29] was performed with KS, SK, T3, T7 (Stratagene, La Jolla, CA, U.S.A.) and specific oligonucleotide primers. The nucleotide sequences were aligned and analysed using The Genetics Computer Group software of the University of Wisconsin Biotechnology Center [30].

PCR amplification

PCR [31] using RNA as the starting template was performed as follows: the first strand was synthesized using the PCR primer which was complementary to the RNA. The incubation mixture contained 10 μ g of total RNA, 0.5 mM-dNTPs, 50 mM-Tris/HCl, pH 8.3, 10 mM-MgCl₂, 150 mM-KCl, 20 mM-2-mercaptoethanol, 50 μ g actinomycin D, 10 mM-dithiothreitol and 200 ng of primer. The primer was extended with 15 units of avian myeloblastosis virus reverse transcriptase (Molecular Genetics Resources, Tampa, FL, U.S.A.). The extended product was then used as the template for the amplification reactions. The reactions were performed in 100 μ l of a buffer containing 50 mM-Tris/HCl, pH 9.0, 50 mM-KCl, 15 mM-(NH₄)₂SO₄ and 7 mM-MgCl₂ with 0.5 mM-dNTPs, 5 units of Taq polymerase and 0.5 μ g of each primer. Amplification was performed for 40 cycles using a thermal cycler (NOLA Sciences Inc., Slidell, LA, U.S.A.). Each cycle consisted of a 1.5 min denaturation at 92°C , a 2 min annealing at 58°C , and a 3 min extension at 72°C .

Southern and Northern blot analysis

Southern blots were prepared with 15 μ g of genomic DNA isolated from human leucocytes. The DNA samples were digested with appropriate restriction endonucleases, fractionated on 0.8% agarose gels and blotted to Nitroplus 2000 (Micron Separations Inc., Westborough, MA, U.S.A.). For Northern blots, total cellular RNA was isolated from human placenta and human fibroblast primary cultures using guanidine thiocyanate [32]. Poly(A)⁺ RNAs were then selected with oligo[dT]-cellulose [33]. Each poly(A)⁺ RNA sample (5 μ g) was electrophoresed in 1.2% agarose/2.2 M-formaldehyde gels [34] and transferred to Nitroplus 2000. Both Northern and Southern blots were hybridized at 42°C in a solution containing 50% formamide, $4 \times \text{SSC}$, 0.1 M-sodium phosphate, pH 6.5, 0.075% sodium pyrophosphate, 0.1% SDS, $1 \times \text{Denhardt's}$, 75 μ g of herring sperm DNA/ml and 10% dextran sulphate. The prehybridization solution was prepared identically, but without dextran sulphate. Filters were washed twice in $2 \times \text{SSC}/0.5\%$ SDS at room temperature for 15 min each, once at 65°C for 20 min in the same solution and, lastly, in $0.2 \times \text{SSC}/0.1\%$ SDS at 65°C for 20 min. The GM2A-specific probe used for hybridization was a 266 bp *PstI/EcoRI* fragment from the 3' untranslated region of the clone (Fig. 1). The GAP probe was the entire pGAP2 cDNA (Fig. 2). Probes were labelled to a specific radioactivity of approximately 10^8 c.p.m./ μ g.

GM2 activator oligonucleotide probe

An oligonucleotide probe, 5'-CATGTCCAGCACATCACAGAAGGCCTCAAAGGTGCAGGAGCCAATGTAGTCTGTGCAAGG-3' (60 bases), complementary to the amino acid

sequence of tryptic fragment no. 2 (Table 1) from proline to methionine, was synthesized on the basis of codon usage frequencies [35].

RESULTS

Isolation of a human GM2 activator cDNA

Approximately 10^6 plaques of a placenta λ gt11 library were screened with the purified polyclonal anti-(GM2 activator) antibodies. The ten plaques which gave signals were purified and analysed by Southern blotting using a ³²P-labelled 60-base probe derived from amino acid sequence data. Only one clone hybridized with the 60-mer. It did not cross-hybridize with the other nine clones. The positive clone was designated as pGM2A (Fig. 1) and will be discussed in further detail below. Additional screening of the library using GM2A as a probe did not give any positive signals. A second placenta λ gt11 library was therefore screened with the same probe. Two clones, pGAP2 and pGAP3, of 894 and 1037 bp in size respectively, were identified from the second library.

The nucleotide sequences of pGAP2 and pGAP3 are overlapping: pGAP2 extends farther 5' and pGAP3 extends farther 3'. The sequence of the two clones taken together (GAP) yields a size of 1093 bp (Fig. 2). GAP nucleotides 64–884 correspond to the human fibroblast pGAP1 cDNA which has been previously reported [26]. The N-terminal fragment and tryptic fragments 1–7 (Table 1) of the peptide sequence obtained from tryptic peptides of GM2 activator were present in pGAP2 and pGAP3. The protein sequence and the amino acid sequence derived from the cDNA were not 100% identical. Two differences were noted: in the N-terminal peptide, the valine at position 25 is an isoleucine in the cDNA (amino acid 66); at position 14 of tryptic fragment 2, the alanine is a histidine in the cDNA (amino acid 117).

The N-terminus of GM2 activator was a phenylalanine (Table 1) according to the protein sequence analysis. The cDNA extends 5' beyond the phenylalanine, suggesting that GM2 activator exists as a precursor which is processed to its mature form. Methionine codons are present at positions 22 and 34 of the nucleotide sequence (Fig. 2). The surrounding sequence of neither of these methionine codons matches the translation initiation consensus sequence [36].

Four minor heterogeneities exist when comparing the placenta pGAP2 and pGAP3 sequences with the fibroblast pGAP1 sequence (Table 2). The nucleotide changes in three of the four positions (positions 76, 196, and 226) result in conservative alterations of the amino acid sequence and the change at the fourth position (nucleotide 603) gives a TAA stop codon rather than the TAG stop codon found in pGAP1. The two placental clones, pGAP2 and pGAP3, were identical in all of these positions.

Nucleotide sequence of pGM2A

The nucleotide sequence of pGM2A (Fig. 1) revealed a striking difference when compared with the sequence of GAP. Nucleotides 1–302 of pGM2A were almost identical with GAP. The sequences diverged dramatically, however, from position 303 to the 3' end of the clone. The pGM2A sequence has a stop codon at position 311, whereas the GAP open reading frame contains an additional 147 bp. The proteins encoded by the two mRNAs would therefore differ in size by 49 amino acids. The 3' untranslated regions of the two cDNAs did not show any sequence similarity.

PCR amplification of GM2A and GAP fragments

Since only one pGM2A clone was isolated from the λ gt11 libraries, it was possible that the divergent sequence of pGM2A

													K	D	P	A	V		
													1G	AAG	GAC	CCT	GCG	GTG	16
I	R	S	L	T	L	E	P	D	P	I	V	V	P	G	N	V	T		
17	ATC	AGA	AGC	CTG	ACT	CTG	GAG	CCT	GAC	CCC	ATC	GTC	GTT	CCT	GGA	AAT	GTG	ACC	70
L	S	V	V	G	S	T	S	V	P	L	S	S	P	L	K	V	D		
71	CTC	AGT	GTC	GTG	GGC	AGC	ACC	AGT	GTC	CCC	CTG	AGT	TCT	CCT	CTG	AAG	GTG	GAT	124
L	V	L	E	K	E	V	A	G	L	W	I	K	I	P	C	T	D		
125	TTA	GTT	TTG	GAG	AAG	GAG	GTG	GCT	GGC	CTC	TGG	ATC	AAG	ATC	CCA	TGC	ACA	GAC	178
Y	I	G	S	C	T	F	E	H	F	C	D	V	L	D	M	L	I		
179	TAC	ATT	GGC	AGC	TGT	ACC	TTT	<u>GAA</u>	<u>CAC</u>	<u>TTC</u>	<u>TGT</u>	<u>GAT</u>	<u>GTG</u>	<u>CTT</u>	GAC	ATG	TTA	ATT	232
P	T	G	E	P	C	P	E	P	L	R	T	Y	G	L	P	C	H		
233	CCT	ACT	GGG	GAG	CCC	TGC	CCA	GAG	CCC	CTG	CGT	ACC	TAT	GGG	CTT	CCT	TGC	CAC	286
C	P	F	K	E	V	S	T	*											
287	TGT	CCC	TTC	AAA	GAA	GTA	AGT	ACT	TAG	GGA	GGA	GAG	AGC	GTT	ACC	CCT	GTG	GCT	340
341	AAA	GAG	ATG	GGG	TTT	GGA	GAG	AAG	GGT	CTT	TGC	ATT	CTC	CTT	CTG	CAG	ATC	TGC	394
395	ATG	TCT	CTG	GAT	TTG	TAA	GCC	AGT	GTG	ACC	TAT	CAG	GAA	TCA	CTT	ATC	TTC	CGG	448
459	GAG	CCT	CAG	TTA	TCC	ATC	TAC	GAA	ACG	GGA	GAC	TTG	AAC	TTA	GAT	GTG	ATC	TTC	502
503	AGG	GCC	CTT	TAT	CCA	TAA	TAA	TCC	ATG	CTC	TAC	AGT	GCT	ATG	GCC	GTC	TCT	CAT	556
557	CTT	GTG	TGG	CTG	TTT	TGA	GAA	TGG	GAA	GAG	GGG	TGG	TAG	TTC	ATG	GCT	<u>GCC</u>	<u>ATC</u>	610
611	CTA	GCA	GTG	GCT	CTA	GGA	GAA	AGA	CCC	CAT	CAG	GAA	TT	648					

Fig. 1. Sequence of pGM2A

The nucleotide sequence of pGM2A and its derived amino acid sequence are shown. The plasmid containing the *EcoRI* insert and *EcoRI*/*PstI* subclones were used to obtain the complete sequence according to the strategy depicted. The primer 1 designation of the arrow indicates that primer 1 (below) was used as a sequencing primer. The sequence of nucleotides 1–59 was determined on only one strand since this sequence corresponds exactly to that of the previously published pGAP1 [26] and pGAP2 and 3 (see below). The 266 bp *PstI*/*EcoRI* fragment was used as a GM2A-specific probe for Northern and Southern blot analyses. The primers used for PCR amplification are underlined. Primer 1 extends from nucleotides 200 to 220. The 3' PCR primer (primer 2) extends from nucleotides 607 to 626 and was synthesized as the reverse complement. Amino acids in italicized bold type correspond to the protein sequence. The 5' and 3' *EcoRI* sites indicated on the diagram are linkers.

resulted from an artifact generated during subcloning or the construction of the library from which it was isolated. In order to verify the authenticity of pGM2A, PCR using total placental RNA as the initial template was performed as described in the Materials and methods section. The strategy for the amplification of pGM2A and pGAP was as follows: a primer derived from the common nucleotide sequence present in the 5' region of both pGM2A and pGAP (primer 1, Fig. 1 nucleotides 200–220; Fig. 2 nucleotides 346–366) was used as the 5' primer. The 3' primer was either specific for the 3' untranslated region of pGM2A (primer 2, Fig. 1 nucleotides 607–626) or specific for GAP (primer 3, Fig. 2 nucleotides 838–857). Total RNA from placenta was used as templates to generate the first-strand cDNA for the PCR reaction. The expected fragments of 426 or 511 bp were obtained using the pGM2A-specific and GAP-specific 3' primers respectively. The PCR products were sequenced directly using primer 1. The nucleotide sequence of the 426 bp product was identical with that of pGM2A and the 511 bp fragment was identical with pGAP2. In a separate experiment, the same results were observed using fibroblast RNA as the initial template. Fig. 3 shows the sequences of the PCR products in the region of the

transition from identical to divergent sequence. The point of transition is nucleotide 449 of GAP and 303 of pGM2A.

Amplification of an independent GM2A fragment

It is conceivable that pGM2A plasmid DNA might have contaminated the reagents used for the PCR reactions described above and inadvertently served as the template to generate the 426 bp PCR product. To obtain a PCR product whose plasmid counterpart did not exist in our laboratory, an additional PCR amplification using placental RNA as the template was performed to generate a longer pGM2A cDNA. The pGM2A-specific primer (primer 2, Fig. 1) and a newly synthesized 5' primer corresponding to the 5' end of pGAP2 from nucleotides 16 to 36 (primer 4, Fig. 2) were used in the reaction. The selection of the 5' primer was based on an assumption that the nucleotide sequences at the 5' end of pGM2A and pGAP2 are identical. The 757 bp fragment produced by the PCR reaction was subcloned into Bluescript and sequenced. The nucleotide sequence of the fragment was 100% identical with that of the region of pGM2A encompassed by the primers, with the additional 5' sequence corresponding to pGAP2, as expected (Fig. 4). The results of this

R	A	G	P	P	F	P	M	Q	S	L	M	Q	A	P	L	L	I	
1 CGG	GC	GGC	CCA	CCC	TTC	CCG	ATG	CAG	TCC	CTG	ATG	CAG	GCT	CCC	CTC	CTG	ATC	54
A	L	G	L	L	A	A	P	A	Q	A	H	L	K	K	P	S		
55 GCC	CTG	GGC	TTG	CTT	CTC	GCG	GCC	CCT	GCG	CAA	GCC	CAC	CTG	AAA	AAG	CCA	TCC	108
Q	L	S	S	F	S	W	D	N	C	D	E	G	K	D	P	A	V	
109 CAG	CTC	AGT	AGC	TTT	TCC	TGG	GAT	AAC	TGT	GAT	GAA	GGG	AAG	GAC	CCT	GCG	GTG	162
I	R	S	L	T	L	E	P	D	P	I	I	V	P	G	N	V	T	
163 ATC	AGA	AGC	CTG	ACT	CTG	GAG	CCT	GAC	CCC	ATC	ATC	GTT	CCT	GGA	AAT	GTG	ACC	216
L	S	V	M	G	S	T	S	V	P	L	S	S	P	L	K	V	D	
217 CTC	AGT	GTC	ATG	GGC	ACC	AGT	GTC	CCC	CTG	AGT	TCT	CCT		CTG	AAG	GTG	GAT	270
L	V	L	E	K	E	V	A	G	L	W	I	K	I	P	C	T	D	
271 TTA	GTT	TTG	GAG	AAG	GAG	GTG	GCT	GGC	CTC	TGG	ATC	AAG	ATC	CCA	TGC	ACA	GAC	324
Y	I	G	S	C	T	F	E	H	F	C	D	V	L	D	M	L	I	
325 TAC	ATT	GGC	AGC	TGT	ACC	TTT	<u>GAA</u>	<u>CAC</u>	<u>TTC</u>	<u>TGT</u>	<u>GAT</u>	<u>GTG</u>	<u>CTT</u>	GAC	ATG	TTA	ATT	378
P	T	G	E	P	C	P	E	P	L	R	T	Y	G	L	P	C	H	
379 CCT	ACT	GGG	GAG	CCC	TGC	CCA	GAG	CCC	CTG	CGT	ACC	TAT	GGG	CTT	CCT	TGC	CAC	432
C	P	F	K	E	G	T	Y	S	L	P	K	S	E	F	V	V	P	
433 TGT	CCC	TTC	AAA	GAA	GTA	ACC	TAC	TCA	CTG	CCC	AAG	AGC	GAA	TTC	GTT	GTG	CCT	486
D	L	E	L	P	S	W	L	T	T	G	N	Y	R	I	E	S	V	
487 GAC	CTG	GAG	CTG	CCC	AGT	TGG	CTC	ACC	ACC	GGG	AAC	TAC	CGC	ATA	GAG	AGC	GTC	540
L	S	S	S	G	K	R	L	G	C	I	K	I	A	A	S	L	K	
541 CTG	AGC	AGC	AGT	GGG	AAG	CGT	CTG	GGC	TGC	ATC	AAG	ATC	GCT	GCC	TCT	CTA	AAG	594
G	I	*																
595 GGC	ATA	TAA	CAT	GGC	ATC	TGC	CAC	AGC	AGA	ATG	GAG	CGG	TGT	GAG	GAA	GGT	CCC	648
649 TTT	TCC	TCT	GTT	TTG	TGT	TTG	CCA	AGG	CCA	AAC	TCC	CAC	TCT	CTG	CCC	CCC	TTT	702
703 AAT	CCC	CTT	TCT	ACA	GTG	AGT	CCA	CTA	CCC	TCA	CTG	AAA	ATC	ATT	TTG	TAC	CAC	756
757 TTA	CAT	TTT	AGG	CTG	GGG	CAA	GCA	GCC	CTG	ACC	TAA	GGG	AGA	ATG	AGT	TGG	ACA	810
811 GTT	CTT	GAT	AGC	CCA	GGG	CAT	CTG	CTG	<u>GGC</u>	<u>TGA</u>	<u>CCA</u>	<u>CGT</u>	<u>TAC</u>	<u>TCA</u>	<u>TCC</u>	CCG	TTA	864
865 ACA	TTC	TCT	CTA	AAG	AGC	CTC	GTT	CAT	TTC	CAA	AGC	AGT	TAA	GGA	ATG	GGA	ACC	918
919 AGA	GTG	TTT	TAG	GAC	CTG	AAG	AAT	CTT	TAT	GAC	TCT	CTC	TCT	TTC	ACT	CTT	TTT	972
973 TTT	TTT	TTG	TCA	CTA	AGT	TAA	AAG	CGA	AGT	GAG	AGT	ATT	AAC	GTT	TTT	GTT	CTC	1026
1027 CTC	CGG	CCC	CCT	GTT	ACA	ATG	AAG	GGG	CAA	AAG	TAT	TTG	CTC	TTA	GTC	TAT	TCC	1080
1081 TCC	CTT	AAA	AAA	A	1093													

Fig. 2. Sequence of GAP cDNA

The nucleotide sequence and the derived amino acid sequence of GAP are shown. The sequence was derived from two overlapping cDNAs, pGAP2 and pGAP3. pGAP2 extends from nucleotide 1 to 894; pGAP3 extends from nucleotide 57 to 1093. The complete sequence of each clone was obtained by sequencing *EcoRI* subclones using Bluescript primers and primers 1, a and b (described below) as indicated in the diagram. The regions of pGAP3 (nucleotides 895–1093) and pGAP2 (nucleotides 1–56) which are unique sequences were sequenced on both strands. Primers used for PCR amplification are underlined. Primer 1 extends from nucleotides 346 to 366. The 3' PCR primer (primer 3) extends from 838 to 857 and was synthesized as the reverse complement. Primers a (nucleotides 838–857) and b (reverse strand of nucleotides 163–180) are sequencing primers. Amino acids which correspond to the peptide sequence data are in bold italicized type.

experiment in conjunction with the PCR experiments described above strongly support the existence of at least two GM2 activator mRNAs in both fibroblasts and placenta.

Northern blot analysis of placental and fibroblast RNA

In order to identify the mRNAs corresponding to the pGM2A and GAP cDNAs, Northern blots were prepared using placental and fibroblast poly(A)⁺ RNA (Fig. 5). Identical amounts of each sample were electrophoresed in two sets on the same gel and transferred to a filter. The filter was cut in half to separate the

two sets. One-half was hybridized with a pGM2A-specific fragment, the other with pGAP2 cDNA (see the Materials and methods section). The probes were labelled to a comparable specific radioactivity (approximately 10⁸ c.p.m./μg). Both pGM2A and pGAP2 (Fig. 5a) hybridized to one major mRNA species of approximately 2.3 kb in size. The intensity of the transcript that hybridized with the pGM2A-specific probe was greatly decreased compared with that identified with pGAP2 in both placenta and fibroblast. Hybridization of the same filters with a β-actin probe (Fig. 5b) showed that the variation in the

Table 1. Amino acid sequence of *N*-terminal and tryptic fragments of human kidney GM2 activator

<i>N</i> -Terminal:
FSWDNCDEGKDPAVIRSL(X)LEPDIVVPG(X)VTLS
Tryptic fragments:
1. LGCIK
2. IPCTDYIGSCTFEAFCDVLDMLIPTGE
3. VDLVLEK
4. FSWDNCDEGKDPAVIR
5. EVAGLWIK
6. SEFVVDPDLELPSWLTGNYR
7. TYGLPCHCPFK

(X), Unidentifiable amino acid.

Table 2. Heterogeneities between GAP and pGAP1

Position of nucleotide	Codon and corresponding amino acid	
	GAP*	pGAP1
76	GCC, Ala	ACC, Thr
196	ATC, Ile	GTC, Val
226	ATG, Met	GTG, Val
603	TAA, stop	TAG, stop

* Combined sequence of pGAP2 and pGAP3 (see Fig. 2).

Fig. 3. Sequence of PCR amplification products

GM2A-specific and GAP-specific PCR fragments were generated with primer 1/primer 2 and primer 1/primer 3 respectively using placental total RNA as the initial template for the reverse transcriptase reaction (see the Materials and methods section). The points of divergence of GM2A and GAP are indicated by an arrow on each panel.

Fig. 5. Expression of GM2A and GAP mRNA in placental and fibroblast RNA

Poly(A)⁺ RNAs from placenta (P) and primary fibroblasts (F) were analysed for the presence of GM2A and GAP mRNAs. Northern blots were prepared with 5 μg of RNA. Samples were electrophoresed in duplicate to obtain two identical blots. (a) One of the filters was hybridized with the GM2A-specific *Pst*I/*Eco*RI fragment and the second filter was hybridized with pGAP2. (b) The same filters shown in (a) were stripped and probed with chicken β-actin as a control to demonstrate the amount of RNA in each lane.

						F	P	M	Q	S	L	M	Q	A	P	L	L	I		
						1	TTC	CCG	ATG	CAG	TCC	CTG	ATG	CAG	GCT	CCC	CTC	CTG	ATC	39
40	A	L	G	L	L	C	A	A	P	A	Q	A	H	L	K	K	P	S		
	GCC	CTG	GGC	TTG	CTT	CTC	GCG	GCC	CCT	GCG	CAA	GCC	CAC	CTG	AAA	AAG	CCA	TCC	93	
94	Q	L	S	S	F	S	W	D	N	C	D	E	G							
	CAG	CTC	AGT	AGC	TTT	TCC	TGG	GAT	AAC	TGT	GAT	GAA	GG	131	----					

Fig. 4. Sequence of an extended GM2A fragment

A PCR product longer than the pGM2A cDNA was generated using primer 2 (Fig. 1) and primer 4 which includes nucleotides 16–36 in Fig. 2. The fragment was sequenced to confirm its identity. The sequence which extended beyond the 5' end of pGM2A is shown.

Fig. 6. Hybridization of GM2A and GAP to genomic DNA

Duplicate Southern blots from one electrophoretic gel were prepared with genomic DNA digested with *Bam*HI, *Eco*RI or *Hind*III. Each lane contained 15 µg of DNA. The filters were hybridized with the GM2A-specific probe or pGAP2 cDNA. Lane B, *Bam*HI; lane E, *Eco*RI; lane H, *Hind*III.

intensities of the two mRNAs was not due to differences in the amount of RNA loaded in each lane. This suggests that the pGM2A mRNA is much less abundant than pGAP2 RNA. The 2.3 kb size of the mRNA is about 1.2 kb longer than the longest cDNA clone that has been identified, although the protein sequence of the purified GM2 activator is contained within the clones.

Southern blots of genomic DNA

Southern blots were prepared with 15 µg of genomic DNA digested with *Eco*RI, *Bam*HI or *Hind*III. Duplicate sets of the three digests were electrophoresed on the same gel and transferred to a filter. The filter was then cut in half to divide the two sets and hybridized with either a GM2A-specific probe or the pGAP2 probe (Fig. 6; see the Materials and methods section). Both probes hybridized to a single fragment of approximately 12 kb in the *Hind*III digest (lane H), a 3.5 kb *Eco*RI fragment (lane E) and a 5 kb *Bam*HI fragment (lane B). This result suggests that the two RNA populations defined by the cDNA probes are the products of the same unique gene. Additional faint bands were also observed in the *Hind*III and *Bam*HI lanes despite stringent washing conditions. These bands may be representative of pseudogenes or may be related genes sharing DNA sequence similarity to the GM2 activator gene.

DISCUSSION

The data described above support the existence of at least two mRNAs encoding GM2 activator. Two cDNAs, GAP and GM2A, represent the two mRNA populations. The cDNAs shared a common 5' region, but diverged at their 3' ends. GM2A mRNA was present in much lower abundance than GAP mRNA in both primary fibroblasts and placenta. Minor heterogeneities were also observed when comparing the reported human fibro-

blast cDNA, pGAP1 [26], with the pGAP2 and pGAP3 cDNAs that we identified in placenta. The source of these differences is unclear; they may be attributable to the existence of polymorphisms in the GM2 activator gene or the presence of two highly similar genes. Alternatively, the differences may result from mistakes generated by the reverse transcriptase during the cDNA synthesis reaction. It is unlikely that this occurred during the generation of the placental cDNAs, since the sequence was verified by comparing the sequences of independent clones. The fibroblast sequence, however, represented only one clone [26].

Intriguing questions arise regarding the derivation of these two RNAs. Southern blot analysis of genomic DNA revealed that both GM2A and GAP hybridized to identical bands in *Bam*HI, *Eco*RI, or *Hind*III digests, suggesting that the mRNAs are the product of the same gene. It is therefore likely that the two forms of GM2 activator mRNA result from alternative splicing, especially in view of the dramatic divergence of the sequences of GAP and GM2A. It is also possible that two genes encoding the two forms of GM2 activator mRNA are positioned in tandem in the genome such that probes to each gene would hybridize to the same genomic bands. The origin of the two mRNAs will be resolved by the isolation and characterization of genomic GM2 activator clones.

Northern blot analysis indicates that the GM2A and GAP mRNAs are about 1.2 kb longer than the longest cDNA we obtained. However, as reported by Schröder *et al.* [26], by Fürst *et al.* [37], and from our peptide sequencing, the complete protein sequence of GM2 activator is present within the clones we described. It is probable that GM2 activator has a precursor form which is then processed to the mature protein. This suggestion is supported by the observation that the pGAP2 cDNA sequence has additional 5' coding sequence which extends beyond the N-terminal amino acid identified by protein sequencing. No definitive ATG initiation codon is present, although two methionine codons exist at positions 8 and 12 of the derived amino acid sequence of pGAP2. The nucleotide sequence surrounding neither of the ATG codons matches the consensus sequence characteristic of many translation initiation sites [36]. The mRNAs probably extend 3' as well, since there was no obvious polyadenylation signal, AATAAA, present in either GM2A or pGAP3.

Recently, more evidence has accumulated to show that saposins A–D, which are generated from the prepro-SAP-1 cDNA [38–41], correspond to three of the four classes of activator proteins (see the Introduction). Saposins A and C [39,42] are the activators for the hydrolyses of GlcCer and GalCer; saposin B [39] is a non-specific activator which stimulates the hydrolysis of a number of glycolipids including GalCerSO₄, GM1 and GbOse₃Cer, and saposin D [16] is the activator for the hydrolysis of sphingomyelin. Our results and that of Schröder *et al.* [26] show that GAP and GM2A share no sequence homology with the cDNAs for the prepro-SAP-1 protein. Therefore GM2 activator is not a product of the prepro-SAP-1 gene.

Our results also suggest that there are at least two species of mRNA which encode GM2 activator. Whether both GAP and GM2A can encode biologically active activator proteins requires investigation. It has been noticed that the highly purified GM2 activator isolated from human tissues always appears as a doublet band during electrophoresis [21,43]. Whether the doublet protein is the product of the two species of mRNA is also unknown at this point. The pGAP and pGM2A cDNAs described in this report will provide the tools for expression studies to address these questions.

This work was supported by Public Health Service grant NS09626 from the National Institutes of Health to Y.-T.L. and S.-C.L.

REFERENCES

1. Li, Y.-T. & Li, S.-C. (1982) *Adv. Carbohydr. Chem. Biochem.* **40**, 235–286
2. Li, Y.-T. & Li, S.-C. (1984) in *Lysosomes in Biology and Pathology* (Dingle, J. T., Dean, R. T. & Sly, W., eds.), pp. 99–117, Elsevier, Amsterdam
3. Conzelmann, E. & Sandhoff, K. (1986) *Methods Biochem. Anal.* **32**, 1–23
4. Mehl, E. & Jatzkewitz, H. (1964) *Hoppe-Seyler's Z. Physiol. Chem.* **339**, 260–276
5. Li, S.-C. & Li, Y.-T. (1976) *J. Biol. Chem.* **251**, 1159–1163
6. Li, S.-C., Wan, C.-C., Mazzotta, M. Y. & Li, Y.-T. (1974) *Carbohydr. Res.* **34**, 189–193
7. Li, S.-C., Kihara, H., Serizawa, S., Li, Y.-T., Fluharty, A. L., Mayes, J. S. & Shapiro, L. J. (1985) *J. Biol. Chem.* **260**, 1867–1871
8. Vogel, A., Fürst, W., Abo-Hashish, M. A., Lee-Vaupel, M., Conzelmann, E. & Sandhoff, K. (1987) *Arch. Biochem. Biophys.* **259**, 627–638
9. Li, S.-C., Sonnino, S., Tettamanti, G. & Li, Y.-T. (1988) *J. Biol. Chem.* **263**, 6588–6591
10. Ho, M. W. & O'Brien, J. S. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 2810–2813
11. Peters, S. P., Coyle, P., Coffee, C. J., Glew, R. H., Kuhlenschmidt, M. S., Rosenfeld, L. & Lee, Y. C. (1977) *J. Biol. Chem.* **252**, 563–573
12. Berent, S. L. & Radin, N. S. (1981) *Arch. Biochem. Biophys.* **208**, 248–260
13. Kleinschmidt, T., Christomanou, H. & Braunitzer, G. (1987) *Biol. Chem. Hoppe-Seyler* **368**, 1571–1578
14. Wenger, D. A., Sattler, M. & Roth, S. (1982) *Biochim. Biophys. Acta* **712**, 639–649
15. Christomanou, H. & Kleinschmidt, T. (1985) *Biol. Chem. Hoppe-Seyler* **366**, 245–256
16. Morimoto, S., Martin, B. M., Kishimoto, Y. & O'Brien, J. S. (1988) *Biochem. Biophys. Res. Commun.* **156**, 403–410
17. Li, Y.-T., Mazzotta, M. Y., Wan, C.-C., Orth, R. & Li, S.-C. (1973) *J. Biol. Chem.* **248**, 7512–7515
18. Conzelmann, E. & Sandhoff, K. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3979–3983
19. Hechtman, P. (1977) *Can. J. Biochem.* **55**, 315–324
20. Conzelmann, E. & Sandhoff, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 1837–1849
21. Li, S.-C., Hirabayashi, Y. & Li, Y.-T. (1981) *J. Biol. Chem.* **256**, 6234–6240
22. Conzelmann, E., Burg, J., Stephen, G. & Sandhoff, K. (1982) *Eur. J. Biochem.* **123**, 455–464
23. Meier, E. M., Schwarzmann, G., Fürst, W. & Sandhoff, K. (1991) *J. Biol. Chem.* **266**, 1879–1887
24. Hechtman, P., Gordon, B. A. & Ng Ying Kin, N. M. K. (1982) *Pediatr. Res.* **16**, 217–222
25. Hirabayashi, Y., Li, Y.-T. & Li, S.-C. (1983) *J. Neurochem.* **40**, 168–175
26. Schröder, M., Klima, H., Nakano, T., Kwon, H., Quintern, L. E., Gartner, S., Suzuki, K. & Sandhoff, K. (1989) *FEBS Lett.* **251**, 197–200
27. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1194–1198
28. Mierendorf, R. C., Percy, C. & Young, R. A. (1987) *Methods Enzymol.* **152**, 458–469
29. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
30. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
31. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, N. A. & Arnheim, N. (1985) *Science* **230**, 1350–1354
32. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
33. Jacobson, A. (1987) *Methods Enzymol.* **152**, 254–261
34. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5201–5205
35. Lathe, R. (1985) *J. Mol. Biol.* **183**, 1–12
36. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148
37. Fürst, W., Schubert, J., Machleidt, W., Meyer, H. E. & Sandhoff, K. (1990) *Eur. J. Biochem.* **192**, 709–714
38. Dewji, N. N., Wenger, D. A. & O'Brien, J. S. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8652–8656
39. O'Brien, J. S., Kretz, K. A., Dewji, N. N., Wenger, D. A., Esch, F. & Fluharty, A. L. (1988) *Science* **241**, 1098–1101
40. Nakano, T., Sandhoff, K., Stümper, J., Christomanou, H. & Suzuki, K. (1989) *J. Biochem. (Tokyo)* **105**, 152–154
41. Fürst, W., Machleidt, W. & Sandhoff, K. (1988) *Biol. Chem. Hoppe Seyler* **369**, 317–328
42. Morimoto, S., Martin, B. M., Yamamoto, Y., Kretz, K. A., O'Brien, J. S. & Kishimoto, Y. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3389–3393
43. DeGasperi, R., Li, Y.-T. & Li, S.-C. (1989) *Biochem. J.* **260**, 777–783

Received 28 May 1991/31 July 1991; accepted 7 August 1991