Human Acid Sphingomyelinase

ISOLATION, NUCLEOTIDE SEQUENCE, AND EXPRESSION OF THE FULL-LENGTH AND ALTERNATIVELY SPliced cDNAs*

Edward H. Schuchman†§, Mariko Suchi†, Tsutomu Takahashi†, Konrad Sandhoff§, and Robert J. Desnick†

From the †Division of Medical and Molecular Genetics, Mount Sinai School of Medicine, New York, New York 10029 and §Institut für Organische Chemie und Biochemie der Universität Bonn, Bonn 1, Federal Republic of Germany

(Received for publication, October 15, 1990)

Two types of partial cDNAs encoding human acid sphingomyelinase (EC 3.1.4.12; ASM) were recently isolated from fibroblast and placental cDNA libraries (Quintern, L. E., Schuchman, E. H., Levrán, O., Suchi, M., Ferlinz, K., Reineke, H., Sandhoff, K., and Desnick, R. J. (1989) EMBO J. 8, 2469–2473). The cDNA inserts had identical sequences with the exception of an internal region; type 1 cDNAs (representing ~90% of the ASM cDNAs isolated) had 172 in-frame base pairs (bp), which were replaced in the type 2 cDNAs by a 40-bp in-frame sequence. Northern hybridization and RNase protection studies indicated that both type 1 and 2 transcripts were ~2.5 kilobases; therefore, efforts were directed to isolate full-length type 1 and 2 cDNAs by screening human placental, testis, hepatic, and retinal cDNA libraries. In addition to type 1 and 2 cDNAs, a new type of ASM cDNA (type 3), which did not contain the type 1- or 2-specific regions, was isolated and sequenced. The full-length type 1 and the reconstructed full-length type 2 and 3 cDNAs were transiently expressed in COS-1 cells. Only the full-length type 1 transcript encoded catalytically active human ASM, demonstrating its functional integrity. The 2347-bp full-length type 1 placental cDNA (pASM-1FL) had an 87-bp 5'-untranslated region, an 1890-bp open reading frame encoding 629 amino acids, and a 370-bp 3'-untranslated sequence. The predicted location of the signal peptide cleavage site was after alanine 46. Two base differences were identified in codons 322 and 506 and shown to be polymorphisms with the common alleles having frequencies of 0.6 and 0.7, respectively. To determine the genomic organization of the type 1, 2, and 3 sequences, a 1665-bp genomic region containing both the unique type 1 (172 bp) and type 2 (40 bp) sequences was amplified by the polymerase chain reaction and sequenced. The 172-bp sequence was exonic, flanked by 5'- and 3'-intrinsic sequences of 1052 and 229 bp, respectively. The 40-bp type 2 sequence was intronic, occurring at the 5'-end of the 1052-bp intron due to the use of a cryptic 5'-donor splice site, which deleted the entire 172-bp exon and both flanking intronic sequences. The type 3 cDNA resulted from an alternative splicing event, which excised the 172-bp exon. These studies demonstrate the occurrence of alternative splicing of the ASM transcript, but the existence of only one functional mRNA.

*This work was supported by Research Grant 5 RO1 DK34045 from the National Institutes of Health, and Basil O'Conner Starter Scholar Research Award 5-640, and Basic Research Grant 1-1224, to whom correspondence should be addressed: Division of Medical and Molecular Genetics, Mt. Sinai School of Medicine, 106th St. and Fifth Ave., New York, NY 10029.

††To whom correspondence should be addressed: Division of Medical and Molecular Genetics, Mt. Sinai School of Medicine, 106th St. and Fifth Ave., New York, NY 10029. Tel.: 212-241-9198; Fax: 212-360-1809.

†The abbreviations used are: ASM, acid sphingomyelinase; NPD, Niemann-Pick disease; PCR, polymerase chain reaction; bp, base pairs(s); kb, kilobase(s); nt, nucleotide; Y, a pyrimidine nucleoside; N, a nucleoside; R, a purine nucleoside.
within an internal region of the coding sequence. Type 1 cDNAs contained an in-frame 172-bp sequence encoding 57 amino acids, while in type 2 cDNAs this sequence was replaced by an in-frame 40-bp sequence encoding 13 different amino acids. Screening of both fibroblast and placental cDNA libraries revealed that about 90% of the cDNAs isolated had the type 1 sequence. These findings suggested the existence of two different ASM transcripts, presumably the result of alternative splicing of a common hRNA.

In this communication, we report the isolation and sequencing of a third ASM cDNA and show by transient expression studies that only the full-length type 1 transcript is functional. In addition, PCR amplification and sequencing of a 1665-bp genomic region containing the alternatively spliced regions definitively demonstrated that the type 1, 2-, and 3-specific transcripts result from alternative splicing of a single hRNA. Analysis of this genomic sequence indicated that a weak 5' donor splice site was responsible for the generation of the type 2 and 3 transcripts.

**EXPERIMENTAL PROCEDURES**

**Materials**

Normal human placental tissue was frozen at -70°C within 30 min of delivery and stored until use. Agt11 human placental, testis, and hepatoma cDNA libraries were obtained from Clontech Laboratories (Palo Alto, CA). A Agt11 human retinal cDNA library was kindly donated by Dr. Jeremy Nathans (Johns Hopkins University, Baltimore, MD). Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, the Klenow fragment of DNA polymerase I, RNA molecular weight markers, and cDNA synthesis kits were obtained from New England Biolabs (Beverly, MA) and/or from Boehringer Mannheim. Taq polymerase was purchased from Perkin-Elmer Cetus Instruments, and Sequenase DNA sequencing kits were from U. S. Biochemical Corp. Bluescript vectors and helper phage, RNA transcription kits, Proteinase K, and RNase-free DNase I were kindly provided by Dr. Randal Kaufman, Hadassah Medical Center (Jerusalem, Israel). The eukaryotic expression vector p91023(B) was obtained from Dr. Randal Kaufman, Hadassah Medical Center (Jerusalem, Israel). The eukaryotic expression vector p91023(B) was obtained from Dr. Randal Kaufman, Hadassah Medical Center (Jerusalem, Israel). The eukaryotic expression vector p91023(B) was obtained from Dr. Randal Kaufman, Hadassah Medical Center (Jerusalem, Israel).

**Methods**

**Northern Hybridization, RNase Protection Analyses, and Type 2-specific PCR Amplification—Total cellular RNA from human placenta (~5 g) was prepared by a modification of the guanidine isothiocyanate procedure (20), and poly(A)+ RNA was isolated by oligo(dT)-cellulose chromatography. Aliquots of total (~10 μg) and poly(A)+ (~3 μg) RNAs were analyzed by electrophoresis through denaturing formaldehyde-agarose gels. Northern hybridizations were performed by standard techniques (21) using the radiolabeled pASM-1 insert as a probe. RNase protection experiments were performed according to the method of Zinn et al. (22). A 333-bp BamHI-SacI restriction fragment isolated from pASM-1FL was subcloned into the SK+ Bluescript vector in order to prepare the radiolabeled riboprobe. After RNase treatment, the protected fragments were electrophoresed in 6% polyacrylamide gel and the autoradiogram was determined by densitometry. For the amplification of type 2-specific cDNAs, a type 2-specific antisense oligonucleotide (5'-ATCATTGTTCCACGAGATAGTAC-3') was used with a sense oligonucleotide (5'-ATCATTGTTCCACGAGATAGTAC-3') constructed from the 5' end of the pASM-1FL cDNA insert. For the template, cDNA was prepared from total placental RNA using a cDNA synthesis kit according to the manufacturer's instructions. Alternatively, the pASM-1FL insert was used as a template to demonstrate the type 2 specificity of the PCR amplification.

**cDNA Library Screening and Isolation of the Full-length Human Type 1 ASM cDNA—For library screenings, human placental, testis, hepatoma, and retinal Agt11 cDNA libraries were plated at densities of ~10,000 plaques/150-mm Petri dish. Initially, the placental library was screened using a 404-bp BsrEII fragment isolated from the type 2 human ASM partial cDNA, pASM-2 (19). This fragment contained the type 2-specific 40-bp region, as well as flanking sequences common to the type 1 and 2 cDNAs. The retinal library was screened with an oligonucleotide (5'-GTTCTGCTCTTCACCGCCG-3') constructed from the 5' end of the longest partial type 1 cDNA previously isolated and then was analyzed for the presence of type 2 cDNAs as described below. The testis library was screened first with the type 2-specific oligonucleotide (40-mer) and then with a 608-bp FstI-SacI fragment from the type 2 cDNA (19). The retinal library was screened with the full-length type 1 placental cDNA (pASM-1FL). Random primer labeling of the cDNA probe with [α-32P]dCTP (~3000 Ci/mmol), 5' end-labeling of the oligonucleotides with T4 polynucleotide kinase and [α-32P]dATP (~8000 Ci/mmol), and filter hybridizations were performed by standard methods (25). After three rounds of purification, DNA was isolated from putative positive plaques by the plate lysate method, and the cDNA inserts were analyzed on 1% agarose gels by Southern hybridization (24) with oligonucleotides constructed from the type 1- and 2-specific regions and ASM intronic sequences.

**DNA Sequencing and Computer-assisted Analyses—Dideoxy sequencing was performed by the method of Sanger et al. (26). A putative positive full-length human type 1 ASM cDNA insert (pASM-1FL) and a representative partial type 3 cDNA insert (pASM-3) were isolated after digestion with EcoRI and electrophoresis through agarose gels. Briefly, the inserts were cut out of the gel and electrodes were added to the bands, which were then stained in 0.5 mM ethidium bromide. Following centrifugation at ~10,000 × g, the aqueous phases were re-extracted with phenol and phenol/chloroform (1:1, v/v), and the DNA was then isolated by ethanol precipitation. The purified inserts were subcloned in both orientations into the Bluescript vector SK+ for sense and antisense strands, and single-stranded template was rescued using VCS13 helper phage for dideoxy sequencing according to the manufacturer's instructions. Sequencing primers were synthesized on an Applied Biosystems model 380B DNA synthesizer using phosphoramidite chemistry (26). Computer analyses were performed using the Wisconsin Genetics Computer Group DNA sequence analysis software (version 6.2) and GenBank (release 64) and NBRF (release 25) DNA and protein data bases, respectively.

**Analysis of Polymorphic Sites in the Human ASM Coding Region—** To determine the population frequency of the base differences in codons 322 and 506 of the full-length ASM transcript, PCR amplification of genomic DNA from 20 normal Caucasian individuals was performed on a Perkin Elmer-Cetus thermal cycler using Taq polymerase according to the method of Saiki et al. (27). For the codon 322 base difference, sense (5'-AGATGTCGACTGCTAGACAACTCAG-3') and antisense (5'-AGTGTGCATCGCTTGGACACCCC-3') PCR primers were constructed as described above to amplify a 567-bp genomic DNA fragment. For the codon 506 base difference, sense (5'-AGATTGCACATGGGAGATGTTGCG-3') and antisense (5'-AGATTGTCGACATGGGAGATGTTGCG-3') primers were constructed to amplify a 567-bp genomic DNA fragment. For these studies genomic DNA was rapidly isolated from whole blood by the following procedure. About 0.5 ml of whole blood and 0.5 ml of lysis buffer (10 mM Tris/HC1 buffer, pH 7.5, containing 5 mM MgCl2, 0.32 M sucrose, and 1% Triton X-100) was mixed at room temperature. Following centrifugation at 13,000 × g, the supernatant was removed, and 0.5 ml of PCR buffer (10 mM Tris/HCl buffer, pH 8.3, containing 50 mM KCl, 2.5 mM MgCl2, 0.1 mg/ml gelatin, 0.45% Nonidet P-40, 0.45% Tween 20, and 0.1 mg/ml Proteinase K) was added. The samples were incubated at 60°C for 1 h and boiled for 10 min to inactivate the protease, and then 25 μl was removed for PCR amplification. Following agarose gel electrophoresis of the PCR products, the concentrations of each product was estimated by ethidium bromide staining.

For the dot-blot hybridization analyses (23), about 0.5 μg of DNA was used. The sequence-specific oligonucleotide probes were 5'-AT-GAGACAAATACCGTCGCT-3' (Ile-322), 5'-ATGAAAGAACACCTGT-TC-3' (Thr-322), 5'-ATGACTTCCAGAGACGAGATGAC-3' (Arg-506), and 5'-ATGCTGCGACATGTCGAGAATGAC-3' (Glu-506). Hybridizations were performed for at least 48 h at 42°C. Following hybridization, the blots were washed at room temperature for 15 min in 6 × SSC containing 0.1% sodium dodecyl sulfate and then for 2 h in the same solutions at either 51°C (Ile-322 and Arg-506) or 53°C (Thr-322 and Gly-506).

**Reconstruction of Full-length Type 2 and 3 cDNAs—** A defective cDNA library screening did not identify full-length type 2 or 3 cDNAs, they
were reconstructed as outlined in Fig. 5, A and B. For the full-length type 2 cDNA (pASM-2FL) a 400-bp BstEII restriction fragment containing the type 2-specific 40-bp sequence was isolated from the partial type 2 cDNA, pASM-2, by the phenol/freeze-thaw method described above. The full-length type 1 cDNA, pASM-1FL, was then digested with BstEII to remove the fragment containing the type 1-specific sequence for replacement with the type 2-specific BstEII fragment. Analogously, the full-length type 3 cDNA (pASM-3FL) was reconstructed using a 360-bp type 3-specific BstEII restriction fragment.

**Transient Expression in COS-1 Cells**—For transient expression experiments in COS-1 cells, the full-length pASM-1FL insert and the reconstructed full-length pASM-2FL and pASM-3FL cDNA inserts were subcloned in both orientations into the EcoRI site of the eukaryotic expression vector, p91023(B) (28). DNA (~5–20 μg) from sense and antisense constructs was then transfected into COS-1 cells by the method of Chen and Okayama (29). The transfected cells were harvested after 72 h, and the ASM and β-glucuronidase enzymatic activities were determined using N-[12(pyrenesulfonfonyl)]amidododecanoyl]-sphingomyelin and 4-methylumbelliferyl-β-glucuronide, respectively (30, 31). Neutral sphingomyelinase activities were determined as previously described (32). A unit of enzymatic activity equalled that amount of enzyme which hydrolyzed 1 nmol of substrate/h. Protein determinations were performed by a modified fluorocaine assay (32). In addition to antisense constructs, mock transfections were performed as controls.

**PCR Amplification of Genomic DNA**—For genomic PCR, sense and antisense primers were synthesized as described above. PCR primers 1 and 2 were constructed from exonic sequences which were common to type 1 and 2 cDNAs (see Fig. 6). Genomic DNA was isolated from cultured normal fibroblasts by standard methods (23). PCR amplifications were performed for 30 cycles, and the amplified products were analyzed as described above. For DNA sequencing, the fragments were isolated from the agarose gels, subcloned into Bluescript vectors, and sequenced by the methods described above.

**RESULTS**

**Evidence for the Occurrence of Type 1 and 2 ASM Transcripts**—Since only partial type 1 and 2 ASM cDNAs had been previously isolated from human placental and fibroblast cDNA libraries (19), Northern hybridization analyses were performed to determine the respective sizes and relative amounts of the type 1 and 2 transcripts. As shown in Fig. 1A, a single band of ~2.5 kb was detected when poly(A)⁺ RNA from human placenta was hybridized with the partial type 1 cDNA (which could detect both type 1 and 2 transcripts). Longer exposures of up to 7 days did not reveal additional hybridizing bands (data not shown). Therefore, to demonstrate the occurrence of type 1 and 2 ASM transcripts, RNase protection experiments were carried out. As illustrated in the schematic (Fig. 1B, right), using a type 1 radiolabeled riboprobe (see "Methods"), it was expected that a type 1 transcript would have a 333-bp protected fragment, while a 266-bp fragment would be protected in a type 2 transcript. In human placental poly(A)⁺ RNA, both type 1- and 2-specific transcripts were detected (Fig. 2B). Lanes 2 and 3 showed 16- and 48-h exposures, respectively, while lane 1 was a control protection assay performed without the addition of poly(A)⁺ RNA. Together with the Northern hybridization results, these experiments indicated that human placenta contained type 1 and 2 transcripts of ~2.5 kb. Furthermore, densitometric quantitation revealed that the type 2 transcripts represented from 5 to 10% of the total ASM placental RNAs, consistent with the previous cDNA library screening results.

To further demonstrate the occurrence of full-length type 2 cDNAs, PCR amplification experiments were performed using a sense primer from the 5′ end of the full-length type 1 cDNA (pASM-1FL; see below) and an antisense primer constructed from the type 2-specific 40-bp region (Fig. 2). As shown in panel A, a product of the expected size (~1.1 kb) was amplified, which specifically hybridized with an ASM-specific oligonucleotide probe (panel B). Control experiments also were performed using these primers to PCR-amplify the full-length type 1 cDNA (data not shown). As expected, no amplified products were found, demonstrating the specificity of this PCR for type 2-containing sequences. DNA sequencing of the amplified ~1.1-kb cDNA fragment revealed that the type 1 and 2 cDNAs had identical 5′ sequences.

**Isolation and Characterization of a Full-length Human Type**
Alternative Splicing of Acid Sphingomyelinase Transcripts

1 ASM cDNA—Since the longest type 1 and 2 ASM cDNAs previously isolated were 1879 and 1982 bp, respectively (19), internal cDNA library screenings were undertaken to isolate the respective full-length ASM cDNAs. Screening of ~2 × 10^6 independent recombinants from a human placental cDNA library resulted in the isolation of 84 putative positive human ASM cDNA clones. Agarose gel electrophoresis and Southern hybridization analyses revealed that the cDNA inserts ranged from ~1.2 to 2.4 kb and that ~90% were type 1. Of the nine type 2 cDNAs isolated, the longest insert was ~1.4 kb. Clone pASM-1FL, isolated from the placental library, contained the longest type 1 insert, a 2347-bp sequence which included an 87-bp 5’-untranslated region, an 1890-bp open reading frame encoding 629 amino acids, and a 370-bp 3’-untranslated region (Fig. 3). The coding region contained six N-glycosylation consensus sequences (encoding Asn-X-Thr/Ser) at residues 86–88, 175–177, 335–337, 395–397, 503–505, and 522–524. Although no poly(A) tail was present, a consensus polyadenylation sequence was found at nucleotides 2254–2259, consistent with its position in pASM-1 and pASM-2 (19). There were two in-frame ATGs present in the 5’ region of the pASM-1FL insert, beginning at nucleotides 1 and 97. Using the von Heijne weight matrix method (33), the signal peptidase cleavage site was optimally predicted after residue 46 (von Hejine score = 10.8; Fig. 3, arrow). The next best signal peptide cleavage site was after residue 50 (von Hejine score = 10.1). Interestingly, the predicted signal peptide consisted of a hydrophobic core sequence, which contained five repeats of the amino acid residues leucine and alanine; the corresponding nucleotide sequence contained a 12-nm tandem direct repeat at nucleotides 109–133.

The predicted amino acid sequence of the pASM-1FL insert was collinear with 111 microsequenced residues in trypsin peptides of ASM purified from human urine (2). The four discrepancies between the predicted pASM-1FL amino acid sequence and the microsequenced peptides also occurred in the sequences predicted by the pASM-1 and pASM-2 inserts (19). However, comparison of the predicted amino acid sequences of the full-length type 1 placental cDNA (pASM-1FL) and the previously reported type 1 (and 2) fibroblast cDNAs revealed two other differences. In the placental cDNA, codons 322 (ATA) and 506 (AGG) predicted isoleucine and arginine residues, respectively, whereas in the fibroblast cDNAs the predicted amino acids were threonine (322) and glycine (506) due to single base changes (ACA and GGC, respectively). Dot-blot hybridization studies of 20 normal Caucasian individuals with sequence-specific oligonucleotide probes (data not shown) revealed that the Gly-506 codon had an allele frequency of 0.7 and the Thr-322 codon had an allele frequency of 0.6, indicating that these nucleotide differences are common polymorphisms.

Isolation and Characterization of a Type 3 Human ASM cDNA—Since no full-length type 2 cDNAs were isolated from the placental library, efforts were directed to screen testis, retinal, and hepatoma cDNA libraries. Screening of ~2 × 10^6 independent recombinants from a human testis library with a type 2-specific 40-mer did not detect any type 2 cDNAs. Replica filters were then screened with the type 2 cDNA, pASM-2. Again, no type 2 cDNAs were identified, although 93 type 1 clones were isolated and analyzed. Next, a human retinal cDNA library was intensively screened with pASM-1FL. From ~5 × 10^6 independent recombinants screened, 26 putative positive ASM cDNAs were isolated and analyzed by Southern hybridization. Of these, there were 10 type 1 and one type 2 cDNAs. Again, only a partial type 2 cDNA was identified. The remaining cDNAs isolated from this library were too short to determine if they were type 1 or 2. Finally, ~1.0 × 10^6 recombinants in the hepatoma library were isolated with an oligonucleotide constructed from the 5’ end of pASM-1, the longest partial type 1 cDNA previously isolated. Five putative full-length ASM cDNAs were isolated; however, Southern hybridization analysis demonstrated that they had all type 1 cDNA inserts. Notably, restriction enzyme analysis of the 65 partial human ASM cDNAs isolated from the testis library revealed a third type of human ASM cDNA (pASM-3, type 3). As shown schematically in Fig. 4, the pASM-3 cDNA was 1914 bp and did not contain either the type 1-specific 172-bp region or the type 2-specific 40-bp sequences, but had a truncated open reading frame of 934 bp.

Reconstruction of Full-length Type 2 and 3 cDNAs and Transient Expression of the Full-length ASM cDNAs—Since intensive screening of five different cDNA libraries did not identify full-length type 2 or 3 cDNAs, full-length sequences were reconstructed by the procedure shown in Fig. 5 to test their functional integrity. The reconstructions were based on the fact that PCR amplification and DNA sequencing studies (shown in Fig. 2 for the type 2 cDNA) had revealed that the full-length type 2 and 3 sequences existed in human placenta and that the 5’ sequences were identical to that found in the full-length type 1 insert, pASM-1FL.

The pASM-1FL insert and the reconstructed full-length type 2 (pASM-2FL) and 3 (pASM-3FL) cDNAs were inserted into the transient expression vector, p-91023(B) (28) and transfected into COS-1 cells. As shown in Table I, the mean endogenous ASM activity in COS-1 cells toward the fluorescent natural substrate, N-12-[1-pyrenesulfonyl]amidomido-decanoyl]-sphingomyelin was about 7.1 units/mg protein. The ASM activity in COS-1 cells transfected with the antisense constructs ranged from about 6.2 to 6.7 units/mg protein. In contrast, COS-1 cells transfected with the p91023(B) full-length type 1 sense construct had 30.6 units/mg protein of ASM activity (~5-fold over endogenous levels), demonstrating that the pASM-1FL type 1 transcript expressed catalytically active enzyme. The reconstructed type 2 and 3 transcripts did not express catalytically active enzymes in COS-1 cells. None of the ASM full-length transcripts expressed neutral sphingomyelinase activities in COS-1 cells. As an additional control, the activity of another lysosomal enzyme, β-glucuronidase, also was determined and did not vary significantly from the endogenous levels in any of the transfection experiments.

PCR Amplification of ASM Genomic DNA—In order to determine the origin of the type 1, 2, and 3 cDNAs, an ASM genomic region was PCR-amplified with primers constructed from common exonic sequences flanking the type 1 and 2-specific sequences (Fig. 6). A 1665-bp PCR product was isolated and sequenced. This genomic region contained both the 172- and 40-bp type 1 and 2-specific sequences. Interestingly, the 172-bp type 1 sequence was exonic, flanked by 1052-bp and 229-bp introns. The 40-bp type 2-specific sequence was located at the 5’ end of the 1052-bp intron. Within this intron there also were two poly(T) tracts of 20 and 23 nt at positions 313–332 and 469–491 and five potential lariat branch points (labeled a–e, Fig. 6). Within the 229-bp intron there were two potential lariat branch points that fit the consensus sequence, YNYURAY (34, 35), at positions 1574–1580 and 1510–1516 (underlined in Fig. 6). The second potential branch point, 77 nt upstream from the A2 acceptor splice site, is followed by a polyuridylic tract.

Table II shows the donor (D1 and D2) and acceptor (A1
Fig. 3. Nucleotide and predicted amino acid sequences of the full-length human ASM cDNA, pASM-1FL. The pASM-1FL insert was sequenced in both orientations. The unique 172-bp type 1 sequence is bracketed. Underlined amino acid residues represent residues that were colinear with the amino acid sequences from tryptic peptides of the purified enzyme (T-1 to T-12). The boxed amino acid residues are those that were different from the fibroblast cDNAs, pASM-1 and pASM-2. CHO represents potential N-glycosylation sites.
FIG. 4. Schematic representation of human ASM type 1, 2, and 3 cDNAs. The longest type 1, 2, and 3 cDNAs isolated by library screening are shown schematically. The type 1- and 2-specific sequences are indicated (172 and 40 bp, respectively), as are the locations of the stop codons. A and B are common 5' and 3' coding sequences, respectively. Note that the partial type 3 cDNA has a premature termination codon (TAA).

FIG. 5. Reconstruction of full-length type 2 and 3 human ASM cDNAs. The full-length type 2 (A) and 3 (B) cDNAs were reconstructed as described under "Methods." The 172- and 40-bp type 1- and 2-specific sequences are indicated, as are the flanking BstEII restriction sites.

and A2) splice site sequences at the intron/exon boundaries within this ASM genomic region, as well as the sequence of the cryptic donor splice site (D3) located adjacent to the 3' end of the 40-bp type 2-specific sequence. Note that neither of the donor sites within this region perfectly matched the consensus sequence (36) and, in particular, there was a G to A transition within donor splice site D2, located at the 3' end of the type 1-specific 172-bp exon. Compared to the D1 and D2 donor sites, the D3 cryptic splice site adjacent to the 40-bp type 2-specific sequence best matches the donor consensus sequence.

DISCUSSION

Previously, we reported the isolation of partial type 1 and 2 cDNAs for human ASM, the longest inserts being 1879 (pASM-1) and 1382 (pASM-2) bp, respectively. Type 1 cDNAs had a unique 172-bp sequence encoding 57 amino acids which was replaced in the type 2 cDNAs by a 40-bp sequence encoding 13 different amino acids. About 90% of the 113 partial cDNAs isolated from human fibroblast and placental libraries were type 1. In the studies reported here, Northern hybridization analyses revealed the presence of a single ~2.5-kb transcript in placental poly(A)+ RNA. Subsequent RNase protection studies demonstrated the occurrence of both type 1 and 2 transcripts. Thus, efforts were undertaken to isolate full-length type 1 and 2 ASM cDNAs by intensive screening of cDNA libraries from five different tissues.

Full-length type 1 cDNAs were isolated including the 2347-bp cDNA, pASM-1FL. The size of the pASM-1FL insert was consistent with the occurrence of the ~2.5-kb transcript observed in the Northern hybridization experiments, the ~150-bp difference due to upstream 5'-untranslated sequences and the length of the poly(A) tract. The full-length cDNA had an open reading frame of 1890 bp which contained two in-frame potential initiation codons. Since the enzyme's N terminus was blocked (19), it is not known which initiation codon was used in vivo. It is possible that both initiation ATGs could be used as is the case for another hydrophobic lysosomal hydrolase, acid β-glucosidase (37). However, compared to the translation initiation consensus sequence GCC(AG)CCATGG (38), the sequence flanking the second ATG in human ASM is weak, particularly since position -3 contains a thymidine residue. Therefore, it is likely that the first ATG is the in vivo initiation codon. Using the von Heijne weight-matrix method (33), the optimal signal peptide cleavage would occur after amino acid residue 46. The predicted 14 amino acids preceding the signal peptide cleavage site have a particularly hydrophobic core consisting of five leucine/alanine repeats. Six potential N-glycosylation sites were predicted in the mature ASM polypeptide. At present it is not known which of the glycosylation sites are utilized. However, treatment of the purified urinary enzyme with glycopeptidase F reduced the molecular mass from ~72 kDa to ~61 kDa, indicating that perhaps as many as five sites may be glycosylated (19).
Alternative Splicing of Acid Sphingomyelinase Transcripts

Fig. 6. Sequence of the PCR-amplified genomic region containing the unique type 1- and 2-specific human ASM regions. PCR amplification of human genomic DNA was performed using primers 1 and 2. Upper and lower case letters indicate exonic and intronic sequences, respectively. The type 1- and 2-specific genomic sequences are shown in boldface type. Boxed sequences D1 to D3, and A1 and A2 indicate 5‘ donor and 3‘ acceptor splice site sequences, respectively. The potential intron branch point consensus sequences are underlined and designated a-e.

Table II
5’ Donor and 3’ acceptor splice sites in the PCR-amplified genomic ASM region

| Splice site sequence | Nucleotide position donor | Nucleotide position acceptor |
|----------------------|--------------------------|------------------------------|
| Consensus            | CAG gttaag Ag           | ncag G                       |
| D1                   | 136–144                  |                             |
| D2                   | 1360–1368                | AAA gttagg                   |
| D3                   | 176–184                  | aag gtgaat                   |
| A1                   | 1187–1190                | tcaag A                      |
| A2                   | 1586–1592                | ctag G                       |

Efforts to identify a full-length type 2 ASM cDNA included intensive screening of five libraries. Screening of placental and retinal libraries resulted in the identification of fourteen additional type 2 cDNAs (~12% of the total cDNAs analyzed); however, none were longer than the previously obtained pASM-2 insert (1382 bp). Screening of hepatoma and testis libraries did not identify any type 2 clones, but a third 1914-bp ASM cDNA (type 3, pASM-3) was identified which lacked the unique type 1 and 2 sequences.

In order to determine the functional integrity of the three different ASM transcripts, it was necessary to reconstruct full-length type 2 and 3 coding sequences (Fig. 5). Prior to performing these reconstruction experiments, the existence of full-length type 2 and 3 cDNAs was shown by PCR amplification of total placental cDNA (shown in Fig. 2 for the type 2 cDNA) and sequencing of the amplified products. Transient expression in COS-1 cells of pASM-1FL and the reconstructed type 2 and 3 cDNAs demonstrated that only the type 1 transcript was functional. The fact that the type 1 mRNA did not express neutral sphingomyelinase activity is COS-1 cells supports the notion that the acid and neutral sphingomyelinases are encoded by different genes.

Two nucleotide differences were initially identified by sequencing type 1 cDNAs from fibroblast and placental libraries (19). The functional integrity of the full-length pASM-1FL sequence from placenta (i.e. Ile-322 and Arg-506) was demonstrated by the transient expression of active enzyme in COS-1 cells (Table I). Subsequent analysis of genomic DNA from 20 unrelated individuals revealed that the base differences in codons 322 and 506 occurred in the Caucasian population as polymorphisms with allele frequencies of 0.9 and 0.7 for the Thr-322 and Gly-506 codons, respectively. The Gly-506 polymorphism creates MspI and NciI restriction sites.
single ASM hnRNA. As shown diagrammatically in Fig. 7, type 1 transcripts result from normal splicing events which remove the 1052- and 229-bp introns, while type 2 transcripts result from splicing at a cryptic site which excises 1012 bp of the large intron, the 172-bp type 1 exon, and the 229-bp intron. The occurrence of the type 3 cDNA can be explained by alternative splicing to the upstream donor splice site, D1, thereby deleting the 172-bp exon.

Splicing of mammalian transcripts is initiated by cleavage at the 5′ donor splice site followed by lariat formation at a branch point, generally within 50 nt of the 3′ acceptor site (34, 35). Then there is a cleavage of the 3′ exon at the acceptor splice site and ligation of the adjacent exons. However, in the human ASM hnRNA, there is a weak 5′ donor splice site adjacent to the 172-bp exon (D2, Table II) that does not function in about 10% of the splicing events, thereby generating the type 2 or 3 transcripts. As shown in Fig. 7, if the cryptic donor splice site (D3) adjacent to the 40-bp intronic sequence is used, a type 2 transcript is generated. The G to A transition of the nucleotide immediately adjacent to the invariant GT consensus dinucleotide in D2 may be particularly important, since this alteration has been shown to cause abnormal splicing of the proα1(I) collagen gene leading to Ehlers-Danlos syndrome Type VII (39). In fact, the D3 cryptic splice site more closely matches the consensus sequence than either of the other two identified donor splice sites, D1 or D2, which are used to generate type 1 transcripts (Table II). The presence of two poly(T) tracts of 20 and 23 nt at the 5′ end of the large intron may cause abnormal RNA secondary structure, perhaps favorably positioning the cryptic splice site, D3. The rare type 3 transcript also is generated by alternative splicing of the 172-bp exon, but in this case splicing proceeds to the upstream donor splice site D1 rather than to the D3 cryptic donor splice site.

Other features of the 1052-bp intron also deserve note. There are five potential lariat branch point sequences that fit the consensus sequence YNYURAY (34) located near the 3′ end of this intron (labeled a–e in Fig. 6). Only one of these potential branch points is followed by a polypyrimidine tract (b), however this branch point is 336 bp upstream from the 3′ acceptor site. It is generally assumed that the branch point should be within 20–50 nt of the 3′ acceptor and not closer than 70 nt to the 5′ donor site. Perhaps after cleavage at the D3 cryptic donor splice site, a lariat cannot efficiently form at branch points within this intron, and thus, the lariat occurs at the next available branch point, which is located in the 229-bp intron.

Alternative splicing also occurs in the transcripts for two other human lysosomal enzymes, β-glucuronidase (40) and β-galactosidase (41), β-Glucuronidase cDNAs, which had an internal 153-bp deletion, were identified in human fibroblast and placental cDNA libraries. The shorter cDNA had lost an entire exon due to alternative splicing and expressed an enzyme protein that was not catalytically active in COS-1 cells. For β-galactosidase, two distinct groups of cDNA clones were isolated from human fibroblast cDNA libraries. The shorter cDNAs were missing two noncontiguous protein coding regions present in the full-length cDNAs and produced truncated proteins, which were targeted to the perinuclear region in COS-1 cells.

In summary, three types of human ASM transcripts have been identified. Genomic PCR amplification and sequencing studies demonstrated that each of the ASM transcripts were derived from a single hnRNA. The type 1 transcript is the major ASM species and encodes a catalytically active enzyme. The type 2 and 3 ASM transcripts result from alternative splicing, most likely due to the presence of a weak donor splice site (D2) adjacent to the 172-bp type 1-specific exonic sequence. Reconstruction and transient expression of full-length type 2 and 3 transcripts revealed that these sequences did not encode functional enzymes. The availability of the full-length type 1 cDNA will permit characterization of the human ASM gene and structure/function studies of the ASM polypeptide, as well as investigations of the mutations which cause the neuronopathic and non-neuronopathic forms of Niemann-Pick disease.

Acknowledgments—We wish to thank Constantine Zamfirescu for expert technical assistance and Orna Levin for valuable contributions throughout the course of these studies.

REFERENCES

1. Levade, T., Saivayre, R., and Doute-Blazly, L. (1986) J. Clin. Chem. Clin. Biochem. 24, 205–220
2. Quinlen, L. E., Weitz, G., Nehrkorn, H., Tager, J. M., Schram, A. W., and Sandhoff, K. (1987) Biochim. Biophys. Acta 922, 323–336
3. Yamanaka, T., and Suzuki, K. (1982) J. Neurochem. 38, 1753–1764
4. Pena, L., Desnick, R. J., Adler, D., Disteche, C. M., and Schuchman, E. H. (1991) Genomics, 9, 229–234
5. Gatt, S., Dinur, T., and Kopolovic, J. (1978) J. Neurochem. 31, 547–551
6. Chatterjee, S., and Ghosh, N. (1989) J. Biol. Chem. 264, 12554–12561
7. Spence, M. W., Byers, D. M., Palmer, F. B. St. C., and Cook, H. W. (1989) J. Biol. Chem. 264, 5538–5563
8. Kolesnick, R. W., and Clegg, S. (1988) J. Biol. Chem. 263, 6534–6537
9. Hannun, Y. A., and Bell, R. M. (1989) Science 243, 500–507
10. Schneider, P. B., and Kennedy, E. P. (1987) J. Lipid Res. 8, 202–209
11. Brady, R. O., Kanfer, J. N., Mock, M. B., and Fredrickson, D. S. (1966) Proc. Natl. Acad. Sci. U. S. A. 65, 366–369
12. Spence, M. W., and Callahan, J. W. (1989) in The Metabolic Basis of Inherited Diseases (Scrivcr, C. R., Beaudet, A. L., Sly, W. S., Valle, D. eds) 8th Ed., pp. 1655–1676, McGraw-Hill, New York
13. Wengcr, D. A., Sattler, M., Kudoh, T., Snyder, S. P., and Kingsport, R. S. (1986) Science 208, 1471–1473
14. Bunza, A., Lowden, J. A., and Charlton, R. M. (1979) Vet. Pathol. 16, 530–535
15. Miyawaki, S., Yoshida, H., Mitsuoka, S., Enomoto, H., and Ikehara, S. (1986) J. Hered. 77, 379–384
16. Sakiyama, T., Tsuda, M., Owash, M., Kitagawa, T., Miyawaki, S., Shinagawa, T., and Tadokoro, M. (1986) J. Inherited Metab. Dis. 9, 305–308

**Fig. 7. Proposed model for alternative splicing of ASM transcripts.** Type 1 transcripts result from normal splicing events, whereas type 2 RNAs occur due to a single splice which brings together the 3′ acceptor, A2 (nt 1590–1591), with the cryptic 5′ donor site, D3 (nt 179–180). Thus, the type 2 transcript deletes the 172-bp exon which is replaced by 40 in-frame intronic bp. The type 3 transcript results from a splicing event which joins the 3′ acceptor, A2, with the 5′ donor, D1 (nt 139–140).
Alternative Splicing of Acid Sphingomyelinase Transcripts

17. Kitagawa, T. (1987) Jpn. J. Hum. Genet. 32, 55–69
18. Besley, G. T. N., Hoogeboom, A. J. M., Hoogeveen, A., Kleijer, W. J., and Galjaard, H. (1980) Hum. Genet. 54, 409–412
19. Quintern, L. E., Schuchman, E. H., Levan, O., Suchi, M., Perlinz, K., Reinke, H., Sandhoff, K., and Desnick, R. J. (1989) EMBO J. 8, 2469–2473
20. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294–5299
21. Thomas, P. S. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 5201–5205
22. Zinn, K., DiMaio, D., and Maniatis, T. (1983) Cell 34, 865–879
23. Sambrook, J., Fritsch, E. F., and Maniatis, T. A. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
24. Southern, E. M. (1975) J. Mol. Biol. 98, 503–517
25. Sanger, F., Nicklen, J., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
26. Itakura, K., Roasi, J. J., and Wallace, R. B. (1984) Annu. Rev. Biochem. 53, 323–356
27. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) Science 239, 487–491
28. Kaufman, R. J., and Sharp, P. A. (1982) Mol. Cell. Biol. 2, 1304–1319
29. Chen, C., and Okayama, H. (1987) Mol. Cell. Biol. 7, 2747–2762
30. Klar, R., Levade, T., and Gatt, S. (1988) Clin. Chim. Acta 176, 259–268
31. Brot, F. E., Bell, C. E., Jr., and Sly, W. S. (1978) Biochemistry 17, 385–391
32. Bishop, D. F., and Desnick, R. J. (1981) J. Biol. Chem. 256, 2469–2473
33. Sambrook, J., Fritsch, E. F., and Maniatis, T. A. (1989) EMBO J. 8, 2469–2473
34. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986) Annu. Rev. Biochem. 55, 1119–1150
35. Green, M. R. (1986) Annu. Rev. Genet. 20, 671–708
36. Mount, S. M. (1982) Nucleic Acids Res. 10, 439–472
37. Sorge, J., Kuhl, W., West, C., and Beutler, E. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 906–910
38. Kosak, M. (1987) Nucleic Acids Res. 15, 8126–8149
39. Weil, D., D’Alessio, M., Ramirez, F., de Wet, W., Cole, W. G., Chan, D., and Bateman, J. F. (1989) EMBO J. 8, 1705–1710
40. Oshimia, A., Kyle, J. W., Miller, R. D., Hoffmann, J. W., Powell, P. P., Grubb, J. H., Sly, W. S., Tropak, M., Guise, K. S., and Gravel, R. A. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 685–689
41. Morreau, H., Galjart, N., Gillemans, N., Willemsen, R., van der Horst, G. T., and d’Azzo, A. (1989) J. Biol. Chem. 264, 20655–20663