THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 266, No. 12, Issue of April 25, pp. 7556-7560, 1991
Printed in U.S.A.

Sulfatide Activator Protein

ALTERNATIVE SPLICING THAT GENERATES THREE mRNAs AND A NEWLY FOUND MUTATION RESPONSIBLE FOR A CLINICAL DISEASE*

(Received for publication, November 6, 1990)

Hans Holtschmidt‡, Konrad Sandhoff‡§, Hae Y. Kwon‡, Klaus Harzer¶, Takeshi Nakano‖, and Kunihiko Suzuki‖

From the ‡Institut für Organische Chemie und Biochemie, Universität Bonn, W-5300 Bonn 1, Federal Republic of Germany, the ¶Institut für Hirnforschung, Eberhard-Karls Universität Tübingen, W-7400 Tübingen 1, Federal Republic of Germany, and the ‖Brain and Development Research Center, Departments of Neurology and Psychiatry, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

The sulfatide activator protein, also known as SAP-1, is derived from a gene that generates an mRNA coding for four homologous proteins. Its physiological function is to stimulate hydrolysis of sulfatide by arylsulfatase A *in vivo*. A genetic defect in the sulfatide activator results in a metabolic disorder similar to classical metachromatic leukodystrophy, which is itself caused by a genetic defect in arylsulfatase A. In a patient with sulfatide activator deficiency, a nucleotide transversion $G^{722} \rightarrow C$ (counted from A of the initiation codon ATG) was found in the mRNA of the sulfatide activator precursor, resulting in the substitution of serine for Cys^{241} in the mature sulfatide activator. The remainder of the coding sequence was completely normal except for a polymorphism C to T in position 1389, which does not change the amino acid sequence. The patient produces at least three different forms of mRNA for the precursor. Two of them include a stretch of an additional 9 and 6 bases, respectively, within the sulfatide activator coding region. In normal individuals this stretch of additional bases has also been observed. This could be explained by the presence of a small 9-base pair exon which can be introduced, or not, by alternative splicing as a stretch of 9 or 6 bases into the mature mRNA. The shortest form of the mRNA yields an active sulfatide activator (Fürst, W., Schubert, J., Machleidt, W., Meier, H. E., and Sandhoff, K. (1990) *Eur. J. Biochem.* 192, 709–714).

Physiological degradation of sulfatide (galactosylceramide sulfate) requires a lysosomal hydrolase, arylsulfatase A,[1] and an additional small lysosomal glycoprotein, the sulfatide activator protein (SAP-1) (1, 2). Genetic defects in either of these lysosomal proteins cause phenotypically similar disorders. Generically these disorders are classified as metachromatic leukodystrophy, in which an abnormal accumulation of sulfatides occurs in most organs, most prominently in the nervous system. In addition to sulfatide degradation, the sulfatide activator can activate *in vitro* degradation of G_{M1}[2] ganglioside by β-galactosidase (3) and globotriaosylceramide by α-galactosidase A (3, 4). The cDNA coding for the sulfatide activator, first reported by Dewji *et al.* (5), has been recognized more recently to code for four different proteins (6–10). The gene generates processed mRNA, which is translated into a large precursor polypeptide, which is in turn post-translationally processed to yield four small, homologous proteins. The sulfatide activator is the second among the four protein domains from the N terminus of the nascent polypeptide. The third protein (A_1 activator, SAP-2, Gaucher factor) is known to activate degradation of glucosylceramide by glucosylceramidase (11). Its physiological role is established because a Gaucher-like disorder occurs when it is genetically defective (12). Activation of the hydrolysis of sphingomyelin and glucosylceramide by the other two processed proteins has been reported *in vitro* in the presence of detergents (13, 14). The physiological significance of these two proteins has not been established definitively.

The mRNA of the precursor of the sulfatide activator from normal individuals exhibits polymorphism. Beside the shortest mRNA which is colinear with the amino acid sequence of the mature sulfatide activator protein (15), a longer one with an additional stretch of 9 bases within the sulfatide activator coding region has been described (10). In the present paper we found that some mRNA contains only the 6 3′-terminal bases of the 9-base stretch. Furthermore, Zhang *et al.* (16) found, in a patient with a genetic defect in the sulfatide activator, a 33-base insertion with the last 9 bases being identical to those found in normal individuals. The source of these inserted sequences has yet not been clarified. In addition to the 33-base insertion, a point mutation within the sulfatide activator which abolishes the only glycosylation site has been identified in another patient with sulfatide activator deficiency (17, 18).

We describe in this report the source and mechanism for the polymorphism of the mRNA for the sulfatide activator

* This investigation was supported in part by Deutsche Forschungsgemeinschaft, Mental Retardation Research Center Core Grant P30 HD-03110, and National Institutes of Health Research Grants R01 NS-24289 and RO1 NS-28997. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M60255-M60259 and D00422 (DNA Data Bank of Japan).

§ To whom correspondence should be sent: Institut für Organische Chemie und Biochemie, Gerhard-Domagk-Straße 1, W-5300 Bonn 1, FRG. Tel: 49-228-73-77-78; Fax: 49-228-73-56-83.

[1] The enzymes used are: arylsulfatase A (aryl-sulfate sulfohydrolase (EC 3.1.6.1)), also called cerebroside-sulfatase (cerebroside-3-sulfate 3-sulfohydrolase (EC 3.1.6.8)); β-galactosidase (β-D-galactoside galactohydrolase (EC 3.2.1.23)); α-galactosidase (α-D-galactoside galactohydrolase (EC 3.2.1.22)); and glucosylceramidase (D-glucosyl-*N*-acylsphingosine glucohydrolase (EC 3.2.1.45)).

[2] The abbreviated designations of glycolipids are according to Svennerholm's nomenclature for ganglio series gangliosides (Svennerholm, L. (1963) *J. Neurochem.* **10**, 613–623).

precursor and a newly found point mutation responsible for the clinical phenotype of sulfatide activator deficiency.

EXPERIMENTAL PROCEDURES

Patient

The patient is a 7-year-old Arabic boy, the third of four children of parents who are first cousins. There is no family history of neurological illness. The patient was initially described as an abnormal case of metachromatic leukodystrophy with normal arylsulfatase A but disturbed fibroblastic sulfatide turnover.[3] These findings collectively indicated sulfatide activator deficiency as the cause of the disease. The diagnosis was confirmed by the absence of cross-reactive materials against anti-sulfatide activator antiserum in fibroblast extracts.[3] Cross-reactive material, however, was present for the A_1 protein (SAP-2, glucosylceramidase activator, Gaucher factor) (19).

Commercial Materials

Restriction endonucleases, T_4 ligase, polynucleotide kinase, DNA polymerase (Klenow), and the T_7 polymerase sequencing kit were obtained from Pharmacia LKB Biotechnology Inc. Avian myeloblastosis reverse transcriptase was obtained from Promega Biotec (Madison, WI), and Taq polymerase was from Perkin-Elmer Cetus Instruments. [γ-^{32}P]ATP, [α-^{32}P]dATP, and [^{35}S]dATP were obtained from Amersham Corp. The oligonucleotide primers were synthesized in the nucleotide synthesis laboratory of the Program in Molecular Biology and Biotechnology, University of North Carolina, under the supervision of Dr. Dana Fowlkes and in the Institut für Genetik der Universität Bonn under the supervision of Hella Lichtenberg-Frate.

Cell Lines

The cultured fibroblasts of the patient were first established in Tübingen, FRG, from a primary culture provided by Dr. Haas-Andela (Giessen, FRG). They and the normal control skin fibroblasts, obtained from a patient with juvenile form of Sandhoff disease, were maintained in Eagle's minimal essential medium supplemented with 10% fetal calf serum, 0.48% glucose, and 1% penicillin/streptomycin. The cells were harvested in a solution of 150 mM NaCl with a rubber scraper.

Methods

Isolation of RNA—The guanidine isothiocyanate-cesium chloride method was used to isolate RNA from cultured fibroblasts (20). The poly(A)-positive RNA fraction was prepared by affinity chromatography on oligo(dT)-cellulose. Northern transfer of electrophoresed mRNA onto Nylon membrane was done according to Maniatis *et al.* (20).

Isolation of Genomic DNA—Genomic DNA was isolated from cultured skin fibroblasts. After removing the culture medium and washing the monolayer with phosphate- buffered saline the cells were harvested in a solution of 150 mM NaCl with a rubber scraper. After washing three times in phosphate-buffered saline the cells were resuspended in 0.3 ml of 150 mM NaCl solution containing 25 mM EDTA. The cell suspension was frozen at −70 °C and thawed. 50 μl of proteinase K solution (20 mg/ml) and 10% sodium dodecyl sulfate solution were added per 100 μl of suspension and left at 37 °C overnight. Genomic DNA was extracted with phenol/chloroform, precipitated with ethanol in the presence of 0.5 volume of 7.5 M ammonium acetate, and washed twice with 70% ethanol. Dried genomic DNA was dissolved in water by warming to 37 °C for 1–2 h.

Polymerase Chain Reaction—The poly(A)-positive RNA fraction was used as the starting material for amplification of the entire coding sequence of the sulfatide activator precursor. The first (antisense) cDNA strand was synthesized with the Promega cDNA synthesis kit using oligo(dT) as the primer and avian myeloblastosis reverse transcriptase. The entire protein coding sequence including 33 bases of 5′-untranslated region and 77 bases of 3′-untranslated region (nucleotides −33 to 1661) was amplified in overlapping segments using a series of synthetic oligonucleotide primers (Fig. 1) with Taq polymerase in a thermal cycler (Biometra, Göttingen, FRG). The amplification conditions were: initial denaturation at 94 °C for 5 min, 30 cycles of amplification with annealing for 3 min at 10 °C below the lower of the calculated dissociation temperature of the two primers, extension at 72 °C for 2 min, and the final extension at 72 °C for 10 min.

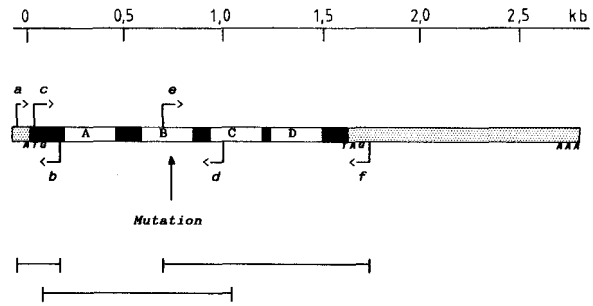

FIG. 1. **Strategy for amplification of the mRNA of the precursor of the sulfatide activator.** The four *white regions* represent the four protein domains (*A*, saposin A; *B*, sulfatide activator; *C*, A_1 activator; *D*, component C). ▮, the coding region of the precursor of the sulfatide activator from ATG to TAG; □, noncoding region. The precursor of the sulfatide activator cDNA covering the entire protein coding sequence was obtained in three overlapping segments (*thin continuous lines*) by polymerase chain reaction. It used three pairs of synthetic oligonucleotide primers (a and b, c and d, e and f). The cDNA was amplified and sequenced from 33 base pairs before ATG and to 77 base pairs after the stop codon (position 1661 numbered from A of the initiation codon). *kb*, kilobases. a = 5′GGC GCA TTG CAG ACT GCG G 5′sense primer, b = 5′GTC TTT GCA TAT GTC GCA GGG 5′antisense primer, c = 5′TCC TTG GAC TGA AAG AA 5′sense primer, d = 5′TTG TTG TTG TCA ATG AG 5′antisense primer, e = 5′CGC CTG GGC CCT GGC ATG GCC 5′sense primer, f = 5′GTG CGT TCA TTC CCC CAG ACA CAC 5′antisense primer.

An intronic segment flanking the variable stretch of 9 bp[4] was amplified from the patient's genomic DNA with a pair of primers synthesized from the intron sequence which had been determined from genomic fragments obtained from a normal human genomic library in λ EMBL-4.[5] The sequence of the primers were 5′-GCTGGCCCAGAGCAGACATTCA (5′ primer, sense strand) and 5′-TGGAGAGTCTCCTAGCCAGAGG (3′ primer, antisense strand). The amplification procedure was as described above.

Sequencing of Polymerase Chain Reaction Products—The amplified DNA segments were subcloned into the plasmid pUC18 using the *Sma*I site. The plasmids containing subclones of interest were identified by hybridization with appropriate radiolabeled oligonucleotides. Sequencing was done by the dideoxy chain termination method (21) on the double-stranded pUC18 DNA. The T_7 polymerase sequencing kit from Pharmacia was used with appropriate sequencing primers and dATPαS.

RESULTS

A preliminary Northern blotting study indicated that fibroblasts of the patient generated stable mRNA for the sulfatide activator precursor of normal size (data not shown). This finding allowed us to approach the gene abnormality in the patient from an examination of mRNA/cDNA.

Sequencing of the cDNA clones covering the entire protein coding region showed that there was a single nucleotide change at position 722 (counted from A of the initiation codon ATG) from the normal G to C (Fig. 2). This transversion was observed in all of the five batches of cDNA clones, each prepared by amplification of different batches of cellular RNA. The remainder of the protein coding sequence was entirely normal except for a polymorphism C to T in position 1398, which does not change the amino acid sequence (not shown). The transversion results in a substitution of serine for normal cysteine (Fig. 2). At the genomic level this mutation was observed after six different polymerase chain reac-

[3] Schlote, W., Harzer, K., Christomanou, H., Paton, B. C., Kustermann-Kuhn, B., Schmid, B., Seeger, J., Beudt, U., Schuster, I., and Langenbeck, U. (1991) *Eur. J. Pediatr.*, in press.

[4] The abbreviations used are: bp, base pair(s); dATPαS, deoxyadenosine 5′-[α-^{35}S]thiotriphosphate.

[5] Holtschmidt, H., Sandhoff, K., Fürst, W., Kwon, H. Y., Schnabel, D., and Suzuki, K. (1991) *FEBS Lett.*, in press.

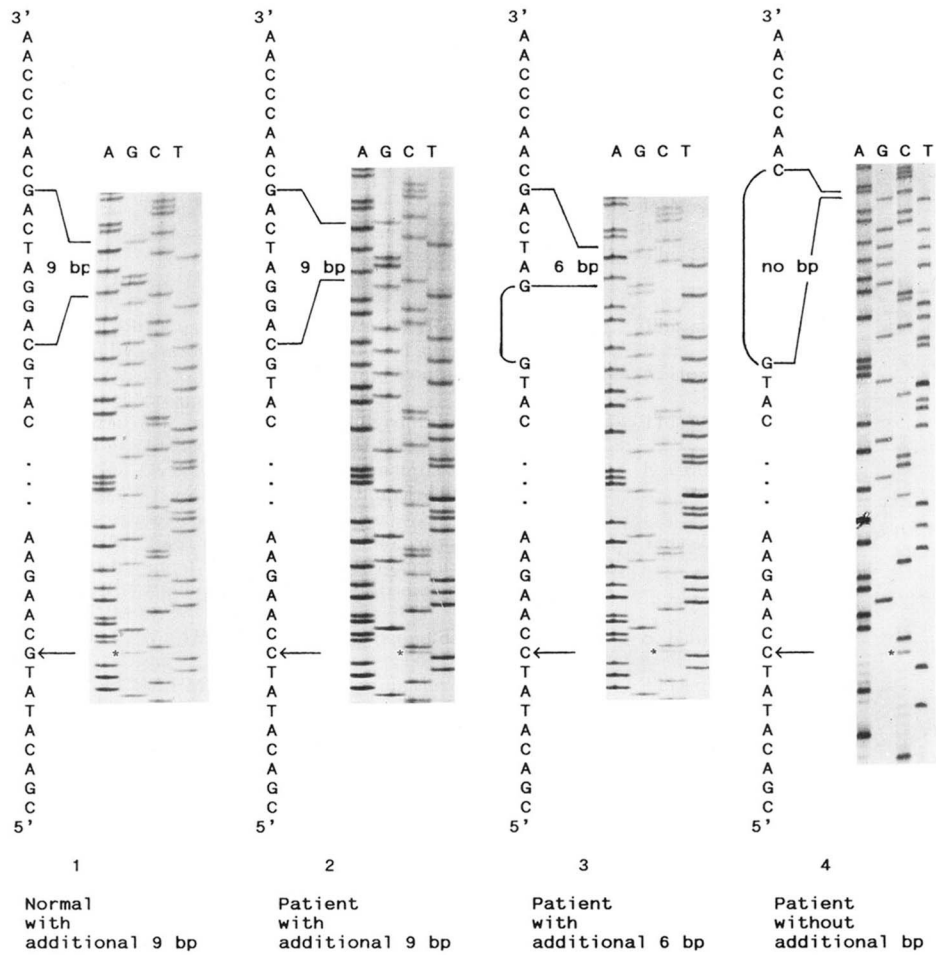

FIG. 2. **Mutation in the patient and the presence of three alternative mRNAs.** The figures represent the coding strand. The mutation at nucleotide 722 (numbered from A of the initiation codon) from the normal G to C in the patient alters the codon 241 from the normal cysteine (TGC) (*sequence 1*) to serine (TCC) (*sequences 2–4*). The *asterisks* and *arrows* denote the location of the base change. This figure also depicts the area representing the 9-base and the 6-base exon. The *left sequence* (*sequence 1*) shows the normal control with an additional 9 bases. The following sequences (*2–4*) show the patient cDNA with an additional 9 bases (*sequence 2*), an additional 6 bases (*sequence 3*), and without additional bases (*sequence 4*).

tion amplifications in three of six sequenced clones. Three clones had the normal sequence in this region. Fig. 3 shows the genomic structure around the respective exon including the mutation. This exon is 57 bp long; the mutated base is the first G of the exon only 2 bases after the 5′ intron-exon junction.

Three alternate forms of mRNA were observed (Fig. 2). Five of the 12 clones sequenced had the largest mRNA containing the same 9-base stretch originally described by Nakano *et al.* (10) in two normal individuals; 2 contained only 6 bases also described by Rafi *et al.* (17) on the 3′ terminus of the 9-base stretch; and 5 had the shortest mRNA that did not contain the 9-base stretch.

To gain insight into the mechanism by which the three alternate forms of mRNA are generated, the genomic segment flanking this region was sequenced (Figs. 4 and 5). It became clear that the stretch of the 9 bases found in the largest mRNA is present in the gene in this region. The stretch also included the additional 24 bases on the 5′ side, comprising the 33-base insertion found by Wenger and colleagues (16) in a patient with sulfatide activator deficiency. The upstream portion of this region is rich in pyrimidines, and the 5′ junctions of both the 9-base and the 6-base stretches have sequences that conform to the consensus sequences for the 3′ acceptor site of an intron-exon junction, CAG CA and CAG GA. At the common 3′ terminus of the 9-base and 6-base stretches there is a typical consensus sequence for the 5′ donor site for an exon-intron junction, CAG GT. Thus, once either of the upstream 3′ acceptor site consensus sequences is recognized as an intron-exon junction the downstream consensus sequence is recognized as the 5′ donor site. This mechanism generates in effect a 6-base or 9-base exon.

DISCUSSION

The stable mRNA of normal size and with a single amino acid substitution within the entire protein coding sequence provides *a priori* evidence that the nucleotide transversion, $G^{722} \rightarrow C$, $Cys^{241} \rightarrow Ser$, could be responsible for the clinical and biochemical phenotype of the sulfatide activator in this patient. Furthermore, all 6 cysteine residues including the one affected by the mutation are strictly conserved among all of the four homologous proteins generated by this gene (10). Sodium dodecyl sulfate-gel electrophoresis of sulfatide activator under nonreducing conditions results in an apparent molecular mass of 16,000 Da and under reducing conditions in an apparent molecular mass of 8,000 Da, suggesting that disulfide bond formation might be functionally critical in these proteins.[6] The failure to form normal disulfide bonds and the resultant disruption of the normal three-dimensional structure are likely to make the mutant sulfatide activator protein unstable, consistent with the lack of cross-reactive material in the patient's tissues. The possibility cannot be completely ruled out, however, that the conformational changes in the mutant sulfatide activator protein make it no longer recognizable by the antibody.

Because of the family history of consanguinity and the consistent finding of the transversion in five successive independent clones we expected the patient to be homozygous

[6] R. Hurwitz and K. Sandhoff, unpublished observations.

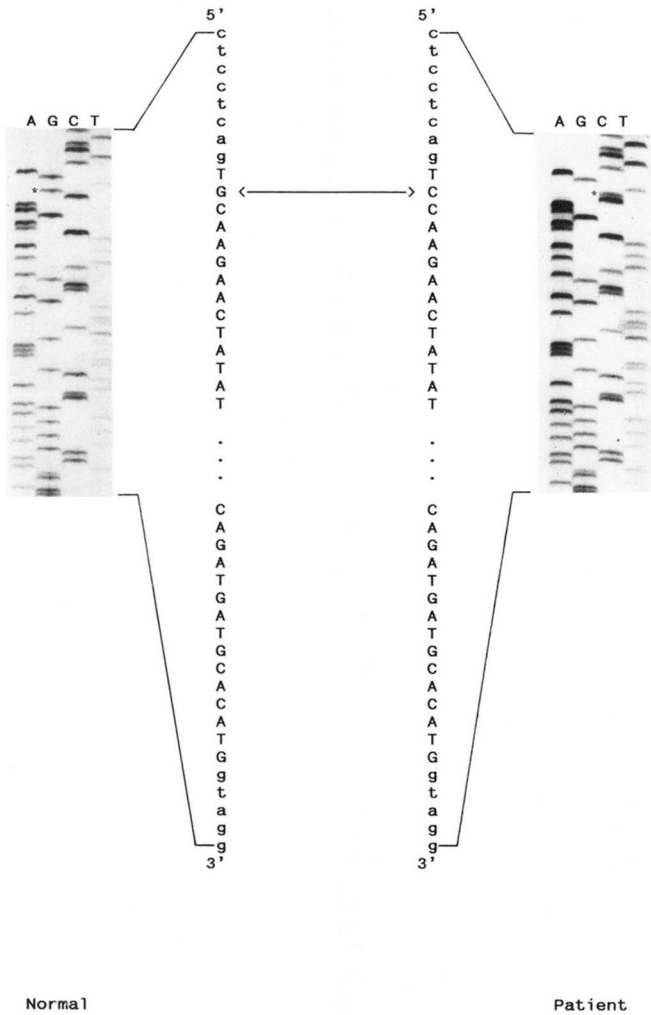

FIG. 3. **Mutation in the patient at the genomic level.** The figure represents the coding strand of the normal control (*left*) and the mutation (*right*). The whole exon including the mutation and parts of the upstream and downstream introns are shown. Amplification and sequencing of the patient's DNA yielded the mutated sequence (three times) as well as the normal sequence (three times). Exon nucleotides are written in *upper case letters*, intron nucleotides in *lower case letters*. The *asterisks* and the *arrows* denote the location of the base change.

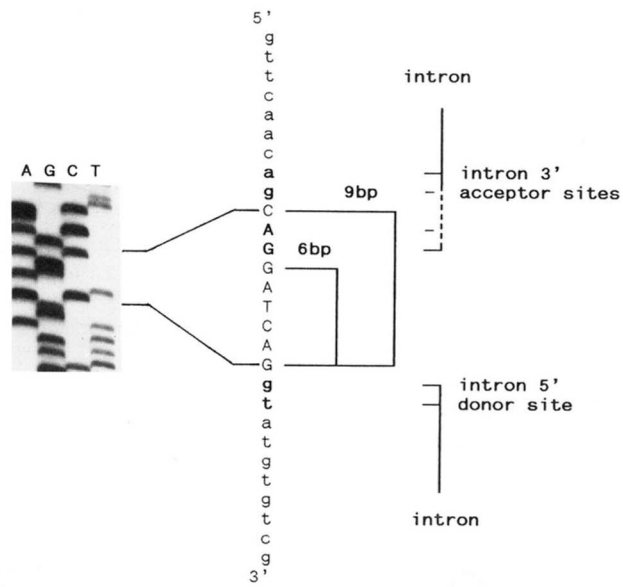

FIG. 4. **The sequence of the genomic structure of a normal individual around the 9-bp exon.** The figure represents the coding strand. Consensus sequences for the 3′ acceptor sites (*ag* and *AG*) and the 5′ donor site (*gt*) exist flanking the exon. Exon nucleotides are written in *upper case letters*, intron nucleotides in *lower case letters*. The two alternative splicing possibilities are indicated.

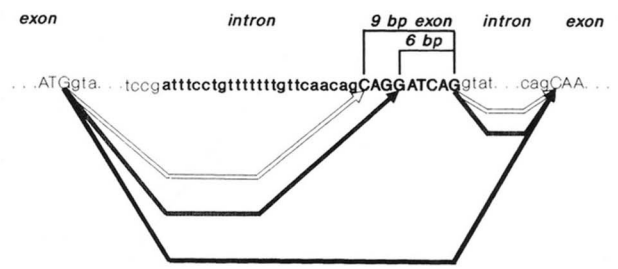

FIG. 5. **The intronic sequence around the observed 9-bp exon.** Consensus sequences for the 3′ acceptor (*ag* and *AG*) and 5′ donor sites (*gt*) exist flanking the exon sequence. Exon nucleotides are written in *upper case letters*, intron nucleotides in *lower case letters*. The insertion of the patient described by Zhang *et al.* (16) is represented as *bold lower case letters*. *Lower solid arrow*, no additionl bp; *upper solid arrows*, 6 additional bp; *open arrow*, 9 additional bp.

for the mutation. Amplification and sequencing of the genomic DNA around the mutation, however, yielded the mutated sequence (three times) as well as the normal sequence (three times) (Fig. 3). This finding suggests that the patient is a compound heterozygote with the identified mutation in one allele and another as yet unidentified mutation in the other allele, which might be of an mRNA-negative type. The mutation found in our patient, $G^{722} \rightarrow C$, $Cys^{241} \rightarrow Ser$, is the third known mutation responsible for sulfatide activator deficiency. The other two are the 33-base insertion at the location of the normal 9-base exon site (16) and a mutation that abolishes the single glycosylation site of the sulfatide activator protein (17, 18).

The finding of three alternative forms of the mRNA for the sulfatide activator precursor which would generate three different forms of the sulfatide activator protein raises two important questions: what is the mechanism that generates the three forms and what is the potential functional significance of the three forms? The observations described in this report clearly answer the first question. An unusual feature in an intronic sequence causes apparently random recognition of the stretches of either 9 or 6 bases as if they were very short exons. When recognized as exons, these stretches of 6 or 9 bases are spliced into the mature mRNA whereas in other cases they are spliced out along with the flanking intron sequences. Furthermore, since the 33-base insertion found by Zhang *et al.* (16) in their patient with sulfatide activator deficiency is also located at the same site as the short exons, one can predict a point mutation immediately upstream of the 33 bases which generates a consensus sequence for a new 3′ acceptor site. We learned recently that Zhang *et al.* (22) has also reached the same conclusion regarding the mechanism that generates the three alternative mRNAs and that his patient has the predicted mutation in the intronic sequence immediately preceding the sequence of the 33-base insertion.

In contrast, there is no experimental basis for answering the question regarding the functionality of the three alternative mRNAs of the precursor of the sulfatide activator protein. The shortest mRNA seems to be functional. It is colinear with the amino acid sequence of the mature sulfatide activator protein (15). Since the additional stretches in the larger mRNAs are in-frame, they will be translated into three dif-

ferent proteins differring from each other only in two or three inserted amino acids. The mRNAs with additional stretches do occur with sufficient frequency so that they too can be of potential functional significance. The sulfatide activator has been reported to stimulate *in vitro* degradation of other glycosphingolipids like G_{M1} ganglioside and globotriaosylceramide. Although it is tempting to speculate that the three alternative forms all have properties of activator proteins with different substrate specificities it is equally possible that only one of these three proteins functions as an activator protein and that the other two have either no, or different, physiological activities. Further experiments are required to clarify this question.

Acknowledgments—We thank Dr. David A. Wenger for allowing us to mention his findings as a personal communication. Joe Langaman, Gaby Weiß, and Andrea Rats contributed to this work with their excellent assistance in cell culture work.

REFERENCES

1. Mehl, E. & Jatzkewitz, H. (1964) *Hoppe-Seyler's Z. Physiol. Chem.* **339,** 260–276
2. Fischer, G. & Jatzkewitz, H. (1975) *Hoppe-Seyler's Z. Physiol. Chem.* **356,** 605–613
3. Li, S.-C. & Li, Y.-T. (1976) *J. Biol. Chem.* **251,** 1159–1163
4. Gärtner, S., Conzelmann, E. & Sandhoff, K. (1983) *J. Biol. Chem.* **258,** 12378–12385
5. Dewji, N. N., Wenger, D. A. & O'Brien, J. S. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 8652–8656
6. Kleinschmidt, T., Christomanou, H. & Braunitzer, G. (1987) *Biol. Chem. Hoppe-Seyler* **368,** 1571–1578
7. Fürst, W., Machleidt, W. & Sandhoff, K. (1988) *Biol. Chem. Hoppe-Seyler* **369,** 317–328
8. O'Brien, J. S., Kretz, K. A., Dewji, N., Wenger, D. A., Esch, F. & Fluharty, A. L. (1988) *Science* **241,** 1098–1101
9. Sano, A., Radin, N. S., Johnson, L. J. & Tarr, G. E. (1988) *J. Biol. Chem.* **263,** 19597–19601
10. Nakano, T., Sandhoff, K., Stümper, J., Christomanou, H. & Suzuki, K. (1989) *J. Biochem.* **105,** 152–154
11. Ho, M. W. & O'Brien, J. S. (1971) *Proc. Natl. Acad. Sci. U. S. A.* **68,** 2810–2813
12. Christomanou, H., Aignesberger, A. & Linke, R. P. (1986) *Biol. Chem. Hoppe-Seyler* **367,** 879–890
13. Morimoto, S., Martin, B., Kishimoto, Y. & O'Brien, J. S. (1988) *Biochem. Biophys. Res. Commun.* **156,** 403–410
14. Morimoto, S., Yamamoto, Y., O'Brien, J. S. & Kishimoto, Y. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 3493–3497
15. Fürst, W., Schubert, J., Machleidt, W., Meier, H. E. & Sandhoff, K. (1990) *Eur. J. Biochem.* **192,** 709–714
16. Zhang, X., Rafi, M. A., DeGala, G. & Wenger, D. A. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 1426–1430
17. Rafi, M. A., Zhang, X.-L., DeGala, G. & Wenger, D. A. (1990) *Biochem. Biophys. Res. Commun.* **166,** 1017–1023
18. Kretz, A. K., Carson, G. S., Morimoto, S., Kishimoto, Y., Fluharty, L. A. & O'Brien, J. S. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 2541–2544
19. Paton, B. C., Hughes, J. L., Harzer, K. & Poulos, A. (1990) *Eur. J. Cell Biol.* **51,** 157–164
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp. 196–203, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
22. Wenger, D. A. (1990) *Am. J. Hum. Genet.* **47,** A172