Volume 280, number 2, 267–270 FEBS 09508 March 1991
© 1991 Federation of European Biochemical Societies 0014-5793/91/$3.50
ADONIS 001457939100234L

The organization of the gene for the human cerebroside sulfate activator protein

Hans Holtschmidt[1], Konrad Sandhoff[1], Werner Fürst[1], Hae Young Kwon[1], Doris Schnabel[1] and Kunihiko Suzuki[2]

[1] *Institut für Organische Chemie und Biochemie der Universität Bonn, D-5300 Bonn 1, Germany* and [2] *Brain and Development Research Center, Departments of Neurology and Psychiatry, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA*

Received 27 December 1990

The organization of 14 exons covering 97% of the cDNA sequence of human cerebroside sulfate activator protein precursor has been determined from two overlapping EMBL-4 human genomic clones extending over 17 kb. All exons and exon/intron splice junctions and five introns were sequenced. Exon 8 consists of only 9 bp and is involved in alternative splicing which generates three different mRNAs of cerebroside sulfate activator precursor.

Cerebroside sulfate activator protein; Genomic structure; Alternative splicing

1. INTRODUCTION

The cerebroside sulfate activator protein (CSAP) stimulates the lysosomal break-down of cerebroside sulfate by arylsulfatase A [1]. It is synthesized as a 65 kDa precursor protein [2] which is proteolytically processed to yield four small homologous sphingolipid activator proteins [3–5]. Alternative cDNAs were observed differing from the one corresponding to the protein sequence by insertions of 9 or 6 bp, respectively, into the cerebroside sulfate activator domain [5], (H. Holtschmidt and K. Sandhoff, unpublished). Here we present the organization of the CSAP gene and offer an explanation for the occurence of the different cDNAs.

2. MATERIALS AND METHODS

2.1. *Isolation of genomic clones*

Two genomic libraries from human kidney and brain tissue in lambda EMBL-4 phages were constructed by the method of Frischauf [6]. 2.3×10^6 phages were screened by hybridization with the cDNA of CSAP [5] which was labeled before with ^{32}P by random priming using a kit from Pharmacia (Freiburg, Germany). Two genomic clones were isolated as described by Maniatis et al. [7].

Correspondence address: K. Sandhoff, Institut für Organische Chemie und Biochemie, Gerhard-Domagk Str. 1, D-5300 Bonn 1, Germany

Abbreviations: CSAP, cerebroside sulfate activator protein

2.2. *Gene mapping and DNA sequencing*

Lambda clones were cleaved by single and double digestions with the corresponding restriction enzymes and the fragments were size-fractionated by agarose gel electrophoresis. Appropriate fragments were extracted from the gel by the glasmilk method (Biogen, LaJolla, CA, USA) and were subcloned to pUC18. Plasmids were prepared with the Quiagen plasmid prep-kit (Diagen, Düsseldorf, Germany). Sequencing was done by the dideoxy chain termination method [8] with the kit from Pharmacia using [α-^{35}S]dATP (22 TBq/mmol, Amersham), T_7-polymerase and a commercial sequencing primer or synthetic oligonucleotides as primers.

3. RESULTS AND DISCUSSION

Two EMBL-4 human genomic libraries were screened with the full-length cDNA of the CSAP precursor described by Nakano et al. [5]. Two overlapping genomic clones, SAP no. 1 and SAP no. 2, were isolated encompassing 17 kb of the genomic sequence and were characterized by restriction mapping (Fig. 1). They cover 14 exons (no. 2 – no. 15) with more than 97% of the cDNA sequence. Only the region coding for the first 13 amino acids of the signal peptide (putative exon +1) and the 5′-untranslated sequence are not represented by these clones and were not yet discovered. About 9.9 kb of the genomic clones including all exons and exon/intron junctions were sequenced, 6.3 kb in both directions. The exon/intron organization of the CSAP gene was deduced from the comparison of the cDNA sequence with the genomic sequence. The positions of the introns show some regularities. The CSAP precursor contains a signal peptide and four homologous sphingolipid activator protein domains.

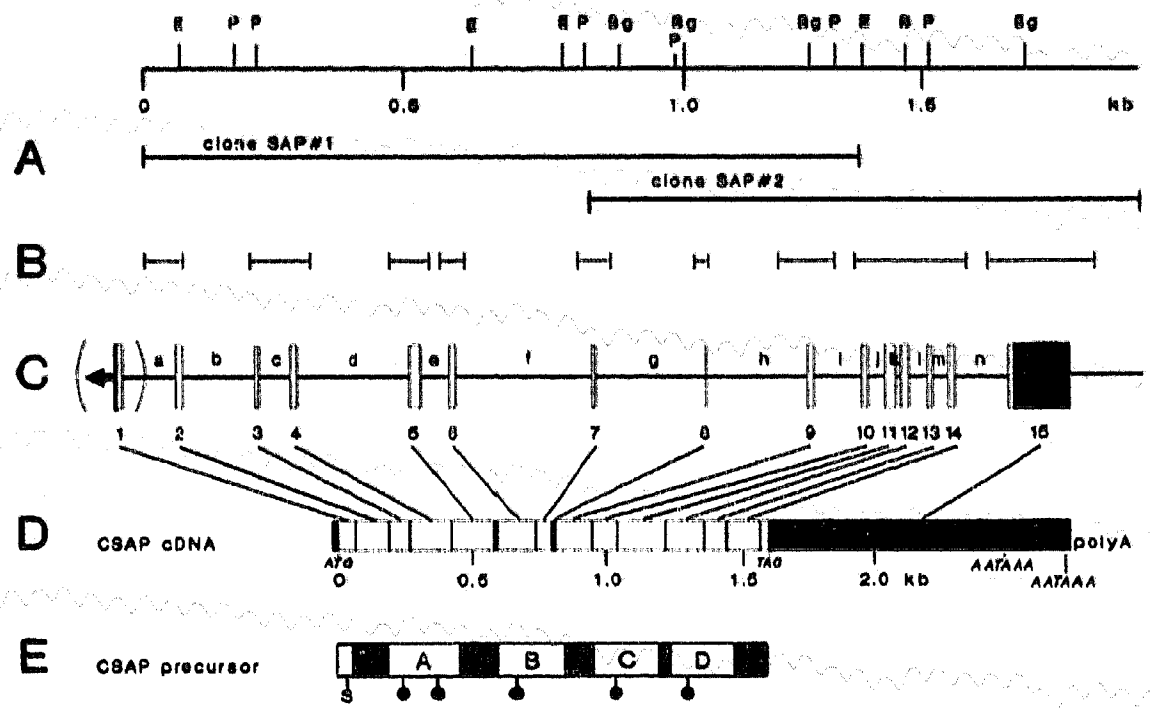

Fig. 1. Map of the CSAP gene. By screening with the cDNA of CSAP the genomic clones SAP no. 1 and SAP no. 2 were isolated from EMBL-4 human genomic libraries. (A) Restriction map of clones SAP no. 1 and SAP no. 2 (B = *Bam*HI, Bg = *Bgl*II, E = *Eco*RI, P = *Pst*I). Exon containing fragments were subcloned to pUC18 and sequenced. (B) Sequenced regions. (C) Exon/intron organization of CSAP gene. Open boxes correspond to exons no. 2 to no. 15 covering the cDNA sequence from the codon for Ala^{14} to the end. Black areas correspond to untranslated regions. The putative exon no. 1 (in brackets) which should cover the missing 5'-end of the gene has not been found yet. Introns are labeled with lower case letters. (D) cDNA of CSAP. Exons and untranslated regions are indicated as in (B). (E) CSAP precursor (527 amino acids). The signal peptide is marked with an S. Domains A-D correspond to the mature activator proteins: A, saposin A [11], B, CSAP [10], C, glucocerebrosidase activator protein [12], D, component C [3]. *N*-glycosylation sites are marked by the points.

Introns b, e, i and k are found just in front of domains A, B, C and D, respectively. Introns d and g, c and j as well as f and l are located at homologous positions of domains A and B, A and C and B and D, respectively. These observations confirm the assumption that the gene of this precursor arose by sequential duplications of an ancestral activator protein gene. All intron sequences comply with the gt/ag rule (Table I) and are in good agreement with the consensus sequences of splice junctions [9]. Introns c, j, k, l and m have been completely sequenced (Fig. 2).

So far, the amino acids encoded by the very short exon 8 (9 bp) were not found in the mature protein [10]. However, in addition to the cDNA corresponding to the protein sequence, other cDNAs were observed including those with nine [5] or six (H. Holtschmidt and K. Sandhoff, unpublished) additional bases. Here we demonstrate exon 8 as the source of these additional bases. The three alternative mRNAs observed may be generated by splicing exon 7 to exon 8, to the last six bases of exon 8 (exon 8) or to exon 9, thereby skipping exon 8 (Table I).

Acknowledgements: This work was supported by grants from the Deutsche Forschungsgemeinschaft.

REFERENCES

[1] Fischer, G. and Jatzkewitz, H. (1975) Hoppe-Seylers Z. Physiol. Chem. 339, 260–276.
[2] Fujibayashi, S. and Wenger, D.A. (1986) Biochim. Biophys. Acta. 875, 554–562.
[3] Fürst, W., Machleidt, W. and Sandhoff, K. (1988) Biol. Chem. Hoppe-Seyler 369, 317–328.
[4] OBrien, J.S., Kretz, K.A., Dewji, N., Wenger, D.A., Esch, F. and Fluharty, A.L. (1988) Science 241, 1098–1101.
[5] Nakano, T., Sandhoff, K., Stümper, J., Christomanou, H. and Suzuki, K. (1989) J. Biochem. (Tokyo) 105, 152–154.
[6] Frischauf, A.-M., Lehrach, H., Poustka and Murray, N. (1983) J. Mol. Biol. 170, 827–831.
[7] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[8] Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[9] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.
[10] Fürst, W., Schubert, J., Machleidt, W., Meyer, H.E. and Sandhoff, K. (1990) Eur. J. Biochem. 192, 709–714.

Table I

Exon/intron boundarie of the CSAP gene. Exon nucleotides are written in upper case letters, intron nucleotides in lower case. The numbers below the exon sequences refer to 5' and 3' nucleotide positions in the CSAP cDNA flanking each intron (1 = A of the initiation codon ATG). Below these numbers are given the amino acids in the CSAP precursor interrupted by, or flanking, each intron. Alternative forms of the mRNA are generated by splicing exon 7 either to exon 8, or to the last six bases of exon 8 (exon 8') or directly to exon 9.

Exon no.	Exon size bp	Sequence of intron exon junctions 5' Border			Intron (no.)		3' Border				Intron size kb
1	>48	GGC	GCG	G 39 A(Ia)[14]	not analaysed	(a) ...	ggtttttcattttag	CT 40 (A)Ia[14]	CTA	GCC	>0.8
2	134	CCA	ACA	GTG 174 Val[58]	gtgagtgcc	(b) ...	tattctttctcttag	AAA 175 Lys[58]	TCC	CTT	1.4
3	75	GCC	ACT	GAG Glu[58]	gtgagcggc	(c) ...	ggtgtcctctcgtag	GAG 250 Glu[84]	GAG	ATC	0.6
4	126	GGA	GAA	ATG 375 Met[125]	gtaagtgat	(d) ...	tttccatcccaccag	AGC 376 Ser[125]	CGT	CCT	2.1
5	201	CAG	CCA	AAG 576 Lys[192]	gtaagacaa	(e) ...	gcctctcttctgtag	GAT 577 Asp[193]	AAT	GGG	0.5
6	144	GCC	GAC	ATA 720 Ile[240]	gtgagcctt	(f) ...	tttctctctcctcag	TGC 721 Cys[241]	AAG	AAC	2.6
7	57	ATG	CAC	ATG 777 Met[259]	gtaggtggc	(g) ...	tttttttgttcaacag	CAG 778 Gln[260]	GAT	CAG	2.1
					or:	(g') ...	tttttttgttcaacagcag		GAT 781 ASP[261]	CAG	2.1
8 or 8'	9 6	CAG	GAT GAT	CAG CAG 786 Gln[262]	gtatgtgtc	(h) ...	ctgtgctctcttcag	CAA 787 Gln[263]	CCC	AAG	1.9
9	132	CCC	ATT	AAG 918 Lys[306]	gtacctggt	(i) ...	ctactctttccacag	AAG 919 Lys[307]	CAC	GAG	0.9
10	96	AAG	ACT	GAG 1014 Glu[338]	gtatgctgt	(j) ...	tctcgtgtgtttcag	AAA 1015 Lys[339]	GAA	ATA	0.35
11	187	CTG	ACC	G 1201 V(al)[401]	gtgagcgct	(k) ...	tgccctcttatgtag	TT 1202 (V)al[401]	CAC	GTG	0.09
12	158	CAG	AAG	CAG 1359 Gln[453]	gtacgcccc	(l) ...	cactttcctttatag	TGT 1360 Cys[454]	GAT	CAG	0.35
13	81	GTG	TGC	TTG 1440 Leu[480]	gtgagcctc	(m) ...	tttctcaccatatag	AAA 1441 Lys[481]	ATT	GGA	0.3
14	108	CAG	TGC	AAT 1548 Asn[510]	gtgagtagc	(n) ...	cctctctcctaccag	GCT 1549 Ala[517]	GTC	GAG	1.0
15	1182										

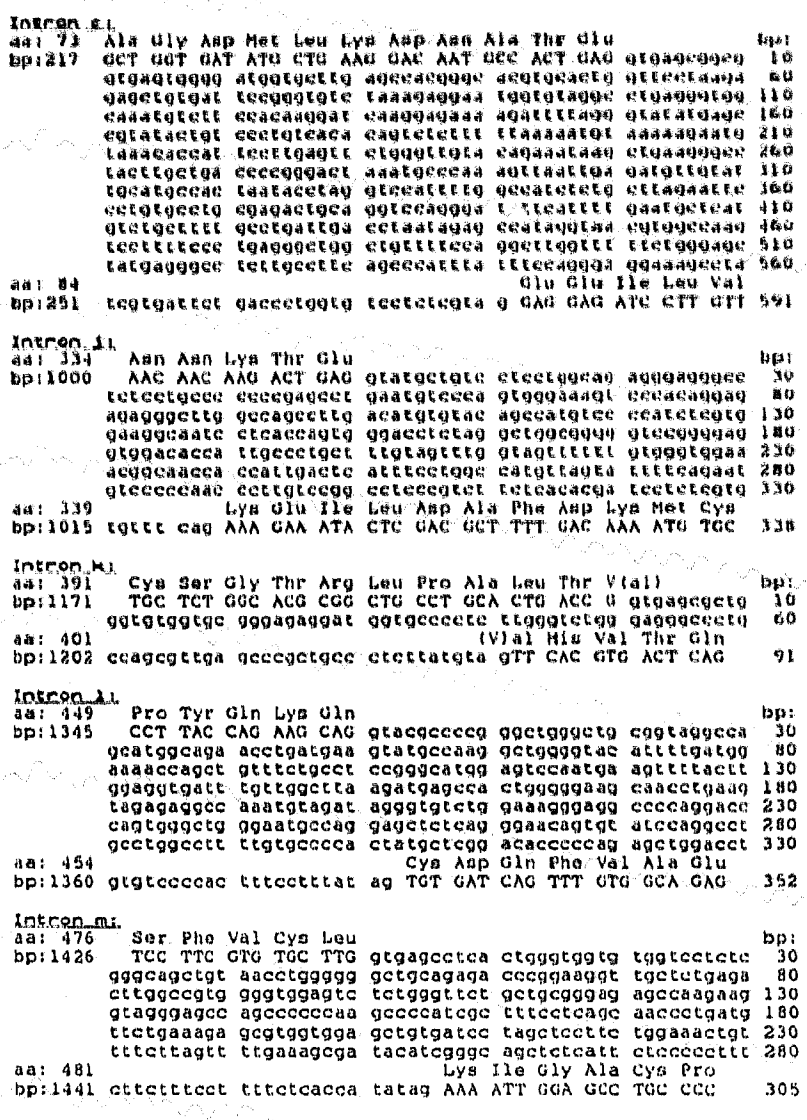

Fig. 2. Nucleotide sequences of introns c, j, k, l and m of the CSAP gene. Exon nucleotide numbers corresponding to the cDNA sequence (1 = A of the initiation codon ATG) and amino acid numbers are given on the left side, intron nucleotide numbers on the right side. Exon nucleotides are written in upper case letters, intron nucleotides in lower case.

[11] Morimoto, S., Martin, B.M., Yamamoto, Y., Kretz, K.A., O'Brien, J.S. and Kishimoto, Y. (1989) Proc. Natl. Acad. Sci. USA 86, 3389–3393.

[12] Kleinschmidt, T., Christomanou, H. and Braunitzer, G.(1987) Biol. Chem. Hoppe-Seyler 368, 1571–1578.