Genomic Organization and Expression of the Human Fatty Aldehyde Dehydrogenase Gene (FALDH)

GERALDINE R. ROGERS,* NEDIALKA G. MARKOVA,† VINCENZO DE LAURENZI,†,‡
WILLIAM B. RIZZO,§ AND JOHN G. COMPTON*,1

*Genetic Studies Section, †Laboratory of Skin Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland 20892; ‡Istituto Dermopatico Dell' Immacolata (IDI-IRCCS), Biochemistry Laboratory at Department of Experimental Medicine, University Tor Vergata, Rome 00167, Italy; and §Departments of Pediatrics and Human Genetics, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298

Mutations in the fatty aldehyde dehydrogenase (FALDH) gene cause Sjögren-Larsson syndrome (SLS)— a disease characterized by mental retardation, spasticity, and congenital ichthyosis. To facilitate mutation analysis in SLS and to study the pathogenesis of FALDH deficiency, we have determined the structural organization and characterized expression of the FALDH (proposed designation ALDH10) gene. The gene consists of 10 exons spanning about 30.5 kb. A TATA-less promoter is associated with the major transcription initiation site found to be 258 bp upstream of the ATG codon. The GC-rich sequences surrounding the transcription initiation site encompassed regulatory elements that interacted with proteins in HeLa nuclear extracts and were able to promote transcription *in vitro.* FALDH is widely expressed as three transcripts of 2, 3.8, and 4.0 kb, which originate from multiple polyadenylation signals in the 3′ UTR. An alternatively spliced mRNA was detected that contains an extra exon and encodes an enzyme that is likely to have altered membrane-binding properties. The FALDH gene lies only 50–85 kb from ALDH3, an aldehyde dehydrogenase gene that has homologous sequence and intron/exon structure. © 1997 Academic Press

INTRODUCTION

Aldehyde dehydrogenases (ALDHs) compose a group of isozymes with the general role of catalyzing the oxidation of a wide variety of aldehydes to their corresponding carboxylic acids. Mammalian ALDHs have been broadly classified as Class 1, 2, or 3 based on subcellular distribution, substrate specificity, and primary sequence similarity (reviewed in Lindahl, 1992).

Fatty aldehyde dehydrogenase (FALDH), a component of the fatty alcohol:NAD$^+$ oxidoreductase complex, catalyzes the oxidation of saturated or unsaturated aliphatic aldehydes of medium- or long-chain length to fatty acids (Kelson *et al.,* in press). Based on its role in lipid metabolism and microsomal localization, it can be classified as a Class 3 ALDH (Lindahl and Petersen, 1991; Kelson *et al.,* in press).

FALDH enzyme deficiency is associated with Sjögren–Larsson syndrome (SLS). SLS is a rare, autosomal recessive disorder with symptoms of congenital, generalized ichthyosis, mental retardation, and spastic diplegia or tetraplegia (Sjögren and Larsson, 1957; reviewed by Rizzo, 1993). A definitive diagnosis of SLS can be made by measuring the activity of the enzyme FALDH in cultured fibroblasts. All SLS patients have profoundly deficient enzyme activity, and their parents and other carriers have intermediate levels of FALDH activity compared to unaffected controls (Rizzo and Craft, 1991; Kelson *et al.,* 1992). Three linkage studies in families of diverse ethnic origins have all mapped the SLS locus to 17p11.2 (Pigg *et al.,* 1994; Rogers *et al.,* 1995; Lacour *et al.,* 1996), thus demonstrating that SLS is a genetically homogeneous disorder. The molecular basis for FALDH deficiency has recently been attributed to homozygous or compound heterozygous mutations in the FALDH gene at 17p11.2 (De Laurenzi *et al.,* 1996).

The cDNA for human FALDH (ALDH10)[2] was cloned with the intent of determining the genetic basis of SLS by identifying FALDH mutations in patients with SLS (De Laurenzi *et al.,* 1996). In screening for mutations we cultured fibroblasts from SLS patients and analyzed their FALDH mRNA sequence by reverse transcription and direct sequencing of the PCR-amplified cDNA. Although this method was successfully used to identify

[2] A convention for naming new ALDH genes in the order they are identified is under discussion. In this nomenclature, the gene encoding FALDH would be designated ALDH10.

the mutations in several patients, it is desirable to develop a DNA-based mutation detection technique to facilitate future analyses. Cell culture is time consuming and not always possible, whereas genomic DNA is readily available from peripheral blood or buccal swabs. Additionally, some disease-causing mutations cannot be readily detected at the RNA level, and deletions/splice variants should be confirmed at the DNA level. In this study we have determined the genomic organization of the FALDH gene and obtained intron sequences that will enable genomic DNA to be used to confirm and/or identify mutations. FALDH was shown to be widely expressed, and we observed differential expression of multiple transcripts between and within tissues. An FALDH variant with additional carboxy-terminal residues was predicted from an alternatively spliced transcript. We also demonstrated that the FALDH gene lies only 50–85 kb from another aldehyde dehydrogenase gene, ALDH3, at 17p11.2. It seems likely that the FALDH and ALDH3 genes have arisen from a common ancestral gene, given their sequence similarity, conservation of gene structure, and physical proximity.

MATERIALS AND METHODS

PCR analysis of the FALDH gene. Primer sequences (5′ to 3′) for amplification of introns from the FALDH gene (identified by sequence position in the FALDH cDNA sequence; GenBank Accession No. L47162) with the sense primer given first in each set were (numbering relative to translation start site) 5′ UTR (−43 to −23)/(95 to 70); int 1 (82 to 104)/(315 to 290); int 2 (310 to 335)/(448 to 415); int 3 (436 to 469)/(615 to 586); int 4 (587 to 614)/(721 to 689); int 5 (689 to 721)/(927 to 897); int 6 (898 to 932)/(1096 to 1063); int 7 (1063 to 1095)/(1194 to 1171); int 8 (1121 to 1154)/(1286 to 1259); int 9 (1397 to 1428)/(1601 to 1566). Standard PCRs in a 30-μl volume contained 2 μl normal human buccal DNA (Richards *et al.,* 1993), 200 nM each primer, 200 μM dNTPs, and 1 U AmpliTaq DNA polymerase (Perkin–Elmer) with the supplied manufacturer's buffer (1.5 mM MgCl$_2$) and were thermal cycled at 95°C/1 min, 60°C/1 min, 72°C/1.5 min for 35 cycles followed by an extension at 72°C for 10 min. The XL PCR kit (Perkin–Elmer) for amplification of longer fragments was used according to the manufacturer's protocol. The sizes of PCR products were determined by comparison with standard molecular weight markers on agarose gels. Each PCR product was gel-purified and sequenced from both ends with the fmol Cycle Sequencing kit (Promega Corp.) to determine the location of the exon/intron junction. The exon sequence was checked against the FALDH cDNA sequence and was found to match exactly in all PCR products.

Primer extension. The AMV Reverse Transcriptase Primer Extension system (Promega) was used to determine the transcription initiation site in the FALDH gene. Briefly, 100 fmol of an antisense primer (−20 to −41) was 5′ end-labeled with [γ-^{32}P]ATP (DuPont) using T4 polynucleotide kinase, annealed to 200 ng poly(A)$^+$ liver RNA (Clontech) at 58°C for 20 min, and then placed at room temperature for 10 min. Annealed primers were extended with AMV reverse transcriptase at 42°C for 30 min.

RNase protection assay. A total of 1590 nucleotides of the 5′-flanking region of the FALDH gene were obtained by walking upstream from an antisense primer (−20 to −41) with the PromoterFinder DNA Walking kit (Clontech). The product was cloned into the pCRII vector (TA Cloning kit, Invitrogen) and sequenced with the fmol Cycle Sequencing kit (Promega Corp.). A 454-bp PCR product was amplified with primers (−359 to −334)/(95 to 70), cloned

into the pCRII vector, and used to produce an antisense riboprobe for the RNase protection assay (RPA). RPAs were performed with the RPA II Kit (Ambion) according to the manufacturer's protocol using 500 ng poly(A)$^+$ liver RNA (Clontech).

Band-shift assays. Band-shift experiments to determine the binding interactions of a putative Sp1 binding motif located 51 bp upstream of the transcription initiation site and an initiator region were performed as described by Jang *et al.* (1996).

In vitro transcription. The plasmid described above for riboprobe production was transcribed *in vitro* with the *Drosophila* Embryo Nuclear Extract *in vitro* Transcription system (Promega) using the manufacturer's protocol. An antisense primer residing at (−125 to −144) was used for primer extension analysis of the newly transcribed RNA.

Northern analysis and RT-PCR. A 255-bp 3′ UTR probe was prepared from PCR product amplified with a sense intron 9 primer (5′-AGGTGTCTGAGTCCCAGTGCTGAGC-3′) and an antisense 3′ UTR primer (1601 to 1566). The probe was radioactively labeled with [α-^{32}P]dCTP (3000 Ci/mmol) using the Prime-It RmT Random Primer Labeling kit (Stratagene) and hybridized with Northern blots using ExpressHyb solution (Clontech). Bands were visualized with autoradiography. RT-PCR was performed with the rTth Reverse Transcriptase RNA PCR kit (Perkin–Elmer) according to the manufacturer's protocol.

3′ RACE. 3′ RACE was performed with the Marathon cDNA Amplification kit (Clontech). Template was produced according to the supplier's protocol from 1 μg poly(A)$^+$ liver RNA (Clontech), and a sense primer located in exon 9 (1397 to 1428) was used as the gene-specific primer. The 3′ RACE products were cloned into the pCRII vector and sequenced from both ends using the M13 Forward Primer and Sp6 Promoter Primer contained within the vector sequence. The same methods were used to determine the 3′ end of the alternatively spliced product containing exon 9′ with the sense primer 5′-GAACACCAGATTTCAAAACCCAGCCCTGTCTG-3′.

YAC analysis. YACs containing DNA from 17p11.2 were obtained from Research Genetics Inc., and yeast chromosomal DNA plugs were prepared in Bio-Rad Sample Plug Molds according to the manufacturer's protocols. Plug DNAs were used as templates in PCR analyses (1 μl melted plug per 30 μl PCR volume) to test for the presence of the FALDH gene (primers (310 to 335)/(448 to 415)) and the ALDH3 gene (primers (5′-GCTTATGGGGGCACCGGGGATGCC-3′)/(5′-TGGGGGTCCACGTCCGTGAGGATG-3′)). Standard PCRs were performed as described previously. For Southern analysis the YAC DNAs (undigested or endonuclease-digested) were separated by PFGE on the Bio-Rad CHEF Mapper and blotted onto Zeta-Probe membrane (Bio-Rad) according to provided protocols. Probes for FALDH and ALDH3 were prepared from the above PCR products, and the 340-bp probe spanning FALDH exons 4 through 7 was prepared from an RT-PCR fragment amplified with primers (587 to 614)/(927 to 897). The probes were radioactively labeled with [α-^{32}P]dCTP (3000 Ci/mmol) using the Prime-It RmT Random Primer Labeling kit and hybridized with the blotted YAC DNA in ExpressHyb solution.

RESULTS

Exon/Intron Organization of the FALDH Gene

Alignments of the FALDH and ALDH3 coding sequences revealed a striking degree of identity (66%, excluding the final 35 codons unique to FALDH) even though the enzymes differ in their substrate specificity and subcellular compartmentalization. In view of this evolutionary conservation, we used the known gene structure of ALDH3 (Hsu *et al.,* 1992) to predict the positions of the exon/intron junctions in the FALDH gene and designed PCR primer pairs to amplify across

TABLE 1

Exon/Intron Organization of the Human FALDH Gene

Exon (bp)	5′ splice junction	Intron (kb)	3′ splice junction	Exon
1 (411)	TGC AAG **gt**anca. . . .	1 (~2.4)	. . . .taac**ag** [154]AGT GAA	2
2 (232)	GCA G **gt**ctgg. . . .	2 (~0.9)	. . . .attt**ag** [385]GA AAT	3
3 (86)	GAC CAG **gt**aaga. . . .	3 (~4.2)	. . . .tttg**ag** [471]GAT CTC	4
4 (209)	TGC AG **gt**gagt. . . .	4 (~1.3)	. . . .cctc**ag** [680]A CGC	5
5 (118)	GTG AAG **gt**ttgt. . . .	5 (~3.3)	. . . .aaat**ag** [798]GAA TTT	6
6 (142)	ATA G **gt**aatg. . . .	6 (~2.4)	. . . .tttt**ag** [940]CC CCA	7
7 (167)	CAT AAG **gt**aagc. . . .	7 (~1.5)	. . . .cctc**ag** [1107]CTC ATC	8
8 (100)	GTG G **gt**gagt. . . .	8 (~7.0)	. . . .tttc**ag** [1207]GT TCC	9
9 (236)	GTC AAG **gt**gagtc. . . .	9 (~3.8)	. . . .ctcc**ag** [1443]GCA GAA	10

Note. Exon sequence is shown in the reading frame in uppercase letters; intron sequence is shown in lower case letters. Superscript numbers refer to nucleotide position in the cDNA sequence (GenBank L47162). Size of exon 1 includes 5′ UTR. Exon 10 is 182 bp including shortest 3′ UTR. The genomic sequence of FALDH with partial intron sequence can be found in GenBank, entries U75286 to U75297.

putative introns in the FALDH gene. The standard PCR method was employed to amplify over the 5′ UTR and introns 2, 4, and 7 successfully, whereas introns 1, 3, 5, 6, and 9 required an amplification protocol for longer fragments. Putative intron 8 could not be amplified with either protocol. Instead, its exon/intron junctions were determined by amplifying from both putative flanking exons into intron 8 from a cloned library (PromoterFinder DNA Walking kit (Clontech)).

Like the ALDH3 gene, the coding sequence of the FALDH gene is interrupted by nine introns (Table 1). The first eight introns occur at the precise nucleotide positions predicted from the alignment with the ALDH3 gene sequence, confirming the postulated conservation of gene structure between the ALDH3 and the FALDH genes. Exon 9 of FALDH is larger than exon 9 of ALDH3 (236 and 131 bp, respectively) owing to the presence of unique sequences in FALDH that encode the hydrophobic carboxy-terminal domain. This domain is important for targeting the enzyme to the microsomal membrane (Masaki *et al.,* 1994). However, the 3′ boundary of intron 9 remains conserved since exon 10 of both genes carries the sequence for the carboxy-terminal four residues and the 3′ UTR. All FALDH introns (0.85 to 7 kb) were larger than the corresponding ALDH3 introns. The calculated size of the FALDH gene, excluding the 5′ and 3′ UTRs, is approximately 28.2 kb compared to 7.2 kb for the same region of the ALDH3 gene. All exon/intron boundaries conformed to the GT/AG consensus for intronic donor and acceptor splice signals, respectively (Shapiro and Senapathy, 1987). Exon sequences and partial FALDH intron sequences, along with 5′ UTR and 3′ UTR sequences, have been deposited in GenBank (U75286 to U75297). This information can be used to design primer sets for the amplification of individual FALDH exons for mutation analysis.

Determination of Transcription Start Site

The FALDH cDNA clone described by De Laurenzi *et al.* (1996) contained 163 nucleotides of 5′ UTR. However, primer extension with an antisense primer (−20 to −41) produced a major band of 238 nucleotides corresponding to transcription initiation occurring at position −258 (Fig. 1a). Since the transcription initiation site clearly lay upstream of the sequences contained in the cDNA clone, we cloned and sequenced a further 1447 nucleotides of genomic 5′-flanking sequence. To confirm the start position of transcription, we performed an RNase protection assay using an antisense riboprobe derived from this clone (Fig. 1b). From the single protected RNA fragment of approximately 350 nucleotides, we deduced that approximately 255 nucleotides of 5′ UTR sequence had been protected. This was in good agreement with the length of the primer extension product and showed that the 5′ UTR was not interrupted by intronic sequences.

Analysis of the sequence around −258 revealed no TATA box. Therefore, we tested whether the sequence immediately around the putative transcription start site could promote faithful transcription *in vitro.* A plasmid containing 101 nucleotides of sequence upstream from the transcription start site at −258 and 352 nucleotides downstream was subjected to *in vitro* transcription in a *Drosophila* embryo nuclear extract. The newly transcribed RNA was used as a template in primer extension analysis with an FALDH-specific primer located at −125 to −144. As shown in Fig. 1c, a major product of the expected length of 133 nucleotides was obtained, indicating that transcription had initiated at −258 in this *in vitro* system.

The region around the site of transcription initiation contained multiple CpG islands that are characteristic of a number of TATA-less promoters (Fig. 1d). Consensus Sp1 DNA binding motifs were observed at positions −2, −51, −134, −192, and −480 relative to the transcription start site. Furthermore, a sequence conforming to the consensus for one group of initiators—5′-YYCAYYYY-3′ where Y is pyrimidine (Azizkhan *et al.,* 1993)—was located on the complementary strand 13 nucleotides downstream of the transcription start site (5′-CTCACCCCT-3′; Fig. 1d). We used gel-shift
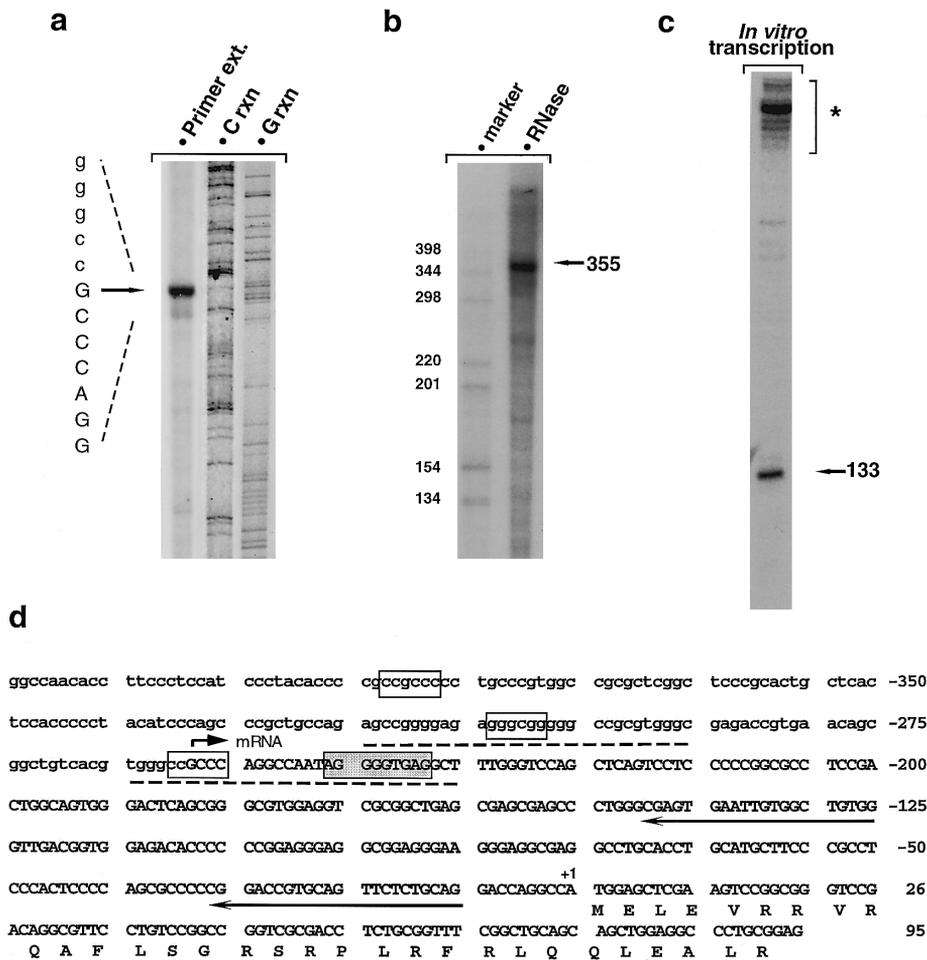
130 ROGERS ET AL.

FIG. 1. Determination of the transcription start site in the FALDH gene. (**a**) Primer extension of liver poly(A)$^+$ RNA with an antisense oligonucleotide located at −20 to −41 of the FALDH gene. The primer extension product is shown next to the C and G reaction sequencing ladders of the corresponding region in the FALDH gene from the same antisense primer. The sequence of the sense strand is given. (**b**) RNase protection assay of liver poly(A)$^+$ RNA with an antisense riboprobe derived from the FALDH gene. The RNase protection product was sized by comparison with a 5′-end-labeled molecular weight marker (nucleotide sizes shown). (**c**) *In vitro* transcription of a plasmid containing the FALDH sequence 101 nucletotides upstream and 353 nucleotides downstream of the apparent transcription initiation site. Primer extension was performed with an antisense primer located at −125 to −144. The asterisk indicates nonspecific products. (**d**) Nucleotide sequence of the 5′-flanking region and partial exon 1 of the FALDH gene. The first nucleotide of the translation initiation codon is numbered as 1. The transcription start site is arrowed, and transcribed sequence is capitalized. Consensus Sp1 binding sites are shown in open boxes, while the putative initiator sequence is shown in the shaded box. Antisense primers used for the primer extensions are shown with a solid underline, while the oligonucleotides used in band-shift assays are depicted with a broken underline.

analysis to make a preliminary assessment of whether two of the protein-recognition motifs around the FALDH transcription start site were indeed sites of protein–DNA binding, interactions that would be consistent with those of similar motifs found to be important in TATA-less promoters. Double-stranded oligonucleotide probes (Fig. 2a) corresponding to the Sp1 motif at position −51 (relative to the transcription start site) and to the initiator region at position +13 (relative to the transcription start site) were each radioactively labeled and used in protein binding reactions (Figs. 2b and 2c). The Sp1 oligonucleotide (Fig. 2b, lane 1) interacted strongly with recombinant Sp1 protein to produce a retarded complex (Fig. 2b, lane 2). A complex with similar mobility was formed when the probe was combined with a nuclear extract derived from HeLa cells in which the FALDH gene is actively transcribed (Fig. 2b, lane 6). The specificity of the bound complex was confirmed in competitions with a 100-fold excess of unlabeled probe oligonucleotide (Fig. 2b, lanes 3 and 8) and consensus Sp1 oligonucleotide (Fig. 2b, lanes 5 and 10) used as specific competitors and with a 100-fold excess of dI/dC used as a nonspecific competitor (Fig. 2b, lanes 4 and 9). An antibody against Sp1 interacted with this bound complex, as shown by the decrease in band intensity and appearance of a weak supershifted band (Fig. 2b, lane 7). Nonimmune goat serum had no affect on the bound complexes. These experiments clearly demonstrated that the region encompassing the Sp1 motif at position −51 bound nu-
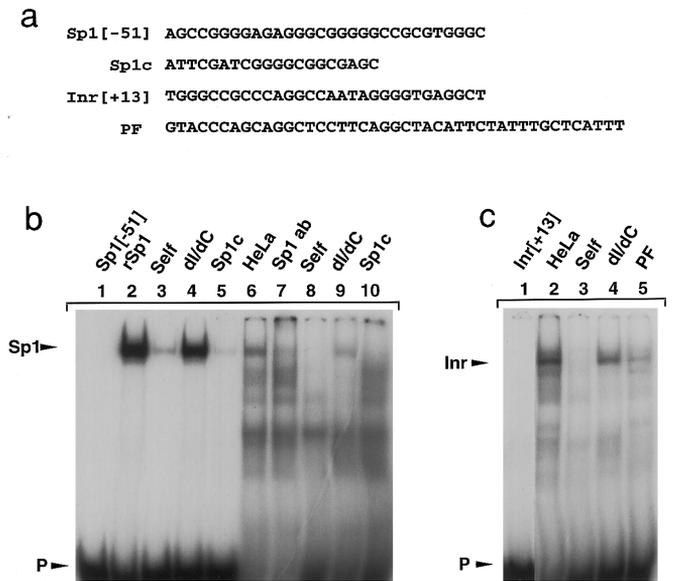
FIG. 2. Band-shift analysis of the putative Sp1 and initiator motifs. (a) Oligonucleotide sequences. The locations of Sp1[−51] and Inr[+13] in the FALDH gene are depicted in Fig. 1d. Numbering in these probes is relative to the transcription start site. (b) Band-shift experiments with radioactively labeled Sp1[−51] probe (P). The retarded protein-DNA complexes are labeled Sp1. (Lanes 2–5) Recombinant Sp1 binding in the absence (lane 2) and presence of competition with unlabeled Sp1[−51] oligonucleotide (lane 3), dI/dC (lane 4), and the consensus Sp1 oligonucleotide (lane 5). (Lanes 6–10) HeLa nuclear extracts in the absence (lane 6) and presence of Sp1 antibody (lane 7) and in competition with unlabeled Sp1[−51] oligonucleotide (lane 8), dI/dC (lane 9), and the consensus Sp1 oligonucleotide (lane 10). (c) Band-shift experiments with radioactively labeled Inr[+13] probe (P). The retarded protein–DNA complexes are labeled Inr. (Lanes 2–5) HeLa nuclear extracts with labeled probe in the absence (lane 2) and presence of competition with unlabeled Inr[+13] oligonucleotide (lane 3), dI/dC (lane 4), and an oligonucleotide derived from the profilaggrin gene and known to carry an initiator-like motif (lane 5).

clear proteins from HeLa cells. Further analysis will be required to confirm the identity of the protein factor(s) binding at this site.

Factors in HeLa extract were also found to interact with an oligonucleotide containing the initiator motif (Inr[+13]; Fig. 2a). The bound complex had a greater mobility than the complex bound to the Sp1 oligonucleotide of the same size (compare Fig. 2b, lane 6, with Fig. 2c, lane 2). This binding was successfully competed by the unlabeled probe (Fig. 2c, lane 3), but was not competed with dI/dC (Fig. 2c, lane 4). Significantly, this bound complex was also competed for by the PF oligonucleotide (Fig. 2c, lane 5), which is derived from the profilaggrin gene (Markova *et al.,* 1993) and known to carry an initiator-like motif, but no Sp1 binding site (Fig. 2a). Although the initiator oligonucleotide Inr[+13] also contains an Sp1 motif located at −2 (relative to the transcription start site), the different mobility shift of the observed complex and the competition by PF indicate that other sequence motifs in the probe sequence interact with nuclear proteins.

We explored whether interactions at the initiator sequence involved the YY.1 protein but could not detect binding to recombinant YY.1, and an antibody against YY.1 was not able to interfere with complex formation observed in HeLa extracts (data not shown). However, there are a number of proteins other than YY.1 that bind to initiator sequences (Azizkhan *et al.,* 1993), and one of these may be responsible for transcription initiation in the FALDH gene.

Expression Analysis

Human FALDH was shown to be widely expressed in Northern blot analysis of poly(A)$^+$ mRNA with a probe from the unique 3′ UTR. Transcripts of about 2 and 4 kb were observed in seven different tissues (Fig. 3a). The larger band could be resolved into 3.8- and 4-kb transcripts after longer electrophoresis, as shown in Fig. 3b for keratinocyte and epidermal poly(A)$^+$ mRNA. Differential expression of the short (2 kb) and the longer (3.8 and 4 kb) FALDH transcripts between and within tissues was also evident. The relative band intensities suggested an excess of the longer transcripts in heart, brain, skeletal muscle, and pancreas, comparable expression in epidermis, and an excess of the 2-kb transcript in liver. The longer transcripts predominated in all areas of the brain as demonstrated on a Northern blot containing poly(A)$^+$ mRNA from eight different regions of the brain (data not shown).

Multiple Polyadenylation Sites Are Responsible for the Variation in FALDH Transcript Length

We hypothesized that the variation in FALDH transcript size observed by Northern blot analysis may be caused by multiple polyadenylation sites, and we tested this with 3′ RACE experiments. Using liver RNA tem-
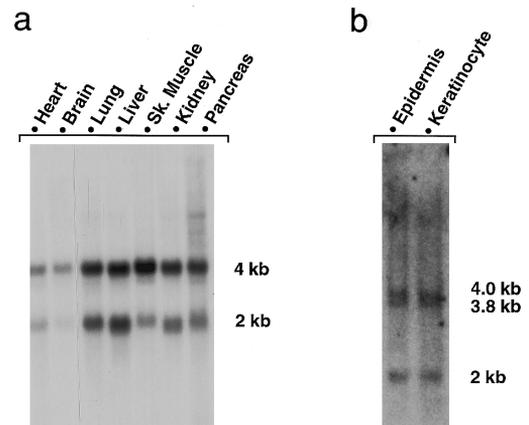

FIG. 3. Northern analysis with FALDH 3′ UTR probe showing that FALDH is widely expressed as multiple transcripts. All lanes have comparable amounts of RNA as judged by probing with a glyceraldehyde-3-phosphate dehydrogenase-specific probe. (a) Multiple tissue poly(A)$^+$ RNA Northern blot (Clontech). (b) Northern blot with foreskin epidermis and normal human epidermal keratinocyte poly(A)$^+$ RNA.

plate and a sense primer in exon 9 of the FALDH gene, three 3′ RACE products of approximately 300, 2000, and 2200 bp were generated. The sequence of the 300-bp product showed that it contained only the known terminal FALDH coding sequence and the previously defined 170-bp 3′ UTR (GenBank L47162). The 5′ end of both larger RACE products also matched these sequences but contiguous sequences beyond the known 3′ end of the FALDH cDNA were present. The 2200-bp RACE product extended a further 199 nucleotides beyond the 2000-bp product. A perfect consensus poly-adenylation signal (AATAAA) was found 12 nucleotides upstream of the poly(A)$^+$ tract in the shortest 3′ RACE product and imperfect signals (NATAAA) 17 and 20 nucleotides upstream in the 2- and 2.2-kb 3′ RACE products, respectively. The absence of introns inter-rupting the larger 3′ UTRs of the FALDH gene was demonstrated by equal sized products following PCR amplification from either genomic DNA or 3′ RACE DNA templates.

Alternatively Spliced Products Detected by RT-PCR

A comparison of the FALDH cDNA with the sequence databases identified a partial cDNA clone from brain (GenBank H29247) that contained FALDH exon 10 but diverged from the FALDH sequence at the position of the exon 9/exon 10 junction. We tested whether this clone represented an alternatively spliced product of the FALDH primary transcript by performing RT-PCR with primers located in exons 9 and 10. In addition to the expected FALDH fragment, a larger RT-PCR product that contained an additional 125 nucleotides (designated exon 9′) between FALDH exon 9 and exon 10 was obtained (Fig. 4). Its sequence essentially matched that of the partial cDNA from brain (GenBank H29247). The presence of transcripts containing the extra exon as well as all 10 FALDH exons was demon-strated by RT-PCR using primers in exon 9′ with up-stream and downstream FALDH primers. Sequencing the 3′ RACE product amplified from an exon 9′ sense primer showed that it contained the identical 170 nu-cleotides of 3′ UTR sequence as FALDH transcripts lacking exon 9′. PCR amplification from genomic DNA demonstrated that the nucleotide sequence of exon 9′ was contained within intron 9, approximately 1.1 kb from the exon 9/intron 9 splice junction. It was flanked by consensus donor and acceptor splice junctions. The alternatively spliced product could be amplified from all tissue RNAs we tested (liver, normal human epider-mal keratinocyte, fibroblast, brain) with a similar low abundance (<10%) compared to RT-PCR products lack-ing the extra exon. Translation of this alternatively spliced transcript would terminate within exon 9′, pro-ducing a protein containing an additional 27 amino acids in place of four carboxy-terminal residues of the more abundant FALDH protein (Fig. 4). Fourteen of these amino acids are commonly found in transmem-
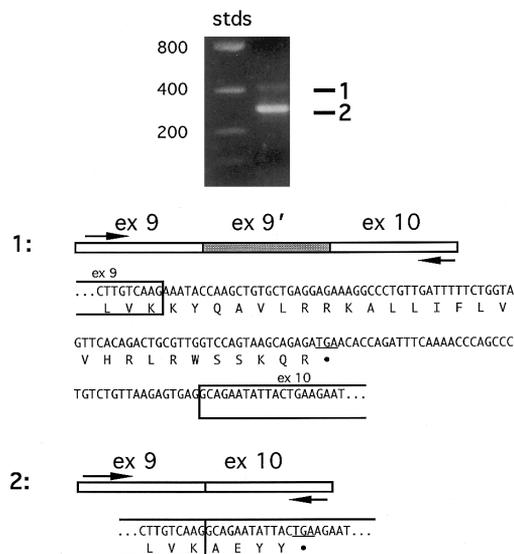

FIG. 4. Alternatively spliced product (1) observed following agar-ose gel electrophoresis of products from RT-PCR with primers located in exons 9 and 10. The nucleotide sequence, with translation, of the alternatively spliced product is shown and can be compared with the sequence of the product of expected size (2).

brane domains (e.g. valine, leucine), and nine are basic (e.g. lysine, arginine).

RT-PCR from exons 3 to 6 on fibroblast RNA gave an unexpectedly short product of very low abundance (<1%) in addition to one with the predicted length. The shorter product lacked the 327 nucleotides encoding exons 4 and 5. If translated, the resulting protein would resemble FALDH with an in-frame deletion of 109 amino acids—a potentially catastrophic effect on FALDH function. Further studies are required to deter-mine whether this is an alternatively or aberrantly spliced product.

Organization of the Aldehyde Dehydrogenase Cluster Containing the FALDH and ALDH3 Genes

In addition to the extensive similarities in their nu-cleotide sequence and gene structure, the FALDH and ALDH3 genes both map to 17p11.2 (De Laurenzi *et al.,* 1996; Hiraoka *et al.,* 1995). Close physical proximity of the two genes was established by PCR amplification of both genes from three YACs—965d9 (900 kb), 781d10 (1100 kb), and 855h1/1 (650 kb). We further examined the complexity and arrangement of the ALDH genes in this cluster by restriction analysis of the YACs.

To test for additional related ALDH genes, a 340-bp FALDH cDNA probe spanning exons 4 through 7 with 71% nucleotide similarity to ALDH3 was used to probe *Eco*RI-digested YAC DNA at reduced stringency (data not shown). The observed bands of 2.6, 4.2, and 10 kb in 781d10, 965d9, and 855h1/2 (a spontaneous deletion variant shown to be only positive for FALDH by PCR) were all accounted for by sequential hybridization with FALDH-specific (4.2 and 10 kb) and ALDH3-specific

(2.6 and 10 kb) probes under high-stringency conditions. Six other 17p11.2 YACs that did not contain either the FALDH or the ALDH3 genes were negative when hybridized with the 340-bp probe. These results indicated that there were no other related ALDH genes located at 17p11.2.

A physical map was constructed from additional Southern analysis of YAC 781d10 DNA digested with rare-cutting restriction endonucleases and probed with FALDH- and ALDH3-specific probes. Both genes were found on one 120-kb *Nru*I restriction fragment, but on different *Mlu*I restriction fragments of 80 kb (FALDH) and 25 kb (ALDH3). There is a recognition sequence for *Nru*I in exon 1 and for *Mlu*I at the 5′ end of intron 1 in the FALDH gene and a recognition sequence for *Mlu*I in the 5′ region of the ALDH3 gene (Hsu *et al.*, 1992). Together with the restriction data and the sizes of FALDH (30.5 kb) and ALDH3 (8 kb ; Hsu *et al.*, 1992), we could deduce that the two genes were separated by at least 50 kb but not more than 85 kb.

DISCUSSION

A critical role for FALDH in the normal functioning of the human nervous system and epidermal tissue was recently established when mutations in the FALDH gene were shown to cause SLS (De Laurenzi *et al.*, 1996), a disease with both neurological and dermatological manifestations (reviewed Rizzo, 1993). We have extended these studies by characterizing the genomic organization of the FALDH gene and investigating its expression pattern. This information will facilitate mutation analysis in SLS patients as well as provide the first steps toward understanding gene regulation and elucidating the pathological mechanism of SLS.

The high nucleotide and amino acid sequence identity between FALDH and ALDH3 is paralleled by the virtually identical exon/intron structures of the two genes. Both coding sequences are partitioned into 10 exons by intronic interruptions at the same positions in the two aligned sequences, with just the one exception of the exon 9/intron 9 junction. Interestingly, the unique FALDH sequence encoding the membrane targeting domain at the carboxy-terminal end of the enzyme is contiguous with the sequence that is homologous to ALDH3 exon 9, rather than being encoded by a separate exon. Similarities have been noted in the gene structures of other Class 3 ALDHs—ALDH3, ALDH7 and ALDH8 (Hsu *et al.*, 1995). A comparison of exon/intron structures of all four Class 3 ALDHs suggests that merging and partitioning of exons have been common events in the evolution of this gene family.

The tandem organization of the Class 3 genes is of evolutionary interest. Among these genes, FALDH and ALDH3 most resemble each other, whereas ALDH7 and ALDH8 share the greatest homology. The related genes within each pair are also linked in the genome.

FALDH and ALDH3 lie between 50 and 85 kb apart at 17p11.2 while ALDH7 and ALDH8 both map to 11q13 (Hsu *et al.*, 1995). The mouse homologs of FALDH (AHD3) and ALDH3 (AHD4) are also tightly linked, with an estimated separation of between 10 and 200 kb (Vasiliou *et al.*, 1996). The close linkage, sequence similarity, and structural conservation of the ALDH3 and FALDH genes indicate that they share a common origin.

The primary FALDH transcript was shown to be processed into three major mature mRNA species. The likely causes of variation in transcript length, i.e., use of multiple transcription initiation sites, multiple polyadenylation sites, or alternative splicing of the primary transcript, were all investigated. A single major site of transcription initiation was determined to be at nucleotide −258 by combining data from both primer extension and RNase protection experiments. Sequence analysis around the transcription initiation site revealed no consensus TATA motif. Many genes with TATA-less promoters have been described, including ubiquitously expressed "housekeeping" genes, oncogenes, growth factors and their receptors, and transcription factors (reviewed Azizkhan *et al.*, 1993). FALDH may qualify as a housekeeping enzyme considering its function and that we detected FALDH gene expression in all of the diverse tissue and cell types tested. In common with other TATA-less promoters, the FALDH 5′ UTR and promoter region contained multiple CpG islands and several Sp1 binding motifs (Azizkhan *et al.*, 1993). The position of one such motif at 51 nucleotides upstream of the transcription initiation site is optimally placed for transcription activation (Smale and Baltimore, 1989), and we confirmed that it binds nuclear proteins by a series of band-shift assays. We also identified a putative initiator sequence near the transcription initiation site that conformed to the 5′-YYCAYYYY-3′ (Y, pyrimidine) consensus (Azizkhan *et al.*, 1993) and interacted with nuclear proteins in band-shift experiments. Furthermore, the results of our *in vitro* transcription experiment showed that the sequence in the vicinity of the defined transcription initiation site was sufficient to promote faithful transcription. Taken together, our results demonstrate that FALDH gene transcription *in vivo* and *in vitro* originates from a TATA-less promoter. The exact nature of the factors and the extent of the sequences controlling the level and specificity of FALDH gene expression from this promoter remain to be studied.

The possibility of multiple polyadenylation sites was tested by 3′ RACE experiments. Three polyadenylation sites were identified, and the predicted lengths of the 3′ UTRs accounted for the variation in FALDH transcript length. 3′ UTR sequences can have wide-ranging effects on gene expression by regulating localization, stability, and translation of mRNAs (reviewed by Decker and Parker, 1995). Since we observed tissue-specific differences in the relative amounts of FALDH tran-

scripts, we can speculate that the variation in 3′ UTR influences FALDH gene expression, and this warrants further investigation. It is interesting to note that the rat and mouse equivalents of FALDH are expressed as a single 3-kb transcript with relatively long 3′ UTRs of 1402 and 1429 bp, respectively (Miyauchi *et al.,* 1991; Vasiliou *et al.,* 1996). Brain is one of the tissues most affected in SLS, and Northern analysis showed that the larger FALDH transcripts were predominant in all regions of the brain. Variation in the 3′ UTR sequence could result in different pathological consequences of mutations in the FALDH gene by altering translation, turnover, or localization.

We identified an alternatively spliced FALDH product, raising the possibility that different FALDH-like proteins may be synthesized. The inclusion of the extra exon between exons 9 and 10 would result in a protein with an increased putative transmembrane domain at the carboxy-terminal region, since many of the 27 additional amino acids were hydrophobic. This may have important ramifications for localization or stability of the FALDH enzyme. However, we know that this alternatively spliced product was of low abundance given the visual comparisons following gel electrophoresis of the RT-PCR products. Further studies are required to determine whether it is indeed translated and to assess the function of the alternative protein.

The wide tissue distribution of FALDH expression contrasts with the relatively restricted clinical pathology associated with SLS. The neurological and cutaneous defects consistently observed in SLS are associated with widespread demyelination of brain and spinal cord and hyperkeratotic alterations in the epidermis (Rizzo, in press). FALDH is expressed as multiple transcripts in these tissues. On the other hand, the high levels of expression in liver, lung, and skeletal muscle do not correlate with the presence of pathology. It has not been established how the enzymatic defect leads to tissue defects, but it is likely that accumulation of fatty alcohol or its metabolic products will be involved in pathogenesis (Rizzo, 1993). Therefore, in tissues where the loss of FALDH function has no apparent consequence, substrate accumulation may be mitigated by other enzymatic activities or tissue function may not be sensitive to the lipid imbalance. Clearly, more detailed understanding of the spectrum of fatty aldehyde dehydrogenase activities in the various tissues and the role of multiple FALDH transcripts in enzymatic function is needed. If other genes are compensating for FALDH deficiency, then upregulation of their expression in neurological and epidermal tissues may be a conceivable therapeutic strategy.

ACKNOWLEDGMENTS

REFERENCES

Azizkhan, J. C., Jensen, D. E., Pierce, A. J., and Wade, M. (1993). Transcription from TATA-less promoters: Dihydrofolate reductase as a model. *Crit. Rev. Eukaryotic Gene Expression* **3**(4): 229–254.

Decker, C. J., and Parker, R. (1995). Diversity of cytoplasmic functions for the 3′ untranslated region of eukaryotic transcripts. *Curr. Opin. Cell Biol.* **7**: 386–392.

De Laurenzi, V., Rogers, G. R., Hamrock, D. J., Marekov, L. N., Steinert, P. M., Compton, J. G., Markova, N., and Rizzo, W. B. (1996). Sjögren–Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene. *Nature Genet.* **12**: 52–57.

Hiraoka, L. R., Hsu, L., and Hsieh, C.-L. (1995). Assignment of ALDH3 to human chromosome 17p11.2 and ALDH5 to human chromosome 9p13. *Genomics* **25**: 323–325.

Hsu, L. C., Chang, W.-C., Shibuya, A., and Yoshida, A. (1992). Human stomach aldehyde dehydrogenase: DNA and genomic cloning, primary structure, and expression in *Escherichia coli. J. Biol. Chem.* **267**: 3030–3037.

Hsu, L. C., Chang, W.-C., Lin, S. W., and Yoshida, A. (1995). Cloning and characterization of genes encoding four additional human aldehyde dehydrogenase isozymes. *In* "Enzymology and Molecular Biology of Carbonyl Metabolism" (H. Weiner *et al.,* Eds.) Vol. 5, pp. 159–168, Plenum, New York.

Jang, S.-I., Steinert, P. M., and Markova, N. G. (1996). Activator Protein 1 activity is involved in the regulation of the cell type specific expression from the proximal promoter of the human profilaggrin gene. *J. Biol. Chem.* **271**(39): 24105–24114.

Kelson, T. L., Craft, D. A., and Rizzo, W. B. (1992). Carrier detection for Sjögren-Larsson syndrome. *J. Inherit. Metab. Dis.* **15**: 105–111.

Kelson, T. L., Secor McVoy, J. R., and Rizzo, W. B. Human liver fatty aldehyde dehydrogenase: Micosomal localization, purification, and biochemical characterization. *Biochim. Biophys. Acta,* in press.

Lacour, M., Middleton-Price, R., and Harper, J. I. (1996). Confirmation of linkage of Sjögren–Larsson syndrome to chromosome 17 in families of different ethnic origins. *J. Med. Genet.* **33**: 258–259.

Lindahl, R. (1992). Aldehyde dehydrogenases and their role in carcinogenisis. *Crit. Rev. Biochem. Mol. Biol.* **27**: 283–335.

Lindahl, R., and Petersen, D. R. (1991). Lipid aldehyde oxidation as a physiological role for class 3 aldehyde dehydrogenases. *Biochem. Pharmacol.* **41**: 1583–1587.

Markova, N. G., Marekov, L. N., Chipev, C. C., Gan, S.-Q., Idler, W. W., and Steinert, P. M. (1993). Profilaggrin is a major epidermal calcium-binding protein. *Mol. Cell. Biol.* **13**: 613–625.

Masaki, K., Masaki, R., Yamamoto, A., and Tashiro, Y. (1994). Microsomal aldehyde dehydrogenase is localized to the endoplasmic reticulum via its carboxyl-terminal 35 amino acids. *J. Cell Biol.* **126**: 1407–1420.

Miyauchi, K., Masaki, R., Taketani, S., Yamamoto, A., Akayama, M., and Tashiro, Y. (1991). Molecular cloning, sequencing and expression of cDNA for rat liver microsomal aldehyde dehydrogenase. *J. Biol. Chem.* **266**: 19536–19542.

Pigg, M., Jagell, S., Sillén, A., Weissenbach, J., Gustavson, K.-H., and Wadelius, C. (1994). The Sjögren–Larsson syndrome gene is close to D17S805 as determined by linkage analysis and allelic association. *Nature Genet.* **8**: 361–364.

Richards, B., Skoletsky, J., Shuber, A. P., Balfour, R., Stern, R. C., Dorkin, H. L., Parad, R. B., Witt, D., and Klinger, K. W. (1993). Mulitplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. *Hum. Mol. Genet.* **2**: 159–163.

Rizzo, W. B. (1993). Sjögren–Larsson syndrome. *Semin. Dermatol.* **12**: 210–218.

Rizzo, W. B. Sjögren–Larsson syndrome. *In* "Handbook of Clinical Neurology" (P. J. Vinken and G. W. Bruyn, Eds.), Elsevier, Amsterdam, in press.

Rizzo, W. B., and Craft, D. A. (1991). Sjögren–Larsson syndrome: Deficient activity of the fatty aldehyde dehydrogenase component of fatty alcohol:NAD+ oxidoreductase in cultured fibroblasts. *J. Clin. Invest.* **88:** 1643–1648.

Rogers, G. R., Rizzo, W. B., Zlotogorski, A., Hashem, N., Lee, M., Compton, J. G., and Bale, S. J. (1995). Genetic homogeneity in Sjögren–Larsson syndrome: Linkage to chromosome 17p in families of different non-Swedish ethnic origins. *Am. J. Hum. Genet.* **57:** 1123–1129.

Shapiro, M., and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **17:** 7155–7175.

Sjögren, T., and Larsson, T. (1957). Oligophrenia in combination with congenital ichthyosis and spastic disorders. *Acta Psychiatr. Neurol. Scand.* **32**(Suppl.), 1–113.

Smale, S. T., and Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell* **57:** 103–113.

Vasiliou, V., Kozak, C. A., Lindahl, R., and Nebert, D. W. (1996). Mouse microsomal Class 3 aldehyde dehydrogenase: AHD3 cDNA sequence, inducibility by dioxin and clofibrate, and genetic mapping. *DNA Cell Biol.* **15:** 235–245.