Parker, C. S., & Topol, J. (1984) *Cell 36*, 357-369.
Parniewski, P., Kwinkowski, M., Wilk, A., & Klysik, J. (1990) *Nucleic Acids Res. 18*, 605-611.
Plum, G. E., Park, Y.-W., Singleton, S. F., Dervan, P. B., & Breslauer, K. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9436-9440.
Prentki, P., Pham, M.-H., & Galas, D. J. (1987) *Nucleic Acids Res. 15*, 10060.
Saltzman, A. G., & Weinmann, R. (1989) *FASEB J. 3*, 1723-1733.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Strobel, S. A., & Dervan, P. B. (1991) *Nature 350*, 172-174.
Weintraub, H. (1984) *Cell 38*, 17.
Wu, H.-M., & Crothers, D. M. (1984) *Nature 308*, 509-513.

Structural Organization and Regulatory Regions of the Human Medium-Chain Acyl-CoA Dehydrogenase Gene[†,‡]

Zhifang Zhang,[§] Daniel P. Kelly,[§] Jung-Ja Kim,[∥] Yeqing Zhou,[§] Moira L. Ogden,[§] Alison J. Whelan,[§] and Arnold W. Strauss*[,§]

Department of Biochemistry, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, and Departments of Pediatrics, Biochemistry and Molecular Biophysics, and Medicine, Washington University School of Medicine, St. Louis, Missouri 63110

Received July 16, 1991; Revised Manuscript Received September 24, 1991

ABSTRACT: Medium-chain acyl-CoA dehydrogenase (MCAD) is a highly regulated mitochondrial flavoenzyme that catalyzes the initial reaction in fatty acid β-oxidation. Deficiency of MCAD is a common inherited defect in energy metabolism. We have previously shown that the mRNA encoding MCAD in an MCAD-deficient child is homozygous for the point mutation A^{985} to G [Kelly et al. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9236-9420]. To define the molecular basis of MCAD deficiency and as an initial step in the study of the regulation of MCAD gene expression, we determined the structure and organization of the human MCAD gene. The gene is comprised of 12 exons which span 44 kb of DNA. Comparison of the MCAD gene to MCAD mRNAs from the MCAD-deficient child revealed that missplicing was common, resulting in a variety of exon deletions and intron insertions. The MCAD gene promoter region is extremely GC-rich and lacks prototypical TATA and CAAT boxes. Several regions upstream of the promoter are homologous with mitochondrial enhancers purportedly involved in coordinate expression of nuclear genes encoding mitochondrial proteins. Transfection of chimeric plasmid constructs with 299 bp of upstream sequence into HepG2 cells revealed high-level transcriptional activity. We conclude that the precursor MCAD mRNA is misspliced to a high degree and complexity in association with the G^{985} mutation and the MCAD gene contains a strong promoter which shares some structural features with other "housekeeping" genes encoding mitochondrial proteins.

Medium-chain acyl-CoA dehydrogenase (MCAD; 2,3-oxidoreductase, EC 1.3.99.3)[1] is a mitochondrial matrix flavoprotein which catalyzes the first reaction in the β-oxidation of straight-chain fatty acids (Beinert, 1963). It forms an enzyme family with short- and long-chain acyl-CoA dehydrogenases. These three enzymes are homotetramers with subunits of similar size, but they are immunologically distinct with different amino acid sequences (Ikeda et al., 1985). MCAD requires medium-chain-length acyl-coenzyme A substrates, contains FAD, and ultimately transfers electrons to electron-transfer flavoprotein. Like most mitochondrial proteins, MCAD is encoded by a nuclear gene, is synthesized in the cytosol as a larger precursor with an NH_2-terminal transit peptide, and is subsequently imported into mitochondria with proteolytic processing to the mature form (Warren, 1987; Kelly et al., 1987).

Inherited deficiency of human MCAD was first described in 1983 and is now recognized as a common inherited metabolic disorder (Kolvraa et al., 1983; Rhead et al., 1983; Stanley et al., 1983). Clinical manifestations of MCAD deficiency include fasting hypoglycemia, recurrent Reye-like syndrome, or sudden death during the first 2 years of life. MCAD-deficient individuals may also be completely asymptomatic (Bougneres et al., 1985; Treem et al., 1986; Duran et al., 1986). If fasting crises are avoided, MCAD-deficient individuals develop normally.

As an initial step in the molecular analysis of MCAD deficiency and to study the regulation of expression of MCAD, we previously isolated and characterized cDNA clones encoding human and rat MCAD mRNAs (Kelly et al., 1987,

[†] This work was supported by Grants AM 20407 (A.W.S. and Z.Z.) and GM 29076 (J.-J.K.) from the National Institutes of Health. A.J.W. is a Fellow of the American Heart Association, Missouri Affiliate. D.P.K. is the recipient of a Lucille P. Markey Scholar Award.

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05355.

* Address correspondence to this author at the Department of Pediatrics, Division of Cardiology, St. Louis Children's Hospital, 400 S. Kingshighway, St. Louis, MO 63110.
[§] Washington University School of Medicine.
[∥] Medical College of Wisconsin.

[1] Abbreviations: MCAD, medium-chain acyl-CoA dehydrogenase; bp, base pair(s); PCR, polymerase chain reaction; CAT, chloramphenicol acetyltransferase; UTR, untranslated region.

0006-2960/92/0431-81$03.00/0

1989). The MCAD gene has been localized to human chromosome 1 (Matsubara et al., 1986). MCAD mRNA is ubiquitously expressed in many tissues, although the level of expression varies dramatically, with the highest levels in tissues with the greatest fatty acid oxidative capacity (heart, adrenal, and skeletal muscle) (Kelly et al., 1989). In addition, cardiac and hepatic MCAD mRNA expression is regulated during development in parallel with the known changes in oxidative energy substrate utilization.

The molecular basis of MCAD deficiency has recently been determined (Kelly et al., 1990; Matsubara et al., 1990; Yokota et al., 1990a,b; Gregersen et al., 1991). In two unrelated patients, we demonstrated an A to G change at nucleotide 985 of the mutant MCAD cDNA coding region. This point mutation causes substitution of glutamic acid for a lysine at amino acid 304 of the mature protein and has been documented in over 50 additional MCAD-deficient patients from the United States and northern Europe (Kelly et al., 1990; Matsubara et al., 1990).

To define further the molecular mechanism of inherited MCAD deficiency, to delineate mechanisms involved in tissue-specific and developmental regulation of MCAD expression, and to compare the structures and the common cis-acting regulatory DNA elements in human nuclear genes encoding mitochondrial proteins, we report herein the isolation and characterization of the complete human MCAD gene with its promoter and upstream regulatory region.

Experimental Procedures

Isolation and Mapping of the MCAD Gene. To screen the human genomic library and map the MCAD gene, several MCAD cDNA fragments were employed as hybridization probes. In all experiments, probes are designated by nucleotide position within MCAD cDNA, with the transcription start site as +1. The translation start codon begins at nucleotide +193. In addition, synthetic oligonucleotides were made as probes and are also named by nucleotide position within the MCAD cDNA. Some probes were generated by the polymerase chain reaction (PCR) from the cDNA. In these instances, DNA amplification was performed with *Thermus aquaticus* (*Taq*) polymerase in a Perkin-Elmer/Cetus programmable thermocycler by standard protocol (Saiki et al., 1988). All DNA fragment probes were labeled with [α-^{32}P]dCTP by the random primer method (Feinberg & Vogelstein, 1983), and oligonucleotides were end-labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase. Seven probes (bp +184 to +218, +196 to +864, +207 to +242, +224 to +340, +799 to +832, +834 to +1235, and +885 to +1387) were emplyed to screen a human choriocarcinoma genomic library constructed in the *Bam*HI site of λMG_3. Repetitive screenings of duplicate lifts were performed with these different probes in an attempt to isolate phage clones containing all exons of the gene. Unfortunately, we were unable to isolate clones from the choriocarcinoma library containing exon 10. Therefore, a human leukocyte library constructed in the *Bam*HI site of cosmid vector pTCF was screened with probes bp +1051 to +1140 and +1051 to +1235. This cosmid library was a gift from Rick Wetzel (Washington University School of Medicine). Phage and cosmid clones were isolated and purified by standard screening protocols (Kaiser & Murray, 1985).

Restriction Enzyme Mapping and Southern Blotting. Single, double, or triple restriction enzyme digests of phage and cosmid clone DNA were analyzed after agarose gel electrophoresis and transfer to nitrocellulose paper as described previously (Sambrook et al., 1989). Restriction maps were generated by hybridization of these Southern blots with appropriate MCAD cDNA fragments or oligonucleotides as probes.

DNA Sequence Analysis. Overlapping fragments of genomic DNA containing exonic sequences were subcloned into pGEM3Z or pGEM4Z plasmids. Nucleotide sequence analysis was by the dideoxy nucleotide chain termination method (Sanger et al., 1977) with either the appropriate universal primers or synthetic sense and antisense oligonucleotides from both cDNA and MCAD genomic sequences. In all, more than 28 kb of sequence was determined.

Primer Extension. To determine the transcription start site, an antisense, synthetic 34-base oligonucleotide (bp +109 to +142) was employed as a primer. The primer extension reaction was done as described (Haas & Strauss, 1990). A DNA sequence ladder generated by the dideoxy nucleotide chain termination method was fractionated on the adjacent gel lanes to allow determination of the precise length of the primer-extended fragment.

Construction of MCAD–Chloramphenicol Acetyltransferase Gene Chimera. Chimeric plasmids were constructed in the promoterless plasmid pCAT-basic (Promega). One test plasmid contained a genomic *Pst*I/*Bgl*I DNA fragment from bp −1193 to +179. This was designated MCADCAT-long (numbers are reference to the cap site at +1). The *Pst*I site was employed to insert the 5′ end of this fragment into the pCAT-basic plasmid followed by blunting of the free ends with T4 DNA polymerase and ligation to complete the construct. Plasmids designated as MCADCAT-long, MCADCAT-short and MCADCAT-del were made by PCR amplification of MCAD genomic DNA utilizing a plasmid template (pGEM 3Z containing the *Pst*I fragment from −1193 to +221 of the MCAD gene) and oligonucleotide primers designed to amplify the region containing bp −299 to +149 (MCADCAT-short) and −966 to +149 (MCADCAT-del). The PCR conditions included melting at 94° for 1 min, annealing at 58° for 2 min, and extension at 74° for 4 min for 25 cycles in the thermocycler. Both 5′ primers contained a *Hin*dIII site and the 3′ primers contained an *Xba*I site. The amplified DNA was placed into the *Hin*dIII/*Xba*I sites of the PCAT basic vector and subjected to DNA sequence analysis. The entire MCAD genomic DNA inserts were sequenced revealing that MCADCAT-del contained an internal deletion within the MCAD DNA insert from bp −178 to +59. The remainder of the sequence in both constructs was identical to the corresponding gene sequence. The plasmid pMSVβgal was generously provided by Nadia Rosenthal (Boston University).

Cell Transfection and Chloramphenicol Acetyltransferase (CAT) Assays. Human HepG2 cells were maintained in an atmosphere of 5% CO_2 in growth media (Dulbecco's modified Eagle's medium supplemented with 10% Nu-Serum). The cells were plated at a density of approximately 1.5×10^6 on 60-mm dishes on the day prior to the transfections. The next day, the cells were refed with 3 mL of medium containing 10% Nu-Serum and incubated for 3 h. Transfections were then performed by the calcium phosphate coprecipitation method (Gorman, 1985). Each precipitate contained 15 μg of test plasmid and 5 μg of pMSVβgal to normalize for transfection efficiency and cell number among dishes. The cells were incubated with the precipitate for 4 h followed by a 3-min glycerol shock. The cells were then washed and refed with medium containing 10% Nu-Serum. Approximately 48 h later, the cells were harvested, and cell extracts were prepared by three cycles of freeze–thawing in 150 μL of 0.25 mM Tris-HCl, pH 7.8, followed by sonication. To normalize for transfection efficiency and cell pellet size, β-galactosidase

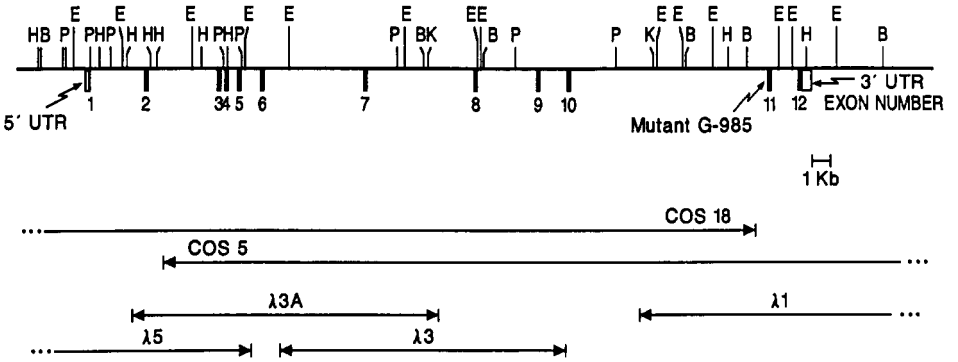

FIGURE 1: Structure of the human medium-chain acyl-CoA dehydrogenase gene. The solid vertical bars denote the 12 exons of the MCAD gene. The open bars indicate the 5'- or 3'-untranslated regions (UTR) in exons. A partial restriction map is shown. The abbreviations of restriction enzymes for the sites illustrated are: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I. The common mutation of human MCAD deficiency, shown as mutant G-985, is located in exon 11 (see text). The sizes and locations of phage and cosmid clones employed in this analysis are indicated at the bottom of the figure.

activity was determined on a 30-μL volume of extract as previously described (Rosenthal, 1987). The volume of extract used for CAT assay was determined by the results of the β-galactosidase activity assays. The CAT assays were performed in 125 μL containing 0.125 μCi of [^{14}C]chloramphenicol and 6 nmol of *n*-butyryl-CoA at 37 °C for 1 h (Seed & Sheen, 1988). The butyrylated chloramphenicol was separated from the free chloramphenicol by xylene extraction and was quantified by liquid scintillation counting. Percent conversion of chloramphenicol was calculated by dividing the counts in the xylene phase by the total counts in the starting material. This assay is linear up to 50% conversion, and the slope of the time–activity curve is linear at 60 min.

Results and Dicussion

Organization and Structure of the Human MCAD Gene. Four λ phage clones were isolated from a human choriocarcinoma genomic library and detailed restriction maps were generated by Southern blot analyses (Figure 1). Intron–exon junctions and a substantial amount of intronic sequence were subjected to nucleotide sequence determination. These results proved that the phage clones contain the 5'-flanking region, 5'-UTR, exons 1–9, exons 11–12, 3'-UTR, and 3'-flanking region. Therefore, the λ phage clones encompassed most of the MCAD gene. Despite repeated screening with exon 10 specific probes, however, we could not locate this exon in the phage library. To overcome this problem, we screened a human cosmid library. Two clones, designated cos 18 and cos 5, were isolated. Detailed restriction mapping with Southern blot analyses employing numerous oligonucleotide and DNA fragment probes, including some subclones from the phage clones, demonstrated that the two cosmid clones overlap and contain the entire MCAD gene (Figure 1).

The intron–exon organization of the human MCAD gene is depicted in Figures 1 and 2. The gene consists of 12 exons interrupted by 11 introns. Exons 1–10 are small, ranging in length from 70 to 140 bp. The two 3' exons are larger, with exon 11 containing 249 bp of coding sequence and exon 12 having 807 bp, incluing the entire 3'-UTR. The composite nucleotide sequence of the exons is identical to our previously published human MCAD mRNA (Kelly et al., 1987). Exons 3–6 and 8–10 are clustered within relatively small segments of the gene. The introns vary in length from 71 bp to more than 10 500 bp. Two large introns (6 and 5.6 kb, respectively) separate exons 6 through 8. Exon 11 is preceded by the 10.5-kb intron, a fact which may relate to the precursor mRNA splicing abnormalities in this gene described below. Exonic sequences encompass only 5% of the gene. The sequences of all exon–intron junctions follow the splicing rule, with GT at the donor site and AG at the acceptor site of the intron (Figure 2). In general, the nucleotide sequences surrounding intron–exon junctions conform to the splice site consensus sequences (Figure 2).

Relationship of Gene and Protein Structures. The crystal structure of tetrameric pig MCAD has been determined at 2.2-Å resolution by X-ray diffraction (Kim & Wu, 1988; J.-J. Kin, unpublished results). Comparison of this structure with the human MCAD gene organization (Figure 3) reveals that the NH_2-terminal α-helical domain (helices A–F) is encoded by exons 2–6; the middle domain of two packed β-sheets correlates with exons 6–8, and the COOH-terminal helical domain (helices G–L) corresponds with exons 9–12. Of the residues which form the fatty acyl binding cavity, Gln-95, Glu-99, Gly-100 (which is an alanine in pig MCAD), and Leu-103 are in exon 5, Phe-252 is in exon 9, Val-259 is in exon 10; and Tyr-375 is in exon 12. The α-proton abstracting base is Glu-376, which is in exon 12. The transit peptide, which is necessary for import of the MCAD precursor into mitochondria, is encoded by exons 1 and 2. There is, therefore, no obvious correlation of structural or functional domains of the MCAD protein with particular exons.

Missplicing of MCAD mRNA. Recently, we reported the molecular characterization of MCAD deficiency in a Dutch infant and his family (Kelly et al., 1990). The single base change (A to G at position 985) results in substitution of glutamic acid for lysine at amino acid 304 of the mature protein. This mutant protein has markedly reduced activity after expression in bacteria or mammalian cells (Bross et al., 1991). We (D. P. Kelly, A. Whelan, M. Ogden, and A. W. Strauss, unpublished results) have recently determined that an MCAD-deficient American child and two unrelated patients from northern Europe are homozygous for the G^{985} mutation.

During the sequence analyses of cDNA clones isolated from a library made from poly(A)-containing RNA from the liver of the Dutch infant, we noted that, in addition to the G^{985} mutation, three of six cDNA clones contained deletions and insertions, as compared with the normal MCAD mRNA (Figure 4). Similarly, after PCR amplification of hepatic MCAD mRNA from the infant and fibroblast mRNA from his father (who is also homozygous for the G^{985} mutation), we observed numerous deletions and insertions in many subclones. Comparison to the normal MCAD genomic sequences revealed that in *all* of these abnormal mutant cDNAs, the deletions corresponded to loss of exons (Figure 4A). For example, we observed that simple deletion of exon 2 (11/85

```
3'-INTRON SPLICE SITE              EXON                    5'-INTRON SPLICE SITE

                #1          ...92 BP...ATGCTGCAGG  gtgagag     ..~3400 BP..
                                                 * *  *****

aagtgttctttacag #2  GTCCTGAGAA...88 BP...TTTAGTTTTG  gtatatg     ..~3800 BP..
   * **********                                    *  ***  *

tctaaataaccttag #3  AGTTCACCGG...98 BP...AACTGGTGAA  gtaggta     .....71 BP..
***   *  ******                                    *   *** **

ttctttttcttctag #4  TATCCAGTCC...70 BP...GAGAACTGTG  gtaagct     ....468 BP..
***************                                   *  *****

attttgatactgtag #5  GAGGTCTTGG..101 BP...TTCTTTGGGG  gtaagtg     ...~900 BP..
 ****  * ******                                    *  ******

aattttcttcggtag #6  CAAATGCCTA...81 BP...ATTGATGTGT  gtgagta     ..~6000 BP..
  ******** ****                                     ******

ttttatatattcaag #7  GCTTATTGTG..131 BP...AAGCTAATTG  gtatgtg     ..~5600 BP..
**** * * *** **                                   *  *** **

atgtgtatctcttag #8  GTATTTTTTA..109 BP...TGGGAGAAAG  gtaaagt     ..~3500 BP..
 * * * ********                                  ***  ****

caatttttcttattag #9  GAATTAAACA..141 BP...CAGACCTGTA  gtaagta     ....883 BP..
*  ******* ****                                     ******

atcttaaaatactag #10 GTAGCTGCTG...96 BP...ACTTGTAGAG  gtaattt     .~10500 BP..
 * **    * ****                                   **  **** *

tctttttaattctag #11 CACCAAGCAA..249 BP...AATCTATCAG  gtaaggt     ...1550 BP..
*******  ******                                   ***  *****

atatttttcttgcag #12 ATTTATGAAG..807 BP...TATTATCATA   poly A addition site
 * ************

                    CONSENSUS SPLICING SEQUENCES
TTTTTTTTTTTNCAG                                   AAG  GTAAGT
CCCCCCCCCCC T                                     C     G
```

FIGURE 2: Intron–exon boundaries in the structure and organization of the human MCAD gene. The 12 numbered exons are interrupted by 11 introns. The sizes in base pairs of each exon or intron are listed. The nucleotide sequences of the consensus splicing sequences are at the bottom. The sequences of intron–exon junctions are compared with the consensus splicing sequences of the gene and identities to the consensus sequences indicated by asterisks.

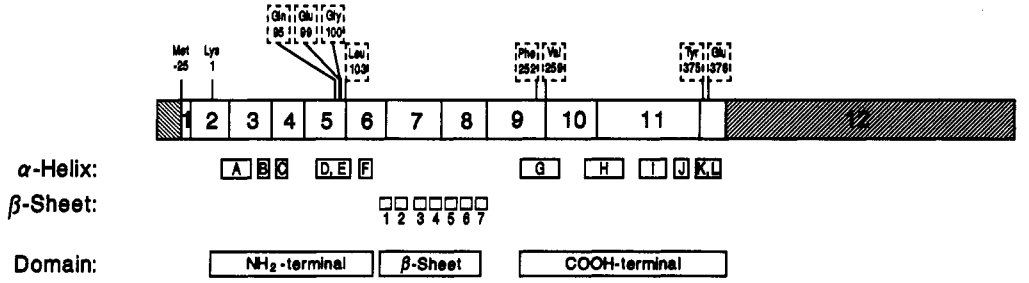

FIGURE 3: Relationship of MCAD gene organization to the MCAD protein structure. The locations of numbered exons in the human MCAD cDNA are indicated on the top line. Untranslated regions are hatched. Above this line, the positions and numbers of essential amino acid residues involved in substrate binding are shown. In the lower portion of the figure, the locations of secondary structural features (α-helices and β-sheets) and the three structural domains are illustrated. The letters denote the regions defined by analysis of the three-dimensional structure of porcine MCAD (Kim & Wu, 1988).

clones) was common. However, deletions of multiple contiguous exons were also noted, particularly of exons 2 through 10 (Figure 4A, line 4). Moreover, multiple noncontiguous exon deletions also occurred; e.g., loss of exon 2, exon 5, and exons 8–10 in a single PCR-derived clone (Figure 4A, line 6). In the exon deletion mutant cDNAs, the normal splicing donor and acceptor sites were used, with one exception (Figure 4, line 5), in which a cryptic acceptor site within exon 11 was employed. This cryptic site (5′-AAATCTTCAG-3′) is 90 bp 3′-ward of the G^{985} point mutation in the normal MCAD gene.

In many instances, the insertions present in mutant MCAD cDNA clones could be explained precisely by retention of entire introns. For example, the 71 bp insertion in one PCR-derived clone (Figure 4A, line 3) is identical to intron 3. Similarly, the 491 bp insertion in a cDNA clone is identical to the 5′ part of intron 11 (not shown). In another cDNA clone, a 12 bp insertion between exons 1 and 2 was present (Figure 4A, line 1). This sequence is identical to the 12 bp which precede exon 2 in the normal MCAD gene (see Figure 2) and must result from aberrant splicing via an alternative acceptor splice site (5′-TTTATTTTAATAAAAG-3′) which immediately precedes the insertion. In another example (Figure 4A, line 2), a 4 bp insertion between exons 3 and 4 of a cDNA clone is identical to the splice donor sequence (GTAG) following exon 3 of the normal MCAD gene, suggesting that this abnormal cDNA resulted from use of an alternative donor splice site (GTATAT) immediately following the 4 bp insert in the normal gene (see Figure 2). In total,

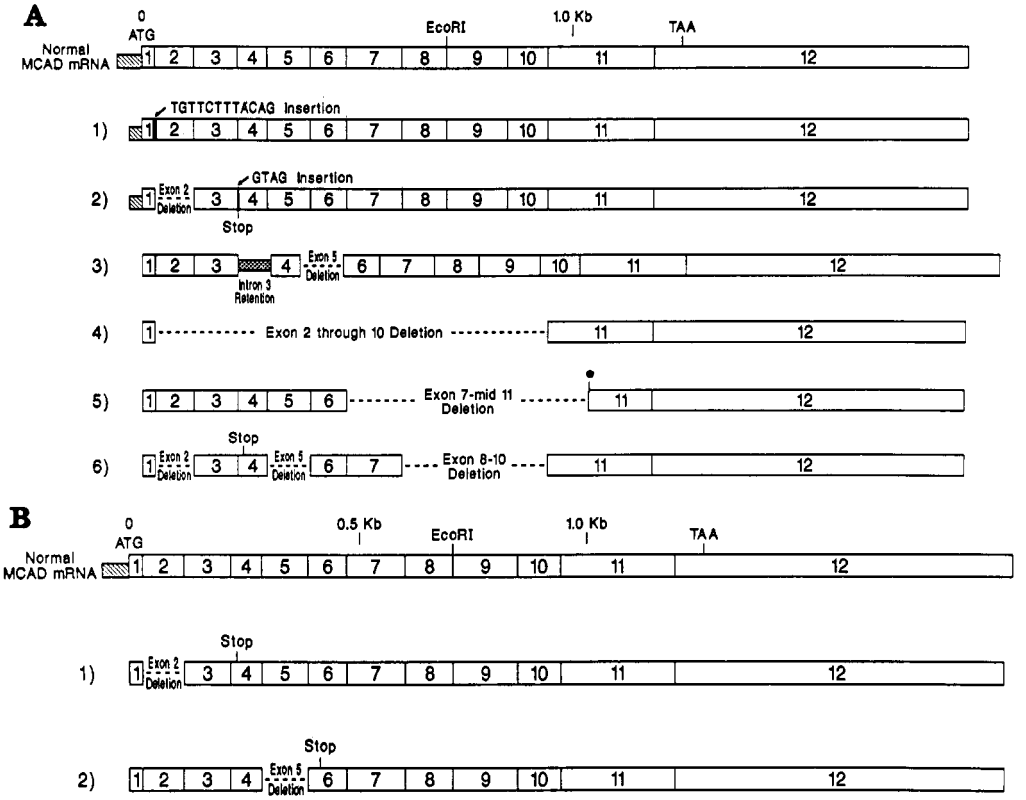

FIGURE 4: Abnormal splicing of MCAD precursor mRNA. The mature MCAD mRNA is illustrated at the top. Numbers in rectangles denote each exon. The translation start codon is indicated as ATG. There is an internal *Eco*RI sites between exons 8 and 9. The TAA stop codon is shown in exon 12. (A) Aberrant splicing of mutant MCAD precursor mRNAs. Six different cDNA clones or PCR-derived subclones generated from hepatic RNA isolated from an MCAD-deficient Dutch patient are illustrated. Exon deletions are schematically indicated as dotted lines and intron retentions as hatched bars. The 12 bp insertion in line 1 and 4 bp insertion in line 2 caused by missplicing (see text) are indicated by solid bars. "Stop" denotes the stop codons which were created by frameshifts due to exon 2 (88 bp) deletion and 4 bp insertion (line 2), and exon 2 deletion (line 6). A cryptic splicing acceptor site is shown as the asterisk within exon 11 (line 5). All clones contain the G^{985} mutation (not shown). (B) Aberrant splicing of normal MCAD pre-mRNA. The symbols are as in Figure 4A.

we analyzed 91 clones from the infant and his MCAD-deficient father. Three of 6 cDNA clones generated from the infant's liver mRNA, 13 of 45 PCR clones derived from the infant's liver mRNA, and 10 of 40 PCR clones from the father's fibroblast mRNA were abnormal, giving a 29% rate of missplicing. Thus, the exon deletions involved many exons and occurred both singly and in combination, resulting in a variety of contiguous and noncontiguous deletions throughout the mRNA coding region. Multiple instances of complete or partial intron retention were also noted. It is important that multiple splicing abnormalities were frequently present within a single clone (Figure 4A, lines 2, 3, and 6).

Most of the misspliced clones would encode truncated proteins with changed amino acid sequences because of altered reading frames. For example, isolated deletion of exon 2 would remove half of the transit peptide coding sequence and alter the reading frame, causing a stop codon to occur within exon 4 (Figure 4A, line 6). The transit peptide is required for import of the MCAD precursor into mitochondria. The 12 bp insertion (line 1) would cause a four amino acid insertion in the middle of the normal MCAD transit peptide.

Our surprising observation of a high rate of aberrant splicing of MCAD precursor mRNA in homozygous mutant individuals prompted us to analyze, in more detail, the MCAD mRNAs from normals (Figure 4B). In RNA derived from different tissues of three normal individuals, 8–10% of PCR-derived MCAD subclones were misspliced. However, only single-exon (either exon 2 or exon 5) deletions existed in these MCAD cDNAs. No intron retentions, multiple exon deletions, combinations of exon deletions and intron retentions, or use of aberrant splice sites were present in normals. These data demonstrate that the homozygous, mutant G^{985} state is associated with a high rate of MCAD pre-mRNA missplicing and that normal individuals have a much lower, but unusual, degree of missplicing.

Because greater than 80% of the splicing abnormalities involved deletion of exon 2, either in isolation or associated with other aberrant splicing, we characterized the splice donor and acceptor sites of the mutant gene. DNA isolated from the MCAD-deficient, Dutch infant was digested with *Eco*RI, fragments of 2–6 kb were selected, and a phage genomic library was constructed in the EMBL vector. Genomic clones were isolated with probes containing exons 1, 2, or 3. DNA sequence analyses of the splice donor, acceptor, and branch-point regions flanking exons 1, 2, and 3 in the mutant gene were identical to our original normal MCAD gene. Southern blot analysis of mutant genomic DNA did not reveal any detectable insertions or deletions in the MCAD gene. Thus, missplicing is not due to mutation in the consensus splicing regions surrounding these exons or to a gross alteration in the mutant gene.

Thus, our detailed structural characterization of the human MCAD gene was essential to prove that the MCAD gene is variably spliced to an unusual extent in both normal and, more dramatically, MCAD-deficient individuals. Complex splicing abnormalities are clearly associated only with the homozygous G^{985} mutation. Gregersen and co-workers (Gregersen et al., 1991) have similarly documented an abnormally high rate of missplicing in five other MCAD-deficient patients from different families. On the basis of the known crystallographic structure of MCAD (Kim & Wu, 1988), the differently spliced mRNAs cannot encode functional enzyme. Thus, this phe-

nomenon truly represents missplicing, not the alternative splicing frequently noted to produce tissue-specific or developmentally regulated isoforms of, for example, muscle contractile proteins. We isolated abnormal MCAD mRNAs both after cDNA cloning of poly(A)-containing RNA and after PCR amplification of total cellular RNA with multiple primers spanning either the entire coding region or shorter segments of the mRNA. Misspliced mRNAs were noted in heart, liver, and fibroblasts. Therefore, MCAD pre-mRNA missplicing is a real phenomenon present in tissues with both high and low levels of MCAD mRNA. We have not observed any unusual features of the normal MCAD gene which might explain this missplicing, although numerous Alu repeats are present, especially in intron 2 and surrounding exons 6 and 10. To our knowledge, this extent of missplicing of the precursor mRNA derived from a normal gene has not previously been documented.

However, many examples of mRNA missplicing associated with human mutations have been reported (Treisman et al., 1983; Padgett et al., 1986). These inherited splicing abnormalities resulted either from mutations in the essential splicing consensus regions or from intronic mutations which create competing cryptic splicing acceptor or donor sites. These mutations usually caused only one species of misspliced mRNA. One other example of multiply misspliced mRNAs has been reported (Ohno & Suzuki, 1988). Several abnormal β-hexosaminidase α-chain mRNAs were documented in a compound heterozygous patient in whom a mutation at the 5′- donor site of an intron was associated with multiple upstream intron retentions and exon deletions, similar to our results in MCAD deficiency. In contrast, however, the mutant MCAD pre-mRNA missplicing is unique because exon skipping is not due to a point mutation in the consensus splice sites and because the high frequency of the aberrant splicing with multiple exon skipping and intron retention occurs throughout the mutant mRNA.

The cause of the greatly accentuated rate of MCAD precursor mRNA missplicing in the homozygous G^{985} point mutation is unclear. Three restriction fragment length polymorphic sites lying within the MCAD gene have recently been described (Blakemore et al., 1990; Kidd et al., 1990). Recently, Kolvraa and co-workers (personal communication) have noted an absolute correlation of the G^{985} mutation with particular alleles of *Ban*II (allele 1), *Pst*I (allele 1), and *Taq*I (allele 2), consistent with a founder effect. We speculate that the genetic background associated with this haplotype of polymorphisms may contain a pre-MCAD mRNA structure which creates vulnerability to missplicing in association with the G^{985} mutation. A second possibility is that another unlinked factor associated with the mutant genetic background acts in trans to accentuate missplicing. However, we have not observed missplicing of other mRNAs in fibroblast mRNA from MCAD-deficient individuals. A third possibility is that accumulation of toxic acylcarnitines or diminished β-oxidation and energy production caused by MCAD deficiency augment missplicing. A fourth potential explanation is that the deficiency increases the MCAD gene transcription rate, thus accentuating pre-MCAD mRNA missplicing.

Localization of the Transcription Start Site of the MCAD Gene. To delineate the transcription initiation site, primer extension was performed with an antisense oligonucleotide primer and human liver total RNA as the template in the presence of [α-^{32}P]dCTP. The major extension product was observed as a band 192 bp upstream of the translation start codon (Figure 5). Upon longer exposure, minor bands within

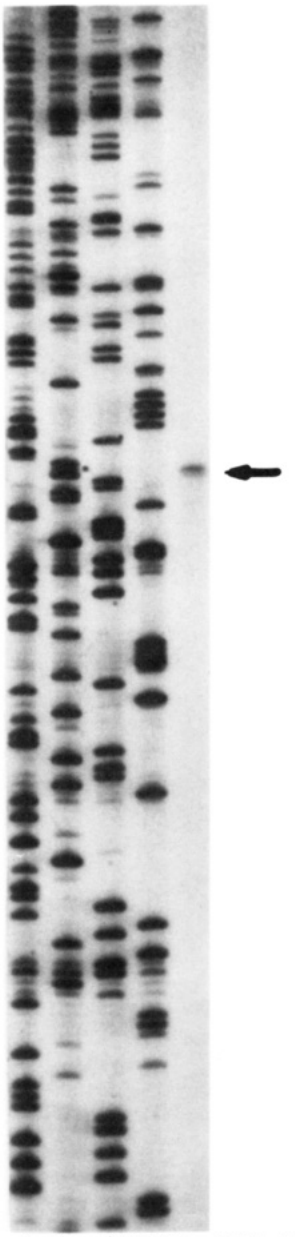

FIGURE 5: Primer extension of MCAD mRNA. The oligonucleotide 5′-GAAATACTTCTCAGGACCCTGCAAGCATCGCC-3′ was annealed to 20 μg of normal total human liver RNA and extended with reverse transcriptase as described under Experimental Procedures. The primer-extended products and dideoxynucleotide sequence reactions were separated on a DNA sequencing gel. The lanes from left to right contain sequencing reactions A, T, G, and C and the primer extension product (far right). The arrow indicates the major primer extension product.

100 bp of this region were also noted and may represent minor alternative transcription start sites (data not shown) of trivial importance in this tissue.

DNA Sequence Analysis of the 5′-Flanking Region of the Human MCAD Gene. To begin the analysis of regulatory elements in the MCAD gene, we determined completely the nucleotide sequence of the 1193 bp upstream of the major transcription start site (Figure 6). Within the putative promoter region of 200 bp upstream of the cap site, the sequence is extremely GC-rich and lacks the prototypical TATA or CAAT boxes. However, this region does contain two Sp1 binding sites typical of promoter regions in "housekeeping genes", protooncogenes, and viral promoters. In addition, four polypyrimidine repeat sequences (CCCTCC) are present, and two are proximate to the Sp1 sites. A dyad sequence ho-

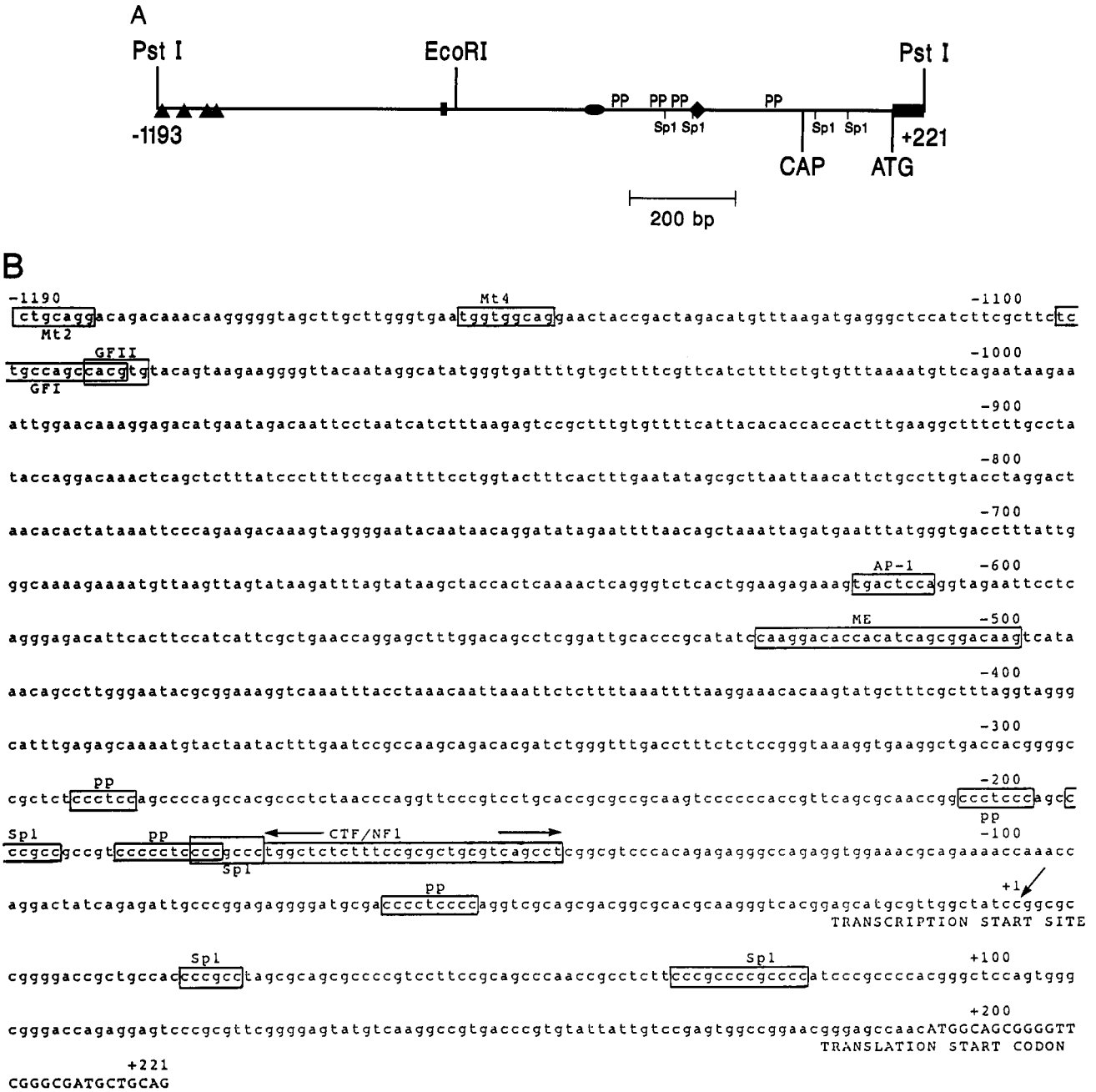

FIGURE 6: 5′-Flanking region of the human MCAD gene. (A) The *Pst*I fragment containing the start codon (ATG), major transcription start site (CAP), and promoter region with putative upstream regulatory elements is illustrated schematically. Putative transcriptional regulatory regions, as defined by DNA sequence homology, are Sp1 binding sites (Sp1), polypyrimidine tracts of CCCTCC (PP), CTF/NF1 binding sites (solid diamond), AP-1 binding sites (short solid bar), and putative mitochondrial enhancers (solid triangles and solid oval). (B) The nucleotide sequence of the 1.4-kb *Pst*I fragment is shown. Numbering is in reference to the major transcriptional start site as +1. The translation start codon, ATG, and the subsequent coding region sequence are in capital letters. The cap site, labeled "TRANSCRIPTION START SITE", is also denoted by the vertical arrow. The boxes surround putative regulatory regions denoted by the abbreviations within each box. The abbreviations are defined in Figure 6A except for ME, GFI, GFII, and Mt4 which indicate different putative mitochondrial enhancer sequences (see text). The horizontal arrows indicate the dyads of the potential CTF/NF-1 sites.

mologous to the CTF/NF-1 binding site is present from −170 to −143 immediately downstream of an Sp1 binding site (Figure 6). The CTF and NF-1 transcription factors bind to CCAAT and similar promoter elements in the several viral genes (Jones et al., 1985; Nagata et al., 1982). CTF/NF-1 sites have also been identified in the genes encoding mouse α-globin (Cohen et al., 1986), the mouse mitochondrial and cytosolic malate dehydrogenases, and cytosolic aspartate aminotransferase (Setoyama et al., 1990), all of which have GC-rich promoters. These data suggest that a factor identical to or similar to CTF/NF-1 may play a role in promoter function of the MCAD and other mammalian genes. A potential binding site for the phorbol ester responsive element, AP-1, is present at −615. Four sequences with homology to retinoic acid responsive elements are also present in this region. These elements may be important in the regulation of MCAD gene expression during cellular differentiation and mitochondrial biogenesis.

Several sequences common to other nuclear genes which encode mitochondrial proteins were also identified in the 5′-flanking region. Tomura and co-workers described an enhancer in the ATP synthase β-subunit gene which is also present in the pyruvate dehydrogenase E1 γ-subunit and cytochrome c_1 genes (Tomura et al., 1990). A sequence homologous to this enhancer exists in the MCAD 5′-flanking region (labeled ME in Figure 6). In addition, we identified

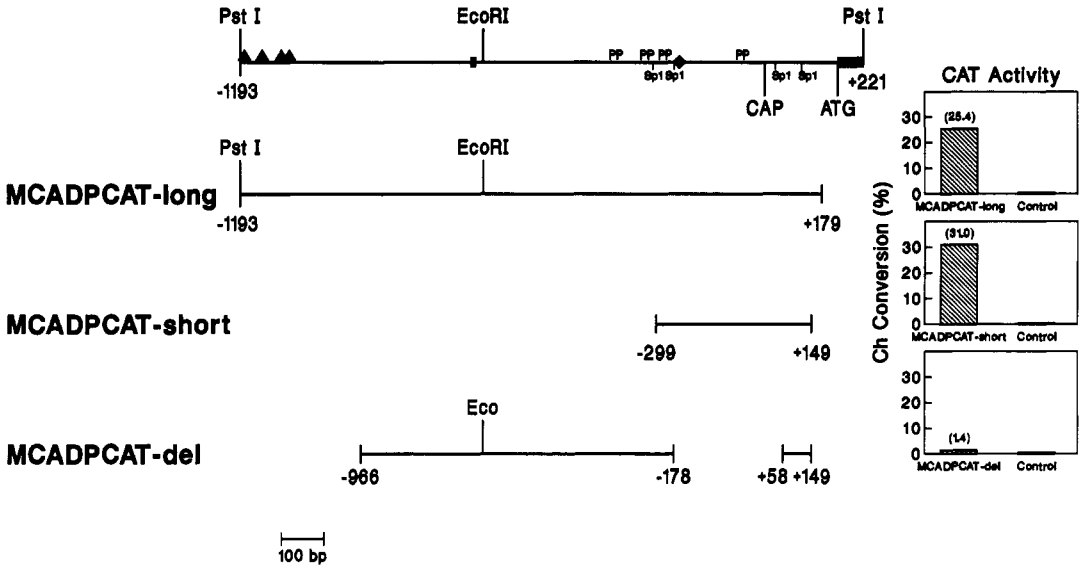

FIGURE 7: Expression of the MCAD promoter–CAT gene chimera. The MCAD–CAT gene constructs MCAD/CAT-long, MCAD/CAT-short, and MCAD/CAT-del are illustrated below the entire *Pst*I 5′-flanking MCAD genomic fragment. Putative regulatory regions are abbreviated as in Figure 6. The constructs were transfected into HepG2 cells. The deletion (−177 to +59) in the construct MCAD/CAT-del is shown. The CAT activity in the transfected cell extracts is indicated to the right of the schematic of each plasmid constrct. The control employed was the pCAT-basic plasmid lacking any promoter or enhancer. The numbers in parentheses indicate the *x*-fold increase in CAT activity as compared to that observed after transfection with the control plasmid.

sequences similar to the GF-I and GF-II binding sites in an overlapping region at bp −1096 to −1082. These factors bind to the 5′-flanking regions of nuclear genes encoding yeast mitochondrial respiratory chain proteins and may participate in the coordinate expression of these genes (Dorsman et al., 1988). Interestingly, sequences homologous to the putative mitochondrial-specific regulatory elements, Mt2 and Mt4 (Suzuki et al., 1990), are present within 100 bp upstream of the GF-I and -II sites. These sequences are present in the 5′-flanking regions of several nuclear genes which encode mitochondrial proteins including cytochrome c_1, F_1-ATPase β-subunit, and ubiquinone binding protein. Thus, the 5′-flanking region of the MCAD gene has several nucleotide sequence homologous which are potentially important in the coordinate expression of nuclear genes encoding mitochondrial proteins.

Transcriptional Activity of MCAD–CAT Gene Chimeric Constructs. To determine if the proposed promoter region of the MCAD gene was sufficient to direct its transcription, a *Pst*I/*Bgl*I fragment containing 1193 bp of DNA upstream of the cap site and 179 bp of 5′-UTR DNA (excluding the translation start codon, ATG) was fused upstream of the chloramphenicol acetyltransferase (CAT) gene in the promoterless pCAT-basic vector (construct MCADCAT-long, Figure 7). This chimeric plasmid was transfected into human HepG2 cells, and CAT assays were performed 48 h later. The transcriptional activity of MCADCAT-long was 25-fold greater than the promoterless control plasmid, pCAT-basic. To determine if the upstream sequences were necessary for the high-level transcriptional activity of MCADCAT-long, a PCR-generated DNA fragment containing bp −299 to +149 was fused to the CAT gene (MCADCAT-short) and transfected into HepG2 cells. The high level transcriptional activity observed with MCADCAT-long was also observed with MCADCAT-short (31-fold over the promoterless plasmid control). Therefore, because bp −299 to bp +149 are sufficient for high-level transcriptional activity in HepG2 cells, the upstream region containing the putative mitochondrial enhancer sequences GFI, GFII, Mt2, and Mt4 and the potential AP-1 site is not necessary for significant MCAD gene promoter activity in these cells.

A third chimeric plasmid, designated MCADCAT-del (Figure 7), was also generated by a PCR designed to amplify a DNA fragment from bp −966 to +149 of the MCAD gene. However, sequence analysis of MCADCAT-del revealed an internal deletion of 236 bp (from bp −177 to +59). The deleted segment contains the transcription start site and flanking DNA. The transcriptional activity of MCADCAT-del was not significantly different than the promoterless control plasmid, providing functional evidence that the MCAD basal promoter active in HepG2 cells lies within the region between −177 and +59. Analysis of the region deleted during the amplification reveals that it can form a stable complex hairpin loop secondary structure (−192 kcal, data not shown). Similar hairpin loops have been predicted in the promoter regions of other "housekeeping" genes encoding mitochondrial proteins, including mitochondrial malate dehydrogenase and aspartate aminotransferase (Takashima et al., 1988; Tsuzuki, 1987). The functional significance of this secondary structure in the transcriptional attenuation of the MCAD gene is under investigation.

Conclusion. We report the characterization and organization of the human MCAD gene. This initial description of a mammalian gene within the β-oxidative pathway will be important in defining DNA sequences required for the coordinate regulation of the genes encoding enzymes in this energy-generating metabolic pathway. Delineation of the genomic structure proved that MCAD precursor mRNA missplicing is the cause of the deletions noted in normal MCAD mRNAs. This unusual degree of missplicing of normal MCAD pre-mRNA is greatly accentuated and much more complex in individuals with homozygous G^{985} mutation, with multiple exon deletions, intron retentions, and use of alternative and cryptic splice sites. Our results show that a GC-rich region within 200 bp upstream of the transcription start site of the MCAD gene is sufficient to confer high-level transcriptional activity in human hepatoma cells and that sequences homologous to putative mitochondrial protein gene enhancers are *not* required for this high-level expression. These studies are an initial step in delineation of the mechanisms

involved in this unusual splicing abnormality and in the regulation of expression of this gene.

Registry No. MCAD, 9027-65-0.

REFERENCES

Beinert, H. (1963) *Enzymes, 2nd Ed. 7*, 447–476.

Blakemore, A. I. F., Engel, P. C., & Curtis, D. (1990) *Nucleic Acids Res. 18*, 2838.

Bourgneres, P. F., Rocchiccioli, F., Kolvraa, S., Hadchouel, M., Lalau-Keraly, J., Chaussain, J. L., Wadman, S. K., & Gregersen, N. (1985) *J. Pediatr. 106*, 918–921.

Bross, P., Krantle, F., Stiemke, J., Ghisla, S., Rasched, I., Gregersen, N., Andresen, B. S., Strauss, A. W., & Kelly, D. P. (1991) in *Flavins and Flavoproteins 1990* (Curti, B., Ronchi, S., & Zanetti, G., Eds.) pp 329–333, de Gruyter, Berlin.

Cohen, R. B., Sheffrey, M., & Kim, C. G. (1986) *Mol. Cell. Biol. 6*, 821–832.

Dorsman, J. C., van Heeswijk, W. C., & Grivell, L. A. (1988) *Nucleic Acids Res. 16*, 7287–7296.

Duran, M., Hofkamp, M., Rhead, W. J., Saudubray, J. M., & Wadman, S. K. (1986) *Pediatrics 78*, 1052–1057.

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.

Gorman, C. (1985) in *DNA Cloning: A Practical Approach* (Glover, D. M., Ed.) Vol. II, pp 143–190, IRL Press, Oxford.

Gregersen, N., Andresen, B. S., Bross, P., Winter, V., Rudiger, N., Engst, S., Christensen, E., Kelly, D., Strauss, A. W., Kolvraa, S., Bolund, L., & Ghisla, S. (1991) *Hum. Genet. 86*, 545–551.

Haas, R. C., & Strauss, A. W. (199) *J. Biol. Chem. 265*, 6921–6927.

Ikeda, Y., Okamura-Ikeda, K., & Tanaka, T. (1985) *J. Biol. Chem. 260*, 1311–1325.

Jones, K. A., Yamamoto, K. R., & Tjian, R. (1985) *Cell 42*, 559–572.

Kaiser, K., & Murray, N. E. (1985) in *DNA Cloning, A Practical Approach* (Glover, D. M., Ed.) Vol. I, pp 1–47, IRL Press, Oxford.

Kelly, D. P., Kim, J.-J., Billadello, J. J., Hainline, B. E., Chu, T. W., & Strauss, A. W. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4068–4072.

Kelly, D. P., Gordon, J. I., Alpers, R., & Strauss, A. W. (1989) *J. Biol. Chem. 264*, 18921–18925.

Kelly, D. P., Whelan, A. J., Ogden, M. L., Alpers, R., Zhang, Z., Bellus, G., Gregersen, N., Dorland, L., & Strauss, A. W. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9236–9240.

Kidd, J. R., Matsubara, Y., Castiglione, C. M., Tanaka, K., & Kidd, K. K. (1990) *Genomics 6*, 89–93.

Kim, J.-J., & Wu, J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6677–6681.

Kolvraa, S., Gregersen, N., Christensen, E., Collet, J. P., Dellamonica, C., & Cottle, J. (1983) *Acta Paediatr. Scand. 72*, 943–949.

Matsubara, Y., Kraus, J. P., Yang-Feng, T. L., Francke, U., Rosenberg, L. E., & Tanaka, K. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6543–6547.

Matsubara, Y., Narisawa, K., Miyabayashi, S., Tada, K., Coates, P. M., Bachmann, C., Elsas, L. J., Pollitt, R. J., Rhead, W. J., & Roe, C. R. (1990) *Biochem. Biophys. Res. Commun. 171*, 498–505.

Nagata, K., Guggenheimer, R. A., Enomoto, T., Lichy, J. H., & Hurwitz, J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6438–6447.

Ohno, K., & Suzuki, K. (1988) *J. Biol. Chem. 263*, 18563–18567.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., & Sharp, P. A. (1986) *Annu. Rev. Biochem. 55*, 1119–1150.

Rhead, W. J., Amendt, B. A., Fritchman, K. S., & Felts, S. J. (1983) *Science 221*, 73–75.

Rosenthal, N. (1987) *Methods Enzymol. 152*, 704–720.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Schare, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487–491.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., pp 9.31–9.58, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Seed, B., & Sheen, J.-Y. (1988) *Gene 67*, 271–277.

Setoyama, C., Ding, S. H., Choudhury, B. K., Joh, T., Takeshima, H., Tsuzuki, T., & Shimada, K. (1990) *J. Biol. Chem. 265*, 1293–1299.

Stanley, C. A., Hale, D. E., Coates, P. M., Hall, C. L., Corkey, B. E., Yang, W., Kelley, R. I., Gonzales, E. L., Williamson, J. R., & Baker, L. (1983) *Pediatr. Res. 17*, 877–884.

Suzuki, H., Hosokawa, Y., Nishikimi, M., & Ozawa, T. (1989) *J. Biol. Chem. 264*, 1368–1374.

Suzuki, H., Hosokawa, Y., Toda, H., Nishikimi, M., & Ozawa, T. (1990) *J. Biol. Chem. 265*, 8159–8163.

Takeshima, H., Joh, T., Tsuzuki, T., Shimada, K., & Matsukado, Y. (1988) *J. Mol. Biol. 200*, 1–11.

Tomura, H., Endos, H., Kagawa, Y., & Ohta, S. (1990) *J. Biol. Chem. 265*, 6525–6527.

Treem, W. R., Witzleben, C. A., Piccoli, D. A., Stanley, C. A., Hale, D. E., Coates, P. M., & Watkins, J. B. (1986) *Hepatology 6*, 1270–1278.

Treisman, R., Orkin, S. H., & Maniatis, T. (1983) in *Globin Gene Expression and Hematopoietic Differentiation* (Stamotoyannopoulos, G., & Niehuis, A. W., Eds.) pp 99–121, Liss, New York.

Tsuzuki, T., Obaru, K., Setoyama, C., & Shimada, K. (1987) *J. Mol. Biol. 198*, 21–31.

Warren, G. (1987) *Nature 327*, 17–18.

Yokota, I., Indo, Y., Coates, P. M., & Tanaka, K. (1990a) *J. Clin. Invest. 86*, 1000–1003.

Yokota, I., Tanaka, K., Coates, P. M., & Ugarte, M. (1990b) *Lancet 336*, 748.