Volume 284, number 1, 34–38 FEBS 09797 June 1991
© 1991 Federation of European Biochemical Societies 00145793/91/$3.50
ADONIS 0014579391005043

Structure of the gene encoding the entire mature E1α subunit of human branched-chain α-keto acid dehydrogenase complex

Nahid Dariush[1], Charles W. Fisher[1], Rody P. Cox[2] and David T. Chuang[1]

Departments of [1]Biochemistry and [2]Internal Medicine, The University of Texas Southwestern Medical Center, Dallas, TX 75235, USA

Received 9 April 1991

We report the isolation of a 22-kb human genomic clone (G7) that contains 8 exons encoding a partial mitochondrial presequence, the entire mature peptide and the complete 3' untranslated region of the E1α mRNA of human branched-chain α-keto acid dehydrogenase complex. Based on this gene structure, exon 9 contains the Tyr^{393} → Asn mutation previously identified in the E1α subunit of Mennonite and other maple syrup urine disease (MSUD) patients. Moreover, the homozygous mutation appears to cause skipping of exon 6 in the mutant E1α transcript. The information on the gene structure for the entire mature E1α subunit will facilitate investigations into the molecular basis of MSUD involving this subunit.

E1α gene structure (human); Maple syrup urine disease; E1α mutation; Exon skipping

1. INTRODUCTION

The mammalian branched-chain α-keto acid dehydrogenase (BCKAD) complex catalyzes the oxidative decarboxylation of the α-keto acids derived from leucine, isoleucine and valine. The mitochondrial multienzyme complex consists of three catalytic components: a branched-chain α-keto acid decarboxylase or E1 comprised two α (M_r = 67000) and two β (M_r = 37500) subunits, a dihydrolipoyl transacylase or E2 with a 24-mer structure (monomer M_r = 46500) forming the core of the enzyme complex, and a dihydrolipoyl dehydrogenase that exists as a homodimer (monomer M_r = 52000) and is a common component shared with pyruvate and α-ketoglutarate dehydrogenase complexes [1]. In addition, the mammalian BCKAD complex contains two regulatory enzymes, a specific kinase and a specific phosphatase that control the activity of the enzyme complex through reversible phosphorylation/dephosphorylation mechanism [2].

Maple syrup urine disease (MSUD) is an autosomal recessive disorder, in which the activity of the BCKAD complex is deficient leading to ketoacidosis, mental retardation and a high mortality among patients. The disease is genetically heterogeneous. Deficiencies in different subunits of the BCKAD complex have been shown using cultured cells from MSUD patients [3,4]. A T⟶A transition resulting in a Tyr⟶Asn missense mutation was described by Zhang et al. to occur in the E1α subunit of a compound heterozygous MSUD patient [5]. The same mutation was later shown to be present in homozygous state in MSUD patients from a Mennonite population [6,7]. To facilitate analysis of mutations in MSUD affecting the E1α subunit, we have undertaken the cloning of the human E1α gene. We report here the gene structure for the entire mature E1α subunit of the human BCKAD complex, and the precise skipping of an exon in the mutant E1α transcript.

Correspondence address: D. Chuang, Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-9038, USA. Fax: (1) (214) 688-8856.

2. MATERIALS AND METHODS

2.1. *cDNA probes*

Partial human E1α cDNA clones used in the present study were described previously [4]. The regions of the human E1α cDNA sequence present in these cDNA clones were as follows: hE1α-1 (0.9 kb), bases 840–1783; hE1α-2 (1.5 kb), bases 191–1783 and hE1α-4 (0.2 kb), bases 1–190. cDNA probes were radiolabeled by the random priming method [8].

2.2. *Genomic screening and subcloning*

A human leukocyte genomic library in λEMBL-3 (Clontech) was screened by plaque hybridization. Plaques producing positive signals were purified, and phage DNA prepared from minilysates as described previously [9]. The purified phage DNA was digested with restriction enzymes, followed by fractionation in agarose gels. Exon-containing fragments were identified by Southern blotting analysis, and subcloned into the Bluescript KS vector (Stratagene).

2.3. *Nucleotide sequencing*

Plasmids containing genomic DNA or cDNA inserts were obtained from the minipreparations of transformed *E. coli* XL-1 Blue cells [9]. Nucleotide sequencing was carried out by the dideoxy chain termination method [10] using T7 DNA polymerase (Pharmacia) on double-stranded templates [11].

2.4. *cDNA synthesis and primers*

Poly(A)$^+$ RNA was isolated from cultured fibroblasts [12] and

Published by Elsevier Science Publishers B.V.

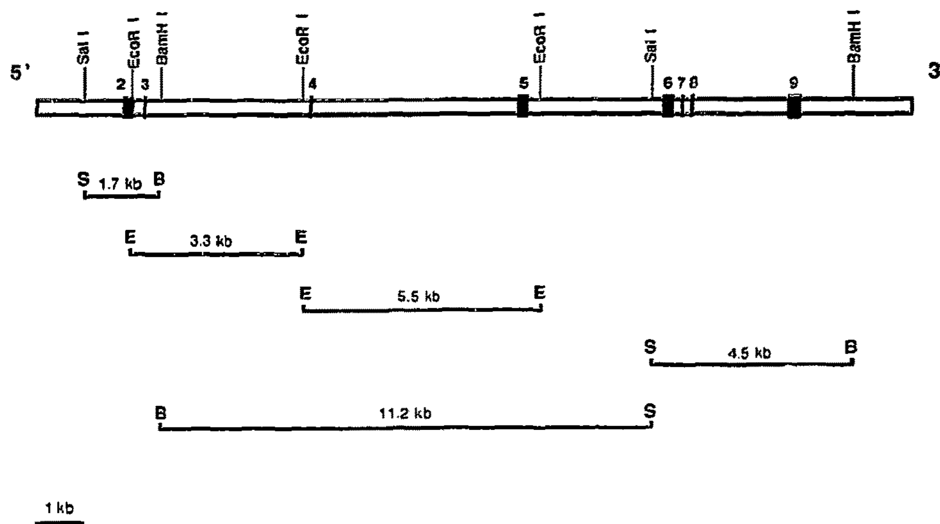

Fig. 1. Restriction map and exon organization of the human E1α gene. The exons (2–9) present in the G7 clone (22 kb) are represented by closed boxes, and numbered after the putative exon 1. Restriction fragments (horizontal bars) were generated with *Sal*I (S), *Eco*RI (E) and *Bam*HI (B).

cDNA synthesized and amplified by the polymerase chain reaction (PCR) [13]. The primers used in the amplification of different regions of the E1α cDNA (5' ⟶ 3') [4] and restriction sites contained were as follows: 1:CGGGGCGATCGATGCAGCGA (*Cla*I) (bases 1–20), 1':CTCATAGAGGATCCGGTCCA (*Bam*HI) (bases 339–358), 2:TGCTGCAGCTCTACAAGAGCA (*Pst*I) (bases 300–320), 2':GTAACAGATGTCGACCCCTGTT (*Sal*) (bases 680–700), 3:ACTGGCCACGCGGATCCCT (*Bam*HI) (bases 618–637), 3':GCTCAGCAGATAGTGCTGCTGCAG (*Pst*I) (bases 1085–1105), 4:GGGTACCACAGCACCAGTGAC (*Kpn*I) (bases 995–1015), 4':TCTCGGGGTACCTGAGGATGG (*Kpn*I) (bases 1356–1376).

3. RESULTS AND DISCUSSION

A total of 2.4 × 10^6 pfu from the EMBL-3 genomic library were screened using hE1α-1 cDNA (bases 840–1783) as a probe. One of the positive clones (G7) purified contained a 22-kb insert for the E1α gene. A restriction map of the entire genomic insert is shown in Fig. 1. Digestion of the G7 clone with *Eco*RI produced gene fragments of 3.3 and 5.5 kb. Hybridization with hE1α-2 cDNA (bases 191–1783) [4], and subsequent nucleotide sequencing showed that these two *Eco*RI fragments in combination contained the 3' end of exon 2 (bases 190 to 283 in the E1α cDNA) and exons 3, 4 and 5. The *Eco*RI site in exon 2 was identical to the one present in the human E1α cDNA [4]. Double digestion of the G7 clone with *Sal*I and *Bam*HI resulted in gene fragments of 1.7, 11.2 and 4.5 kb (Fig. 1). Hybridization with hE1α-4 cDNA (bases 1–190) indicated that the 1.7 kb *Sal*I-*Bam*HI fragment contained exons in the

```
                    350                                          360                                          370
    ValMetGluAlaPheGluGlnAlaGluArgLysProLysProAsnProAsnLeuLeuPheSerAspValTyrGlnGluMet
acagGTGATGGAGGCCTTTGAGCAGGCCGAGCGGAAGCCCAAACCCAACCCCAACCTGCTCTTCTCAGACGTGTATCAGGAGATG    1243

                    380                                          390
ProAlaGlnLeuArgLysGlnGlnGluSerLeuAlaArgHisLeuGlnThrTyrGlyGluHisTyrProLeuAspHisPheAsp
CCCGCCCAGCTCCGCAAGCAGCAGGAGTCTCTGGCCCGCCACCTGCAGACCTACGGGGAGCACTACCCACTGGATCACTTCGAT    1327
400
Lys***
AAGTGAGACCTGCTCAGCCCACCCCCACCCATCCTCAGCTACCCCGAGAGGTAGCCCCACTCTAAGGGGAGCAGGGGGACCTGA    1411

CAGCACACCACTGTCTTCCCCAGTCAGCTCCCTCTAAAATACTCAGCGGCCAGGGCGGCTGCCACTCTTCACCCCTGCTCCTCC    1495

CGGCTGTTACATTGTCAGGGGACAGCATCTGCAGCAGTTGCTGAGGCTCCGTCAGCCCCCTCTTCACCTGTTGTTACAGTGCCT    1579

TCTCCCAGGGGCTGGGTGAGGGCACATTCAGGACTAGAAGCCCCTCTGGGCATGGGGTGGACATGGCAGGTCAGCCTGTGGAAC    1663

TTGCGCAGGTGCGAGTGGCCAGCAGAGGTCACGAATAAACTGCATCTCTGCGCCTGGCTCTCT                         1726
```

Fig. 2. Nucleotide and deduced amino acid sequences of the exons of human E1α gene. Exonic sequences are in capitals, and partial intronic sequences at exon/intron boundaries are in lower case. The approximate sizes of introns (except intron 1) are indicated. The highly conserved region (underlined) flanking the two phosphoserine (asterisk) residues is encoded by exon 8. The complete 3' end of the gene, which includes the polyadenylation signal (AATAAA) (underlined) terminates at base 1726. Nucleotide numbers are according to the human E1α cDNA sequence [4].

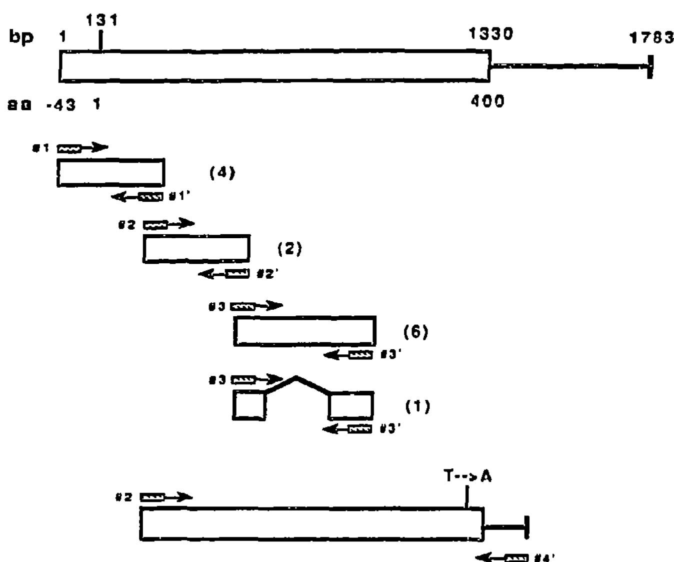

Fig. 3. Subclones of the amplified regions of the E1α cDNA from a Mennonite MSUD patient (P.K.). Open boxes represent the coding region and solid lines the non-coding region of the E1α cDNA. Stippled boxes depict primer pairs used for PCR amplification as described in section 2.4. The number in parentheses indicates the number of subclones isolated and sequenced. The deleted exon 6 is shown by a peak line. The T$\longrightarrow$A conversion at base 1307 is present in the amplified 2–4' region.

5' portion of the gene. Probing with hE1α-1 cDNA showed that the 4.5 kb *Sal*I-*Bam*HI fragment carried exons in the 3' region. Further restriction analysis, subcloning and sequencing established that exons 2 and 3 were located in the 1.7 kb *Sal*I-*Bam*HI fragment and exons 6, 7, 8, and 9 in the 4.5 kb *Sal*I-*Bam*HI fragment. The organization of exons 2 to 9 of the E1α gene is depicted in Fig. 1.

Fig. 2 shows the complete coding sequences for exons 2–9 and exon-intron boundaries of the E1α gene. The 8 exons encode a partial mitochondrial presequence from His-(−9) to Phe-(−1), the entire mature E1α subunit (Ser-1 to Lys-400) and the complete 3' untranslated region including the single polyadenylation signal (AATAAA). The sizes of exon are as follows: exon 2 (180 bases), exon 3 (87 bases), exon 4 (109 bases), exon 5 (106 bases), exon 6 (206 bases), exon 7 (143 bases), exon 8 (172 bases) and exon 9 (564 bases). The introns range from 106 bases (between exons 2 and 3) to 5 kb (between exons 4 and 5) in length. All intronic 5' and 3' splice sites conform to the gt-ag rule [14].

The G7 clone does not contain the genomic sequence encoding the 5' untranslated region of the E1α mRNA that is not present in the cDNA. The upstream mitochondrial presequence from Gly-(−43) to Ala-(−10) in the E1α cDNA is also absent in the genomic clone (Fig. 2). These two regions are putatively encoded by exon 1 by analogy with the gene for the β subunit of human mitochondrial ATP synthase [15]. In the past two years we have isolated and purified approximately 80 additional clones from four different genomic libraries, but none of these clones contained the putative exon 1 and the 5' flanking regions of the E1α gene. The reason for the inability to clone the extreme 5' region is presently unknown. One possibility is that this region contains inverted repeats forming a cruciform structure, which prevents cloning in the λ phage vector [16].

Based on the gene structure for the mature E1α subunit, the homozygous T$\longrightarrow$A transition that causes MSUD in Mennonites [6,7] is present in exon 9. This mutation appears to affect the assembly of the E1α and E1β subunits into an $\alpha_2\beta_2$ structure (Fisher et al., in preparation). Moreover, during amplification of different regions of the mutant E1α cDNA from a Mennonite MSUD patient (P.K.), a deletion that corresponds precisely to exon 6 was found (Fig. 3). The deleted mutant transcript is a minor species, and appears to be a secondary event related to the homozygous primary mutation. The mechanism for this exon deletion in the mutant E1α transcript remains to be elucidated. However, one can speculate that the primary mutation in exon 9 may have disruptive effects on the entire secondary structure of the mutant E1α pre-mRNA. Exon deletions may arise as a result of the failure of the splicing complex to recognize the correct splice site.

Acknowledgements: This work was supported by Grants DK-26758 and DK-37373 from the National Institutes of Health, and Grant 1-1149 from the March of Dimes Birth Defects Foundation.

REFERENCES

[1] Pettit, F.H., Yeaman, S.J. and Reed, L.J. (1978) Proc. Natl. Acad. Sci. USA 75, 4881–4885.
[2] Randle, P.J., Fatania, H.R. and Lau, K.S. (1984) in: Enzyme Regulation by Reversible Phosphorylation-Advances (Cohen, P. ed.) pp. 1–26, Elsevier, Amsterdam, New York.
[3] Danner, D.J. and Elsas, L.J. (1989) in: The Metabolic Basis of Inherited Disease (Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. eds) 6th edn., pp. 671–692, McGraw-Hill, New York.

Fig. 3. Subclones of the amplified regions of the E1α cDNA from a Mennonite MSUD patient (P.K.). Open boxes represent the coding region and solid lines the non-coding region of the E1α cDNA. Stippled boxes depict primer pairs used for PCR amplification as described in section 2.4. The number in parentheses indicates the number of subclones isolated and sequenced. The deleted exon 6 is shown by a peak line. The T$\longrightarrow$A conversion at base 1307 is present in the amplified 2–4' region.

```
 -43      -40                           -30                                 -20
 GlyAlaIleAlaAlaAlaArgValTrpArgLeuAsnArgGlyLeuSerGlnAlaAlaLeuLeuLeuLeuArgGlnProGlyAla
CGGGGCGATCGCTGCAGCGAGGGTCTGGCGGCTAAACCGTGGTTTGAGCCAGGCTGCCCTCCTGCTGCTGCGGCAGCCTGGGGCT      85

               -10                                        1
ArgGlyLeuAlaArgSer                         HisProProArgGlnGlnGlnGlnPheSerSerLeuAspAspLysPro
CGGGGACTGGCTAGATCT........ctcttccccagCACCCCCCCAGGCAGCAGCAGCAGTTTTCATCTCTGGATGACAAGCCC     151

    10                          20                          30
GlnPheProGlyAlaSerAlaGluPheIleAspLysLeuGluPheIleGlnProAsnValIleSerGlyIleProIleTyrArg
CAGTTCCCAGGGGCCTCGGCGGAGTTTATAGATAAGTTGGAATTCATCCAGCCCAACGTCATCTCTGGAATCCCCATCTACCGC     235

        40                     50
ValMetAspArgGlnGlyGlnIleIleAsnProSerGluAspProHis
GTCATGGACCGGCAAGGCCAGATCATCAACCCCAGCGAGGACCCCCACgtgagaggcggcctcccccacttcccgtgcccccca     283


                                                                 LeuProLysGluLysValLeu
cgcccaggcc...../0.1 kb/...ccaactgccccacgtctatctgtgcctccacccgcagCTGCCGAAGGAGAAGGTGCTG     304

  60                         70                          80
LysLeuTyrLysSerMetThrLeuLeuAsnThrMetAspArgIleLeuTyrGluSerGlnArgGln
AAGCTCTACAAGAGCATGACACTGCTTAACACCATGGACCGCATCCTCTATGAGTCTCAGCGGCAGgtgcgtgg..../3.1     370

                                                90                            100
                           GlyArgIlePhePheTyrMetThrAsnTyrGlyGluGluGlyThrHisValGlySerAla
kb/.......ccactccacccccagGGCCGGATCTCCTTCTACATGACCAACTATGGTGAGGAGGGCACGCACGTGGGGAGTGCC     430

                     110
AlaAlaLeuAspAsnThrAspLeuValPheGlyGlnTyrArgGluAlaG
GCCGCCCTGGACAACACGGACCTGGTGTTTGGCCAGTACCGGGAGGCAGgtacgtct...../5.0 kb/......ctcctcccct     479

        120                        130                           140
    lyValLeuMetTyrArgAspTyrProLeuGluLeuPheMetAlaGlnCysTyrGlyAsnIleSerAspLeuGlyLysGly
cctagGTGTGCTGATGTATCGGGACTACCCCCTGGAACTATTCATGGCCCAGTGCTATGGCAACATCAGTGACTTGGGCAAGGGG     559

               150                          160                          170
ArgGlnMetProValHisTyrGlyCysLysGluArgHisPheValThrIleSerSerProLeuAlaThrGlnIleProGlnA
CGCCAGATGCCTGTCCACTACGGCTGCAAGGAACGCCACTTCGTCACTATCTCCTCTCCACTGGCCACGCAGATCCCTCAGGgtg     641

                                                             180
                                        laValGlyAlaAlaTyrAlaAlaLysArgAlaAsnAlaAsnArgVal
aggat...../3.1 kb/.....tctcatcccctgcagCGGTGGGGGCGGCGTACGCAGCCAAGCGGGCCAATGCCAACAGGGTC     668

      190                        200                           210
ValIleCysTyrPheGlyGluGlyAlaAlaSerGluGlyAspAlaHisAlaGlyPheAsnPheAlaAlaThrLeuGluCysPro
GTCATCTGTTACTTCGGCGAGGGGGCAGCCAGTGAGGGGGACGCCCATGCCGGCTTCAACTTCGCTGCCACACTTGAGTGCCCC     772

            220                         230
IleIlePhePheCysArgAsnAsnGlyTyrAlaIleSerThrProThrSerGluGlnTyrArgGlyAspGlyIleA
ATCATCTTCTTCTGCCGGAACAATGGCTACGCCATCTCCACGCCCACCTCTGAGCAGTATCGCGGCGATGGCATTGgtatgggc..     848

                                  240                           250
                                   laAlaArgGlyProGlyTyrGlyIleMetSerIleArgValAspGlyAsnAsp
.../0.3 kb/......tctgtgtccccacagCAGCACGAGGCCCCGGGTATGGCATCATGTCAATCCGCGTGGATGGTAATGAT     901

    260                        270                           280
ValPheAlaValTyrAsnAlaThrLysGluAlaArgArgArgAlaValAlaGluAsnGlnProPheLeuIleGluAlaMetThr
GTGTTTGCCGTATACAACGCCACAAAGGAGGCCCGACGGCGGGCTGTGGCAGAGAACCAGCCCTTTCTCATCGAGGCCATGACC     985

                                                      290    *                    300
TyrAr                                               gIleGlyHisHisSerThrSerAspAspSerSerAlaTyr
TACAGgtgcctgc...../C.2 kb/.....cttgcccctgtgcagGATCGGGCACCACAGCACCAGTGACGACAGTTCAGCGTAC     1030

   *                      310                         320
ArgSerValAspGluValAsnTyrTrpAspLysGlnAspHisProIleSerArgLeuArgHisTyrLeuLeuSerGlnGlyTrp
CGCTCGGTGGATGAGGTCAATTACTGGGATAAACAGGACCACCCCATCTCCCGGCTGCGGCACTATCTGCTGAGCCAAGGCTGG     1114

  330                       340
TrpAspGluGluGlnGluLysAlaTrpArgLysGlnSerArgArgLys
TGGGATGAGGAGCAGGAGAAGGCCTGGAGGAAGCAGTCCCGCAGGAAGgtgagggt...../1.2 kb/.....cccatgtcccc     1162
```

[4] Fisher, C.W., Chuang, J.L., Griffin, T.A., Lau, K.S., Cox, R.P. and Chuang, D.T. (1989) J. Biol. Chem. 264, 3448–3453.
[5] Zhang, B., Edenberg, H.J., Crabb, D.W. and Harris, R.A. (1989) J. Clin. Invest. 83, 1425–1429.
[6] Matsuda, I., Nobukuni, Y., Mitsubuchi, H., Indo, Y., Endo, F., Asaka, J. and Harada, A. (1990) Biochem. Biophys. Res. Commun. 172, 646–651.
[7] Fisher, C.R., Fisher, C.W., Chuang, D.T. and Cox, R.P. (1991) Am. J. Hum. Genet. (submitted).
[8] Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.
[9] Davis, L.G., Dibner, M.D. and Battey, J.F. (1986) Basic Methods in Molecular Biology, Elsevier, New York.
[10] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[11] Chen, E.J. and Seeburg, P.H. (1985) DNA 4, 165–170.
[12] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[13] Saiki, R.K., Gelfand, D.H., Stofeel, S., Scharf, S.J., Highchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science 239, 487–491.
[14] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.
[15] Ohta, S., Tomura, H., Matsuda, K. and Kagawa, Y. (1988) J. Biol. Chem. 263, 11257–11262.
[16] Leach, D.R.F. and Stahl, F.W. (1983) Nature 305, 448–451.