Volume 279, number 2, 229–232 FEBS 09410 February 1991

ADONIS 001457939100118E

Premature termination of transcription and alternative splicing in the human transacylase (E2) gene of the branched-chain α-ketoacid dehydrogenase complex

Kim S. Lau[1], Jun Lee[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 75235-9038, USA

Received 11 December 1990

We have isolated a human genomic clone hgE2-14 containing exons 5, 6, 7 and 8 of the branched-chain α-ketoacid dehydrogenase E2 transacylase gene. Sequencing of exon 8 and its surrounding intronic sequences reveals complete identity with the previously reported truncated E2 cDNA (hE2-1) sequence between nucleotides 938 and 1521. We have identified consensus splice site junctions flanking exon 8 and also a cryptic 3' splice site 370 bases upstream from the start of exon 8 in the gene. In addition, two polyadenylation signals located in the hE2-1 cDNA are also present in the intronic sequence downstream of exon 8 which promote termination of transcription. The data indicate that shortened human liver E2 transcripts undergo alternative splicing to yield mRNA of the hE2-1 type.

Genomic clone; cDNA; E2 gene; Alternative splicing

1. INTRODUCTION

The branched chain α-keto acid dehydrogenase complex (abbreviated branched chain complex) catalyzes the rate-limiting step in the oxidative decarboxylation of the branched-chain α-keto acids derived from leucine, isoleucine and valine. It is a mitochondrial multienzyme complex that is both structurally and mechanistically analogous to the pyruvate and α-ketoglutarate dehydrogenase complexes [1]. The branched-chain complex consists of three catalytic components: a branched-chain α-keto acid decarboxylase or E1, a dihydrolipoyl transacylase or E2 and a dihydrolipoyl dehydrogenase or E3 [2,3]. The E3 component is common to all three α-keto acid dehydrogenase complexes. The branched-chain complex also has two regulatory enzymes, a specific kinase and a specific phosphatase [4–8]. The E2 subunit forms a 24-mer cubic core to which E1, E3, the kinase and the phosphatase are attached through non-covalent interactions. Studies on the human and bovine E2 subunits have revealed that the E2 subunit contains three distinct domains: a lipoyl-bearing domain, an E3-binding domain and an inner-core domain [9,10]. Similar E2 domain structures are also found in other α-keto acid dehydrogenase complexes from both eukaryotic and prokaryotic organisms [11,12].

Correspondence address: K.S. Lau, Department of Biochemistry, The University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-9038, USA

We and others [10,13] have previously isolated a human liver E2 cDNA clone. This clone (hE2-1) is 1577 bp and contains a poly(A)$^+$ tail preceded by 2 polyadenylation signals [10]. The amino acid sequence predicted by its nucleotide sequence codes for the lipoyl-bearing and E3 binding domains of the E2 polypeptide; however, it lacks the inner-core domain of the E2 polypeptide as described later by the isolation of the full-length human E2 cDNA [14]. During the course of cloning the human E2 gene, we isolated a genomic clone (hgE2-14) that contains intronic sequences flanking the eighth exon of the E2 gene which are identical to the nucleotide sequences of hE2-1 bases 938–1521. Within this genomic region is exon 8 consisting of 78 bp coding for Asp-253 to Lys-278 (Lau and Chuang, unpublished results). In addition, two polyadenylation signals are present downstream of exon 8. We propose that premature termination of E2 gene transcription followed by alternative splicing accounts for the occurrence of the truncated hE2-1 mRNA species in the human liver as a normal event.

2. MATERIALS AND METHODS

2.1. *Probe preparation*

A cDNA probe (1436 base pairs) for E2 was prepared from human fibroblast mRNA isolated as described previously [15] using the polymerase-chain reaction (PCR) technique. The primers were designed to allow amplification of the human E2 mRNA from base 1 through to base 1436 [14]. Primer 1: CGCTGCAGTCCGTATGCTGA (5'–3') encoded bases 1–20 and primer 4': CTAGTAGCATAACAGCTGGGT (5'–3') encoded bases 1416–1436 of the human

E2 mRNA. The amplified cDNA product was subcloned into the Bluescript KS vector. Results of nucleotide sequencing established the fidelity of the amplified E2 cDNA. The latter was radiolabeled by the random priming method [16] using [α-^{32}P]dATP and Klenow enzyme.

2.2. *Screening and subcloning*

A human λEMBL-3 genomic library obtained from Stratagene was screened by plaque hybridization using the human E2 cDNA probe. Plaques producing positive signals were purified as described by [17]. Phage DNA was prepared from 10 ml minilysates as described by Davis et al. [17] with the omission of the PEG precipitation step. Restriction mapping of the purified λDNA was performed and Southern analysis carried out to determine fragments containing exon coding sequences [17]. Exon containing genomic DNA inserts were subcloned into the Bluescript KS vector from Stratagene (La Jolla, CA).

2.3. *Nucleotide sequencing*

Plasmids containing amplified E2 cDNA probe or genomic DNA inserts were prepared from minipreps of transformed XL-1 *E. coli* cells [17]. Sequencing was carried out by the dideoxy chain termination method using T7 DNA polymerase (Pharmacia) on denatured double-stranded templates.

3. RESULTS AND DISCUSSION

A total of 5 × 10^5 pfu from the genomic library were plated and screened with the E2 cDNA probe. One of the positive clones purified (hgE2-14) contained a 16.3 kb genomic insert. Within this insert, was a single *Bam*HI restriction site separating the clone into 10.3 kb and 6 kb *Bam*HI-*Sal*I restriction fragments (Fig. 1). Restriction analysis, subsequent subcloning and sequencing of these fragments showed that exons 5, 6 and 7 of the E2 gene were located in the 10.3 kb *Bam*HI-*Sal*I fragment and exon 8 in the 6 kb *Bam*HI-*Sal*I fragment. Using specific sequencing primers based on the hE2-1 cDNA sequence, we sequenced 644 bases in both directions of the genomic region surrounding and including exon 8 (underlined region, Fig. 2). Fig. 3 shows the sequence obtained. Exon 8 (coding for Asp-253–Lys-278) is coded by bases 371–448 and is bounded by the consensus ag-gt 3′–5′ splice sites (points II and III in Figs 2 and 3). Upstream of the ag 3′ splice site (point II) are 353 bases of intronic sequence which are identical to the hE2-1 cDNA sequences 938–1521 (points I–II). Point I is a consensus ag 3′ splice site which is followed immediately by sequence identical to hE2-1 cDNA. Downstream of exon 8 are 153 bases of sequence that are again identical to downstream sequence of hE2-1 point III. Beyond this sequence at point IV there is no consensus gt 5′ splice site in the genomic sequence. However, immediately upstream of point IV are found two polyadenylation signals of the type ATTAAA in the genomic and cDNA sequences.

It was previously suggested by Danner et al. [14] that the presence of noncoding DNA sequences in hE2-1 cDNA might have arisen from failure to correctly remove the intronic sequence from E2 liver transcripts. The results presented here confirm that human liver E2

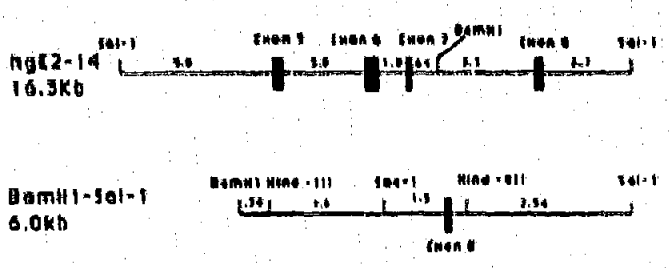

Fig. 1. Diagram of hgE2-14 (16.3 kb) showing position of exons 5–8. The 6.0 kb *Bam*HI-SalI fragment containing exon 8 was subcloned into Bluescript KS vector (Stratagene).

transcripts are capable of producing mRNAs of the hE2-1 type by using the alternative splice site at point I. In Fig. 4, a mechanism is proposed for generation of hE2-1 mRNAs. E2 liver transcripts are synthesized with premature termination of transcription occurring shortly after the two polyadenylation signals. As a result of premature termination of transcription the alternative splice site at point I is used in preference to site II to yield the mRNA containing exons 1–7 correctly spliced and exon 8 attached to upstream intronic sequence. In the prematurely terminated E2 gene transcript, the lack of any consensus 3′ splice site sequence downstream of exon 8 prevents the removal of the intronic sequence (III–IV). The 153 bases of intronic sequence (III–IV) thus remain unspliced in the hE2-1 mRNA (bottom line, Fig. 4). Polyadenylation of the mRNA takes place using the two polyadenylation signals found in the intronic region III–IV. The function of the hE2-1 mRNA and its protein product is not known and remains to be established. The above mechanism explains how the truncated hE2-1 mRNA occurs in normal human liver and adds to the list of eukaryotic genes that also undergo alternative splicing [18].

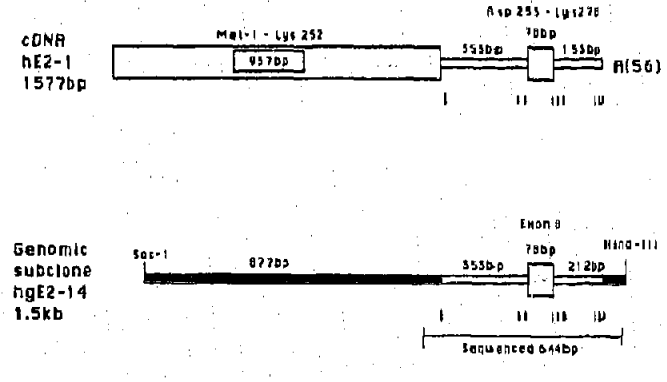

Fig. 2. Schematic representation of hE2-1 cDNA and the 1.5 kb *Sac*I-*Hin*dIII fragment of hgE2-12 genomic clone. Shaded boxes represent exon containing sequences. Thin bar represents intronic DNA. The complete hE2-1 cDNA has been sequenced [10]. The region of hgE2-12 sequenced is underlined. The open white intronic sequence represents identical sequences in hE2-1 and hgE2-14. Points I, II and III contain consensus splice site junctions. Point IV is the point of transcription termination.

```
                                                I
hgE2-12   tttgtgtttgatcctagtattacttctcatggaggtcttaagatcaacatgtcagtgaggaaatcaccta      70
hE2-1                938 tattacttctcatggaggtcttaagatcaacatgtcagtgaggaaatcaccta      990

hgE2-12   gaatttagacattttctttggaacttggctggtctgtatcattatttgaaaatctccttgatgtttaca     141
hE2-1     gaatttagacattttctttggaacttggctggtctgtatcattatttgaaaatctccttgatgtttaca     1061

hgE2-12   agctacctgactactttatctttttattctttaatgattagggtgacctagttccacatggcctgaaggta     211
hE2-1     agctacctgactactttatctttttattctttaatgattagggtgacctagttccacatggcctgaaggta     1131

hgE2-12   acattggcagtaatgtgactaggtacatgtagacttgaaattaaactcagcataaataagagttctct      279
hE2-1     acattggcagtaatgtgactaggtacatgtagacttgaaattaaactcagcataaataagagttctct      1199

hgE2-12   ctctccctccttttatacctcttattgtttcaaattctctatcctttctaagttaaaagtactaaaattgca     351
hE2-1     ctctccctccttttatacctcttattgtttcaaattctctatcctttctaagttaaaagtactaaaattgca     1272

                                    II
hgE2-12   ttctttttctgttaaaacagGCTGCTTCCTTGGGATTACTACAGTTTCCTATCCTTAACGCTTCTGTGGAT     421
hE2-1     ttctttttctgttaaaacagGCTGCTTCCTTGGGATTACTACAGTTTCCTATCCTTAACGCTTCTGTGGAT     1341

                                                III
hgE2-12   GAAAACTGCCAGAATATAACATATAAGgttggctatggaaggagtaaaaatgttcttattattaatag      489
hE2-1     GAAAACTGCCAGAATATAACATATAAGgttggctatggaaggagtaaaaatgttcttattattaatag      1409

hgE2-12   aagatggggcaaatgtgcttgtagattagagacttctattgtacctcaaatagtgatctttttagacta     559
hE2-1     aagatggggcaaatgtgcttgtagattagagacttctattgtacctcaaatagtgatctttttagacta     1479

                                                IV
hgE2-12   gaattaaaggtatggggtaaaacattaaacttaagtctaattcattgttgttgttttgtgttaataata     629
hE2-1     gaattaaaggtatggggtaaaacattaaacttaagtctaattA(56)                           1577

hgE2-12   ctgttttagagattc                                                          644
hE2-1
```

Fig. 3. Nucleotide sequence of hgE2-14 (as underlined in Fig. 2). Consensus splice sites at points I, II and III are indicated. Point IV denotes the transcription termination site. Intronic sequence is in lower case and exon 8 sequence is in upper case. hE2-1 cDNA sequence 938-1521 is presented under the hgE2-14 sequence.

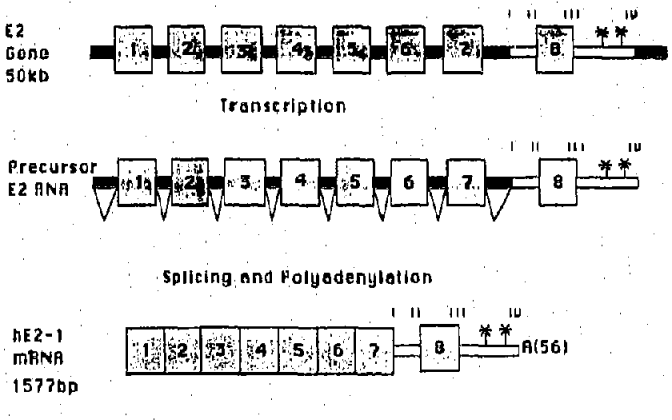

Fig. 4. Alternate splicing of human E2 gene transcripts. Schematic representation of the first 8 exons of the human E2 genes is shown. Consensus splice sites surrounding exon 8 (I, II and III) are shown. The two polyadenylation signals downstream of exon 8 are marked (asterisks) followed by the point of transcription termination IV. The open white intronic regions in the gene are those found in the hE2-1 cDNA. The center figure shows the shortened E2 transcript. Normal splicing of exons 1-7 takes place. The splice site at point I is used alternatively to the splice site of point II. The hE2-1 mRNA is the product of the alternative splicing which is subsequently polyadenylated.

Acknowledgements: This work was supported by Grants DK26758 and DK37373 from the National Institutes of Health, Grant-in-Aid 909-093, American Heart Association, Texas Affiliate and Grant 1-1149 from the March of Dimes Birth Defects Foundation. J.L. was a recipient of a Chilton Foundation Fellowship. We thank Dr R. Padgett for helpful discussions.

REFERENCES

[1] Petit, F.H., Yeaman, S.J. and Reed, L.J. (1978) Proc. Natl. Acad. Sci. USA 75, 4881-4885.
[2] Heffelfinger, S.C., Sewell, E.T. and Danner, D.J. (1983) Biochemistry 22, 5519-5522.
[3] Chuang, D.T., Hu, C.-W.C., Ku, L.S., Niu, W.-L., Myers, D.E. and Cox, R.P. (1984) J. Biol. Chem. 259, 9277-9284.
[4] Odessey, R. (1982) Biochem. J. 204, 353-356.
[5] Lau, K.S., Fatania, H.R. and Randle, P.J. (1982) FEBS Lett. 144, 57-62.
[6] Paxton, R. and Harris, R.A. (1982) J. Biol. Chem. 257, 14433-14439.
[7] Damuni, Z., Merryfield, M.L., Humphreys, J.S. and Reed, L.J. (1984) Proc. Natl. Acad. Sci. USA 81, 4335-4338.
[8] Fatania, H.R., Patston, P.A. and Randle, P.J. (1983) FEBS Lett. 158, 234-238.
[9] Griffin, T.A., Lau, K.S. and Chuang, D.T. (1988) J. Biol. Chem. 263, 14008-14014.

[10] Lau, K.S., Griffin, T.A., Hu, C.-W.C. and Chuang, D.T. (1988) Biochemistry 27, 1972–1981.
[11] Guest, J.S. (1987) FEMS Lett. 44, 417–422.
[12] Reed, L.J. and Hackert, M.L. (1990) J. Biol. Chem. 265, 8971–8974.
[13] Hummel, K.B., Litwer, S., Bradford, A.P., Aitken, A., Danner, D.J. and Yeaman, S.J. (1988) J. Biol. Chem. 263, 6165–6168.
[14] Danner, D.J., Litwer, S., Herring, W.J. and Pruckler, J. (1989) J. Biol. Chem. 264, 7742–7746.
[15] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[16] Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6.
[17] Davis, L.G., Dibner, M.D. and Battey, J.F. (1986) Basic Methods in Molecular Biology, Elsevier, New York.
[18] Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987) Annu. Rev. Biochem. 56, 467–495.