Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 5330–5334, July 1989
Biochemistry

Characterization of cDNAs encoding human pyruvate dehydrogenase α subunit

(deduced amino acid sequence/polymerase chain reaction/S1 nuclease analysis)

LAP HO*, ISAIAH D. WEXLER†, TE-CHUNG LIU*, THOMAS J. THEKKUMKARA*, AND MULCHAND S. PATEL*‡§

Departments of *Biochemistry and †Pediatrics, and ‡Pew Center for Molecular Nutrition, Case Western Reserve University School of Medicine, Cleveland, OH 44106

Communicated by Harland G. Wood, April 28, 1989 (received for review March 17, 1989)

ABSTRACT **A cDNA clone (1423 base pairs) comprising the entire coding region of the precursor form of the α subunit of pyruvate dehydrogenase ($E_1\alpha$) has been isolated from a human liver cDNA library in phage λgt11. The first 29 amino acids deduced from the open reading frame correspond to a typical mitochondrial targeting leader sequence. The remaining 361 amino acids, starting at the N terminus with phenylalanine, represent the mature mitochondrial $E_1\alpha$ peptide. The cDNA has 43 base pairs in the 5′ untranslated region and 210 base pairs in the 3′ untranslated region, including a polyadenylylation signal and a short poly(A) tract. The nucleotide sequence of human liver $E_1\alpha$ cDNA was confirmed by the nucleotide sequences of three overlapping fragments generated from human liver and fibroblast RNA by reverse transcription and DNA amplification by the polymerase chain reaction. This consensus nucleotide sequence of human liver $E_1\alpha$ cDNA resolves existing discrepancies among three previously reported human $E_1\alpha$ cDNAs and provides the unambiguous reference sequence needed for the characterization of genetic mutations in pyruvate dehydrogenase-deficient patients.**

Mammalian pyruvate dehydrogenase complex (PDC) is composed of multiple copies of pyruvate dehydrogenase (E_1; pyruvate:lipoamide 2-oxidoreductase, EC 1.2.4.1), dihydrolipoamide acetyltransferase (E_2), dihydrolipoamide dehydrogenase (E_3), E_1 kinase, phospho-E_1 phosphatase, and a protein, X, of unknown function (1–3). E_1 consists of two nonidentical subunits, α and β, and has a tetrameric ($\alpha_2\beta_2$) structure (1). E_1 catalyzes the irreversible decarboxylation of pyruvate, which is the rate-limiting step in the overall reaction of the complex (1). The activity of E_1 is regulated by the phosphorylation/dephosphorylation of three specific serine residues in the $E_1\alpha$ polypeptide (1). The amino acid sequence of human $E_1\alpha$ has been deduced from the deoxynucleotide sequence of cDNA clones isolated from human fetal liver (4), hepatoma (5), and cultured foreskin fibroblast (6) cDNA libraries. There are major discrepancies among these $E_1\alpha$ cDNA sequences.

Subjects with diminished PDC activity have lactic acidosis and varying degrees of neurologic disability (7, 8). Enzymatic studies of PDC-deficient patients show that the majority of these patients have low levels of E_1 activity but normal levels of E_2 and E_3 activities (8–10). A subset of patients with low E_1 activity may have a defect in E_1 activation (11). Other patients lack one or both of the subunit proteins (10–12). Since many E_1-deficient patients have normal amounts of both E_1 subunit proteins (10), at least one of the two subunits may be catalytically inactive. In many of these patients, there is no definitive way to determine whether the α or β subunit is affected.

Recently our laboratory has begun to analyze $E_1\alpha$ and $E_1\beta$ cDNA sequences of patients with E_1 deficiency. Our analysis has been hampered by the lack of consistency among the published deoxynucleotide sequences for $E_1\alpha$ (4–6). Differences that exist among the published sequences in certain regions compound the difficulty in determining whether the $E_1\alpha$ sequence of E_1-deficient patients at certain sites is normal or mutated. To facilitate the analysis of mutant $E_1\alpha$ cDNAs, we have generated an unambiguous sequence for human $E_1\alpha$ cDNA. The authenticity of this sequence was validated by several techniques including multiple sequencing and restriction endonuclease analysis of independent $E_1\alpha$ cDNA clones, sequence analysis of multiple clones generated by reverse transcription and the polymerase chain reaction (PCR), and S1 nuclease analysis of total human RNA.¶

MATERIALS AND METHODS

Isolation of cDNA Clones. Isolation and partial characterization of a cDNA clone specific for $E_1\alpha$ ($\lambda E_1\alpha 1$) from a human liver cDNA library in bacteriophage λgt11 have been reported (10). A second $E_1\alpha$ cDNA clone ($\lambda E_1\alpha 2$), larger than the $\lambda E_1\alpha 1$ cDNA, was isolated by rescreening the same human liver λgt11 cDNA library with a ^{32}P-labeled oligodeoxynucleotide probe generated from the $\lambda E_1\alpha 1$ cDNA by the random priming method (13).

Sequence Analysis of cDNAs. The two λgt11 recombinants ($\lambda E_1\alpha 1$ and $\lambda E_1\alpha 2$) were digested with *Eco*RI, and the resultant cDNA fragments were subcloned into M13mp19. Overlapping M13mp19 deletion clones were generated by a single-strand directional deletion method (14). Deoxynucleotide sequencing of single-stranded M13mp19 DNA was performed by the dideoxy chain-termination method employing the M13 universal primer and a modified phage T7 DNA polymerase (Sequenase, United States Biochemical) (15). Regions on the two cDNAs that were not conveniently covered by the directional deletion cloning procedure were sequenced by using synthetic oligodeoxynucleotide primers based on the sequence analysis of our $E_1\alpha$ cDNAs. Fig. 1 presents the sequencing strategies as well as a partial restriction endonuclease map of human liver $E_1\alpha$ cDNAs.

Enzymatic Amplification of $E_1\alpha$ RNA by PCR. A human liver specimen was obtained 12 hr after death from a subject with unexplained lactic acidosis in the presence of normal PDC activity. A fresh human foreskin specimen was obtained by circumcision. Total cellular RNA was isolated from human tissue specimens or cultured cells [skin fibroblasts and hepatoma (NPLC) cells] by extraction with guanidinium

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: E_1, pyruvate dehydrogenase; E_2, dihydrolipoamide acetyltransferase; E_3, dihydrolipoamide dehydrogenase; PDC, pyruvate dehydrogenase complex; PCR, polymerase chain reaction.

§To whom reprint requests should be addressed.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. M24848).

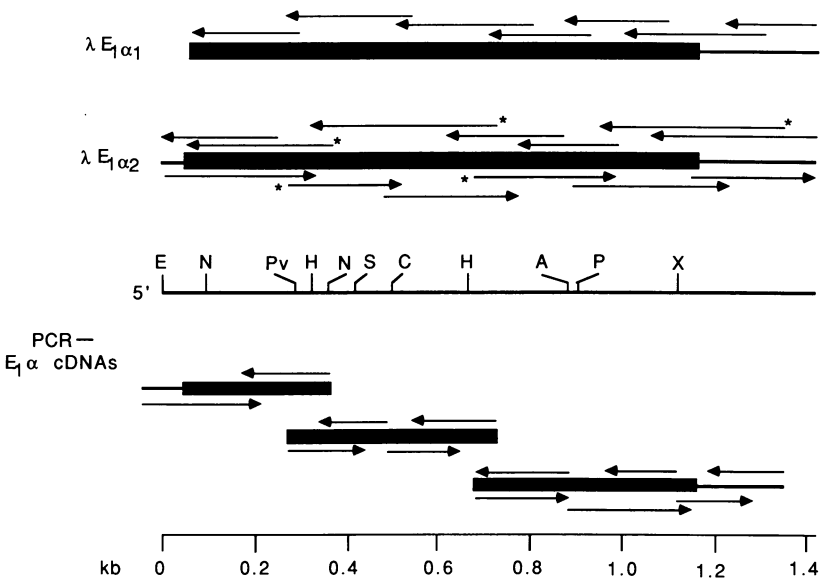

FIG. 1. Sequencing strategies and partial restriction endonuclease map of human $E_1\alpha$ cDNA clones from liver ($\lambda E_1\alpha 1$ and $\lambda E_1\alpha 2$) and skin fibroblasts (PCR-$E_1\alpha$ cDNAs). The three overlapping $E_1\alpha$ cDNA clones were generated from total human liver or skin fibroblast RNA by reverse transcription and PCR. Protein-coding regions are represented by solid boxes and untranslated regions are represented by thin lines. Horizontal arrows indicate the starting position, direction, and extent of sequence determinations. Asterisks indicate sequence analysis using oligodeoxynucleotide primers specific for $E_1\alpha$. Restriction endonuclease sites: A, *Apa* I; C, *Cla* I; E, *Eco*RI; H, *Hin*dIII; N, *Nae* I; P, *Pst* I; Pv, *Pvu* II; S, *Sma* I; X, *Xma* I. The *Eco*RI site represents the piece of linker DNA that was ligated to cDNAs during the generation of the λgt11 library. kb, Kilobases.

isothiocyanate and centrifugation through a cesium chloride gradient (16). One to four micrograms of total RNA was reverse-transcribed in a final volume of 20 μl for 30 min at 42°C in the presence of 13 units of Moloney murine leukemia virus reverse transcriptase (Boehringer Mannheim) and 10 pmol of antisense oligodeoxynucleotide primers complementary to the sequence in $E_1\alpha$ mRNA (17). The reaction mixture was subsequently adjusted to a final volume of 100 μl containing 2.5 units of *Thermus aquaticus* (*Taq*) polymerase and 50 pmol each of the same antisense primer and a specific sense primer (17). Each round of amplification consisted of denaturation at 95°C (1 min), annealing at 42°C (2 min), and primer extension at 72°C (3.5 min). Forty-five to fifty rounds of amplification were performed using a DNA Thermal Cycler (Perkin–Elmer/Cetus). Three sets of sense and antisense primers were utilized in the generation of overlapping cDNA clones from total RNA; the clones spanned the entire coding region as well as portions of the 5′ and 3′ untranslated regions of human $E_1\alpha$ mRNA (Fig. 1). The 5′ → 3′ deoxynucleotide sequences of these primers are identified as bases of the $\lambda E_1\alpha 2$ cDNA sequence in Fig. 2: set 1, antisense primer complementary to bases 1356–1336 and sense primer identical to bases 673–693; set 2, antisense primer complementary to bases 740–717 and sense primer identical to bases 252–272; set 3, antisense primer complementary to bases 397–378 and sense primer identical to a stretch of 21 deoxynucleotides (TCTGCTGGGGCACCTGAAGGA) from the 5′ untranslated region of the human fetal liver $E_1\alpha$ cDNA (4). This sense primer extended the 5′ end of the $\lambda E_1\alpha 2$ cDNA by an additional 27 base pairs (bp). The amplified products were cloned into pBluescript plasmid (Stratagene), and dideoxy DNA sequencing was performed using either double-stranded plasmid DNA or single-stranded phage DNA induced from the pBluescript transformants in the presence of the helper phage M13K07 (18).

S1 Nuclease Protection Assay. The 875-bp *Eco*RI–*Apa* I fragment (bases 1–875) of the $\lambda E_1\alpha 2$ cDNA (Fig. 1) was subcloned into pBluescript. The recombinant plasmid was linearized by digestion with *Xba* I at the pBluescript polylinker region and was used as template in a T3 polymerase reaction in the presence of [α-^{32}P]UTP essentially as described in the Bluescript DNA sequencing system instruction manual (Stratagene). The resultant 925-base labeled antisense RNA transcript (consisting of 875 bases complementary to the $\lambda E_1\alpha 2$ cDNA and 50 bases derived from the vector) was hybridized to total human RNA (19). Single-stranded RNAs, including unhybridized probe or regions on the antisense probe that did not hybridize with human $E_1\alpha$ mRNA, were digested to completion with S1 nuclease (Boehringer Mannheim; 510 units at 45°C for 45 min) (19). The protected RNA products were resolved in a denaturing 8.3 M urea/4% polyacrylamide gel and radioactive bands were visualized by autoradiography.

RESULTS AND DISCUSSION

Characterization of Human $E_1\alpha$ cDNAs. The isolation and partial characterization of an $E_1\alpha$ cDNA clone ($\lambda E_1\alpha 1$) from the human λgt11 cDNA library were reported previously (10). The complete nucleotide sequence analysis of this cDNA revealed an open reading frame encoding a peptide that is identical to the first 20 amino acids from the N terminus of purified bovine heart $E_1\alpha$ (20) and that contains the three phosphorylation sites of porcine heart $E_1\alpha$ (21). However, the translation initiation codon ATG was not present in the $\lambda E_1\alpha 1$ cDNA. A larger $E_1\alpha$ cDNA clone ($\lambda E_1\alpha 2$) was therefore isolated by rescreening the human liver λgt11 library with a ^{32}P-labeled $\lambda E_1\alpha 1$ cDNA probe generated by the random priming reaction. Nucleotide sequence analysis showed that the overlapping regions of $\lambda E_1\alpha 1$ and $\lambda E_1\alpha 2$ cDNAs were identical (Fig. 1).

The complete nucleotide sequence of human liver $E_1\alpha$ cDNA ($\lambda E_1\alpha 2$) as well as the deduced amino acid sequence of the $E_1\alpha$ prepeptide are presented in Fig. 2. The $\lambda E_1\alpha 2$ cDNA contains 1423 bp, including 13 adenines at the 3′ end of the clone. The largest open reading frame, starting at base 44, comprises 1170 nucleotides and encodes an $E_1\alpha$ prepeptide of 390 amino acid residues (Fig. 2). Starting at base 53 within the same open reading frame is a codon for a second potential translation initiation methionine. Translation starting at base 44 or 53 would give an $E_1\alpha$ prepeptide with a calculated molecular mass of 43,246 or 42,831 daltons, respectively. Both these values approximate the estimated molecular mass of 44,500 daltons for $E_1\alpha$ prepeptide from cultured porcine kidney cells, based on sodium dodecyl sulfate/polyacrylamide gel electrophoresis (22). However, nucleotide sequences surrounding either of these potential translation initiation sites do not conform to the hypothesized characteristics for eukaryotic translation initiation (23). Presently, it is not possible to determine which of the two is the translation initiation site *in vivo*. The N-terminal amino acid of the mature $E_1\alpha$ peptide (mitochondrial form) is phenylalanine (amino acid 1 in Fig. 2) based on the amino acid sequence of bovine heart $E_1\alpha$ (20). If translation starts at base

```
CTGGGTTGTGAGGAGTCGCCGCTGCCGCCACTGCCTGTGCTTC ATG AGG AAG ATG CTC GCC GCC GTC TCC      70
                                              M   R   K   M   L   A   A   V   S     -21
                                             -29

CGC GTG CTG TCT GGC GCT TCT CAG AAG CCG GCA AGC AGA GTG CTG GTA GCA TCC CGT AAT     130
 R   V   L   S   G   A   S   Q   K   P   A   S   R   V   L   V   A   S   R   N      -1
                                                                                   -1

TTT GCA AAT GAT GCT ACA TTT GAA ATT AAG AAA TGT GAC CTT CAC CGG CTG GAA GAA GGC     190
 F   A   N   D   A   T   F   E   I   K   K   C   D   L   H   R   L   E   E   G      20
 1

CCT CCT GTC ACA ACA GTG CTC ACC AGG GAG GAT GGG CTC AAA TAC TAC AGG ATG ATG CAG     250
 P   P   V   T   T   V   L   T   R   E   D   G   L   K   Y   Y   R   M   M   Q      40

ACT GTA CGC CGA ATG GAG TTG AAA GCA GAT CAG CTG TAT AAA CAG AAA ATT ATT CGT GGT     310
 T   V   R   R   M   E   L   K   A   D   Q   L   Y   K   Q   K   I   I   R   G      60

TTC TGT CAC TTG TGT GAT GGT CAG GAA GCT TGC TGT GTG GGC CTG GAG GCC GGC ATC AAC     370
 F   C   H   L   C   D   G   Q   E   A   C   C   V   G   L   E   A   G   I   N      80

CCC ACA GAC CAT CTC ATC ACA GCC TAC CGG GCT CAC GGC TTT ACT TTC ACC CGG GGC CTT     430
 P   T   D   H   L   I   T   A   Y   R   A   H   G   F   T   F   T   R   G   L     100

TCC GTC CGA GAA ATT CTC GCA GAG CTT ACA GGA CGA AAA GGA GGT TGT GCT AAA GGG AAA     490
 S   V   R   E   I   L   A   E   L   T   G   R   K   G   G   C   A   K   G   K     120

GGA GGA TCG ATG CAC ATG TAT GCC AAG AAC TTC TAC GGG GGC AAT GGC ATC GTG GGA GCG     550
 G   G   S   M   H   M   Y   A   K   N   F   Y   G   G   N   G   I   V   G   A     140

CAG GTG CCC CTG GGC GCT GGG ATT GCT CTA GCC TGT AAG TAT AAT GGA AAA GAT GAG GTC     610
 Q   V   P   L   G   A   G   I   A   L   A   C   K   Y   N   G   K   D   E   V     160

TGC CTG ACT TTA TAT GGC GAT GGT GCT GCT AAC CAG GGC CAG ATA TTC GAA GCT TAC AAC     670
 C   L   T   L   Y   G   D   G   A   A   N   Q   G   Q   I   F   E   A   Y   N     180

ATG GCA GCT TTG TGG AAA TTA CCT TGT ATT TTC ATC TGT GAG AAT AAT CGC TAT GGA ATG     730
 M   A   A   L   W   K   L   P   C   I   F   I   C   E   N   N   R   Y   G   M     200
                                                                 ==========

GGA ACG TCT GTT GAG AGA GCG GCA GCC AGC ACT GAT TAC TAC AAG AGA GGC GAT TTC ATT     790
 G   T   S   V   E   R   A   A   A   S   T   D   Y   Y   K   R   G   D   F   I     220
 ==================

CCT GGG CTG AGA GTG GAT GGA ATG GAT ATC CTG TGC GTC CGA GAG GCA ACA AGG TTT GCT     850
 P   G   L   R   V   D   G   M   D   I   L   C   V   R   E   A   T   R   F   A     240

GCT GCC TAT TGT AGA TCT GGG AAG GGG CCC ATC CTG ATG GAG CTG CAG ACT TAC CGT TAC     910
 A   A   Y   C   R   S   G   K   G   P   I   L   M   E   L   Q   T   Y   R   Y     260
                                                                         ===

CAC GGA CAC AGT ATG AGT GAC CCT GGA GTC AGT TAC CGT ACA CGA GAA GAA ATT CAG GAA     970
 H   G   H   S   M   S   D   P   G   V   S   Y   R   T   R   E   E   I   Q   E     280
 ==========================================

GTA AGA AGT AAG AGT GAC CCT ATT ATG CTT CTC AAG GAC AGG ATG GTG AAC AGC AAT CTT    1030
 V   R   S   K   S   D   P   I   M   L   L   K   D   R   M   V   N   S   N   L     300

GCC AGT GTG GAA GAA CTA AAG GAA ATT GAT GTG GAA GTG AGG AAG GAG ATT GAG GAT GCT    1090
 A   S   V   E   E   L   K   E   I   D   V   E   V   R   K   E   I   E   D   A     320

GCC CAG TTT GCC ACG GCC GAT CCT GAG CCA CCT TTG GAA GAG CTG GGC TAC CAC ATC TAC    1150
 A   Q   F   A   T   A   D   P   E   P   P   L   E   E   L   G   Y   H   I   Y     340

TCC AGC GAC CCA CCT TTT GAA GTT CGT GGT GCC AAT CAG TGG ATC AAG TTT AAG TCA GTC    1210
 S   S   D   P   P   F   E   V   R   G   A   N   Q   W   I   K   F   K   S   V     360

AGT TAA GGG GAG GAG AAG GAG AGG TTA TAC CTT CAG GGG GCT ACC AGA CAG TGT TCT CAA    1270
 S                                                                                 361

CTT GGT TAA GGA GGA AGA AAA CCC AGT CAA TGA AAT TCA ATG AAA TTC TTG GAA ACT TCC    1330

ATT AAG TGT GTA GAT TGA GCA GGT AGT AAT TGC ATG CAG TTT GTA CAT TAG TGC ATT AAA    1390

AGA TGA ATT ATT GAG TGC TTA AAA AAA AAA AAA                                        1423
```

FIG. 2. Deoxynucleotide and deduced amino acid sequences of human liver $E_1\alpha$ ($\lambda E_1\alpha 2$) cDNA. Nucleotides and deduced amino acids (in single-letter codes) are numbered on the right. If translation is initiated at the first methionine, the 29 amino acids of the deduced signal peptide are amino acids −29 to −1. The broken underline identifies the 20 amino acids that matched perfectly with a sequence of 20 amino acids that we have derived, starting with the N-terminal residue, from purified bovine heart $E_1\alpha$ (10). Two peptides containing the three phosphorylation sites are identified by double underlining. A polyadenylylation signal is underlined.

44 of the $\lambda E_1\alpha 2$ cDNA, the $E_1\alpha$ leader peptide is composed of 29 amino acids, and the mature human $E_1\alpha$ peptide contains 361 amino acids with a calculated molecular mass of 40,183 daltons. This is consistent with the estimated molecular mass of 41,000 daltons for bovine kidney $E_1\alpha$ (1).

The accuracy of the human liver $E_1\alpha$ cDNA sequence shown in Fig. 2 was verified by (*i*) single-strand deoxynucleotide sequencing of the $\lambda E_1\alpha 2$ cDNA in both directions and (*ii*) sequence identity between $\lambda E_1\alpha 1$ and $\lambda E_1\alpha 2$ cDNAs in the overlapping region (1362 bp). In addition, we independently isolated human $E_1\alpha$ cDNA clones that were derived by reverse transcription of total RNA isolated from cultured human skin fibroblasts as well as a human liver specimen from an E_1-deficient patient (10) in the presence of specific antisense $E_1\alpha$ primers followed by DNA amplification using the PCR. Deoxynucleotide sequence analyses of the overlapping $E_1\alpha$ cDNA clones from this patient revealed that the composite human $E_1\alpha$ cDNA sequence was identical to normal human liver $E_1\alpha$ cDNA shown in Fig. 2, indicating that the mutation is not present in the $E_1\alpha$ mRNA of this patient. The consistency of our findings over independent cDNA clones that were derived from different cDNA cloning paradigms and from different tissues argues for the authenticity of the human $E_1\alpha$ cDNA sequence as presented in Fig. 2.

To date, three cDNA sequences for normal human $E_1\alpha$ have been published: a human fetal liver cDNA (4), a human hepatoma cDNA (5), and a cultured human foreskin fibroblast cDNA (6). The sequences of all three cDNAs are significantly different from one another (Fig. 3). Discrepancies ranging from single base alterations (such as insertions, deletions, and substitutions) to major deletions resulting in large reading-frame shifts of up to 64 amino acids as well as a deletion of 93 consecutive bases are distributed at 14 locations in the coding and 3′ untranslated regions (Fig. 3; refs. 4–6). Since the three published $E_1\alpha$ cDNA sequences differ from one another, it was not possible to accurately identify the DNA sequence of normal human $E_1\alpha$ mRNA. Now, with the addition of information we have obtained on the sequence of human liver and skin fibroblast $E_1\alpha$ cDNAs, we are able to resolve the discrepancies among the published $E_1\alpha$ cDNA sequences.

The most prominent discrepancy among the three published sequences of human $E_1\alpha$ cDNA is the absence from the sequence reported for foreskin fibroblast cDNA (6) of a stretch of 93 nucleotides that was present in the human fetal liver and hepatoma $E_1\alpha$ cDNAs (Fig. 3; refs. 4 and 5) and in our human liver $E_1\alpha$ cDNA (Fig. 2, bases 555–647). The loss of 31 amino acids due to the 93-base deletion was compensated in the foreskin fibroblast cDNA by two base insertions (both are guanine, at bases 1120 and 1153 of $\lambda E_1\alpha 2$ in Fig. 2) that caused two reading-frame shifts in the deduced sequence of the 31 C-terminal amino acids in Fig. 2 as well as bypassing

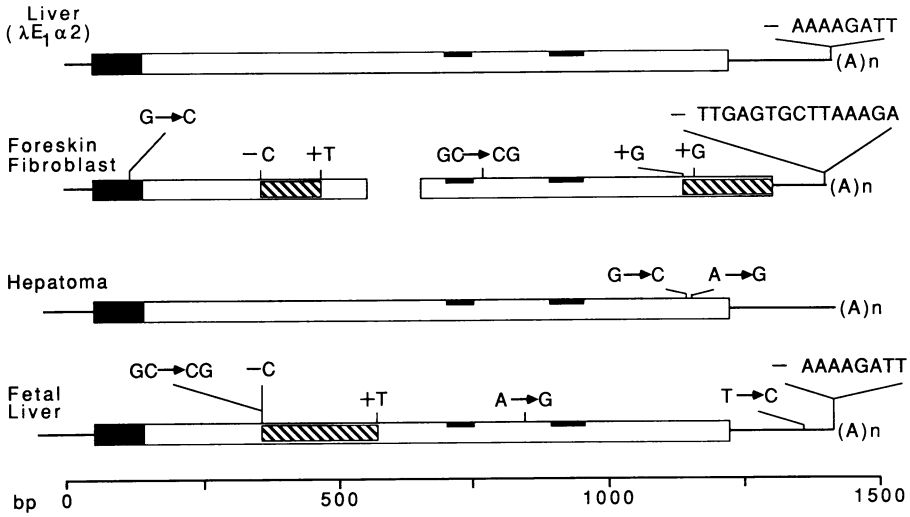

FIG. 3. Schematic representation of the three published human $E_1\alpha$ cDNAs (4–6) as well as our $\lambda E_1\alpha 2$ cDNAs. Boxes represent coding region. Untranslated regions are represented by thin lines. Solid boxes indicate sequence encoding the leader peptide. Hatched boxes indicate altered deduced amino acid sequences due to a change in the reading frame. Heavy lines indicate the two short amino acid sequences containing the three phosphorylation sites. Base insertions and deletions are identified (+ and –, respectively). Discontinuity in the human foreskin fibroblast cDNA (6) represents the absence of 93 consecutive nucleotides.

of the stop codon at bases 1214–1216. Translation in the foreskin fibroblast cDNA would continue to base 1310 for an additional 33 amino acids (6), thus maintaining the overall size of the mature peptide (Fig. 3).

The 93-base deletion found in the cultured foreskin fibroblast $E_1\alpha$ cDNA (6) may reflect the presence of a human $E_1\alpha$ isozyme. We previously reported (24) a case of E_1 deficiency in which the defect was expressed in the patient's liver, heart, and muscle but was only partially expressed in kidney and was not expressed in cultured skin fibroblasts. It is therefore possible that tissue-specific expression of E_1 isozymes may be the basis of this observation. To determine whether the 93-base deletion found in the foreskin $E_1\alpha$ cDNA is present in other human cell types, we surveyed human tissue specimens and cultured cells by two independent experimental approaches. In one set of experiments, a ^{32}P-labeled 925-base antisense RNA probe (complementary to bases 1–875 of the $\lambda E_1\alpha 2$ cDNA plus 50 bases from the vector) was used for S1-protection analysis of total RNA isolated from a human liver specimen, human hepatoma (NPLC) cells, human skin fibroblasts, and human glioblastoma cells (A172). Total RNA from all these specimens protected a sequence of ≈900 bp in the probe (Fig. 4), indicating that the 93-base region is present in the $E_1\alpha$ mRNA from these human cells.

In a separate set of experiments, we amplified this 93-base region from human foreskin $E_1\alpha$ mRNA and analyzed the products by gel electrophoresis (Fig. 5). The stretch of 489 nucleotides corresponding to bases 252–740 of $\lambda E_1\alpha 2$ cDNA (Fig. 2) was cloned and amplified from total RNA of different cell types by using specific $E_1\alpha$ primers (primer set 2 in *Materials and Methods*), reverse transcriptase, and PCR. Absence of this 93-base segment would have resulted in amplification of a 396-bp cDNA. A cDNA fragment of ≈500 bp was amplified from total RNA from cultured human fetal lung fibroblasts, human liver, and foreskin specimens (Fig. 5), indicating that these 93 bases were present in this region of human $E_1\alpha$ cDNA. It is therefore very unlikely that this 93-base deletion represents normal human $E_1\alpha$ mRNA. The large reported deletion (6) probably either resulted from a cloning artifact or was peculiar to the foreskin fibroblasts used in generating the cDNA library.

Another major discrepancy among the three published nucleotide sequences is that, relative to the human hepatoma cDNA (5), the human fetal liver $E_1\alpha$ cDNA contained a set of single-base deletion and insertion (at, respectively, bases 361 and 566 of $\lambda E_1\alpha 2$ in Fig. 2), which introduced a reading-frame shift altering the deduced sequence of 68 amino acids (Fig. 3; ref. 4). Our analysis of $\lambda E_1\alpha 2$ and $\lambda E_1\alpha 1$ cDNAs (Fig. 2) and PCR-generated human liver $E_1\alpha$ cDNAs (results not shown) is in agreement with the nucleotide sequence reported in this region for human hepatoma $E_1\alpha$ cDNA (5).

Among the three published cDNA sequences for human $E_1\alpha$ (4–6), our results (Fig. 2) most closely resemble the sequence reported for the hepatoma $E_1\alpha$ cDNA (ref. 5; Fig. 3). However, our human liver $E_1\alpha$ cDNA sequence differs from the hepatoma sequence at two sites: a guanine at base 1088 and an adenine at base 1103 in Fig. 2 are replaced by a cytosine and a guanine, respectively, in the hepatoma cDNA (5), resulting in two amino acid substitutions, alanine to proline and alanine to threonine, respectively (Fig. 2). The deoxynucleotide sequence of human liver and skin fibroblast cDNAs at these two locations agrees with two other reported sequences (4, 6). In the hepatoma $E_1\alpha$ cDNA, these two bases are part of the recognition sequences for the restriction endonucleases *Bam*HI and *Sac* II, respectively. Consistent with our sequence analysis of human liver $E_1\alpha$ cDNA clones,

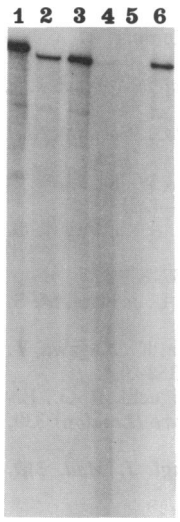

FIG. 4. S1-protection analysis of human $E_1\alpha$ mRNA. An evenly labeled 925-base antisense RNA probe complementary to bases 1–875 of $\lambda E_1\alpha 2$ was annealed to total RNA (50 μg) isolated from various human cell types. After digestion of single-stranded RNA with S1 nuclease, the protected RNA was resolved in a denaturing urea/4% polyacrylamide gel and radioactive bands were visualized by autoradiography. Lanes: 1, labeled probe without added S1 nuclease; 2, human liver; 3, cultured human hepatoma (NPLC); 4, cultured human glioblastoma (A172); 5, cultured skin fibroblasts from an $E_1\alpha$-mRNA-deficient patient (10); 6, cultured skin fibroblasts from a normal subject. In lane 1, the slightly slower migration of the undigested probe reflects the presence of 50 bases derived from pBluescript. In lanes 4 and 5, low signal intensities reflect a low abundance of $E_1\alpha$ mRNA in these cell lines. Total RNA from all cell types protected an ≈900-bp fragment of the labeled probe.

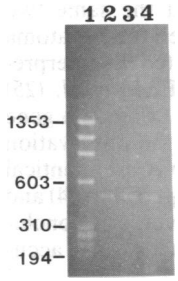

FIG. 5. Agarose gel electrophoresis of amplified cDNA prepared from total RNA isolated from various human cell types. $E_1\alpha$ cDNAs were generated from total RNA by reverse transcription and PCR in the presence of two oligodeoxynucleotide primers (primer set 2, *Materials and Methods*; ref. 17). The amplified products were resolved in a 1% agarose gel and stained with ethidium bromide. Lanes: 1, markers (*Hae* III digest of φX174 replicative-form DNA; lengths indicated in bases); 2, human foreskin; 3, human liver; 4, cultured human hepatoma (NPLC).

restriction digestion analyses of $\lambda E_1\alpha 2$ cDNA with *Bam*HI or *Sac* II confirmed the absence of recognition sites for these two enzymes (data not shown). The two bases in Fig. 2 therefore accurately represent the human $E_1\alpha$ mRNA. Although it is plausible that single-base substitutions at these two locations in the hepatoma $E_1\alpha$ cDNA may represent polymorphism, it is unlikely that two independent polymorphic events are simultaneously present in one $E_1\alpha$ cDNA but are absent from the other six separate $E_1\alpha$ cDNAs ($\lambda E_1\alpha 2$ and $\lambda E_1\alpha 1$ cDNAs isolated from the human liver λgt11 library; cDNAs generated by PCR from cultured skin fibroblast RNA from an E_1-deficient patient; also see refs. 4, 6, and 25). It is possible that these two base changes either are peculiar to the hepatoma cell line used in the isolation of the previously reported hepatoma $E_1\alpha$ cDNAs (5) or are the result of a cloning artifact.

Relevance to the Identification of Genetic Defects. We have presented evidence from different experimental approaches to confirm the human $E_1\alpha$ cDNA sequence (Fig. 2). $E_1\alpha$ cDNA was originally isolated in our laboratory for the purpose of characterizing human E_1 mutations. In the majority of cases of E_1 deficiency, it is not possible to identify mutations specific to either the $E_1\alpha$ or $E_1\beta$ peptide by simple analysis of enzymatic activity or protein or mRNA profiles. Another approach for the identification of the mutations in E_1 subunits is direct nucleotide sequence analysis of patient-specific E_1 mRNAs or genes. Presently, direct analysis of specific mRNA appears to be the method of choice for analysis of human E_1 mutations, in light of the recent advances in the application of the PCR (26, 27). However, this approach for localization of mutation(s) requires the availability of deoxynucleotide sequences that accurately represent normal human $E_1\alpha$ and $E_1\beta$ mRNAs. We have therefore made every attempt to resolve all the discrepancies that have been observed between our human $E_1\alpha$ deoxynucleotide sequence and the previously published sequences in order to present an accurate basis for future analysis of $E_1\alpha$ in both normal and diseased states.

The application of the PCR in the analysis of $E_1\alpha$ mRNA from one of our E_1-deficient patients has provided an independent sequence of human $E_1\alpha$ mRNA. Skin fibroblasts from this patient have low E_1 activity but normal activities for E_2 and E_3 (10). Analysis of this patient's $E_1\alpha$ cDNAs generated by reverse transcription and PCR (results not shown) revealed that the composite deoxynucleotide sequence of the patient's $E_1\alpha$ cDNA is identical to the reported hepatoma $E_1\alpha$ cDNA except for the two deoxynucleotides corresponding to bases 1088 and 1103 of the $\lambda E_1\alpha 2$ cDNA (Fig. 2) as discussed above. However, based on comparison between the patient's $E_1\alpha$ cDNA sequence and the $E_1\alpha$ sequence of our human liver cDNA $\lambda E_1\alpha 2$, we ruled out the possibility of an $E_1\alpha$ defect in this patient. Had we relied on the published human hepatoma $E_1\alpha$ cDNA sequence, we could have made the interpretation that this patient's $E_1\alpha$ mRNA was mutated at these two nucleotides. Similarly, in another E_1-deficient patient recently reported by Endo *et al.* (25), a deletion of four nucleotides upstream from the termination codon resulted in a reading-frame shift generating a new termination codon at the 33rd codon downstream from the normal termination site. The additional differences at the same two bases (bases 1088 and 1103 in Fig. 2) between the hepatoma and patient $E_1\alpha$ cDNAs apparently complicated the interpretation of the mutation in this patient (25). Endo *et al.* (25) attributed these differences to heterogeneity of the $E_1\alpha$ gene unrelated to the patient's defect, based on the observation that these two bases from the patient's cDNA are identical with that reported for the human fetal liver $E_1\alpha$ cDNA (4) and the foreskin fibroblast $E_1\alpha$ cDNA (6). However, this conclusion was based on two cDNA sequences that do not accurately reflect the human $E_1\alpha$ mRNA. The unambiguous human $E_1\alpha$ cDNA sequence in Fig. 2 eliminates the possibility of multiple $E_1\alpha$ mutations in this patient. The availability of the human $E_1\alpha$ cDNA sequence as a reference and the application of the PCR technique should facilitate the analysis of genetic mutations in the entire coding region of the human $E_1\alpha$ mRNA.

We thank Drs. T. Chandra and S. L. C. Woo (Baylor College of Medicine, Houston) for providing the λgt11 human liver cDNA library, Beth Kleinhenz for excellent technical assistance, Dr. D. S. Kerr (Rainbow Babies and Children Hospital, Cleveland) for the human liver specimen, Dr. R. Ortez-Reyes (Institute of Pathology, Cleveland) for the human foreskin specimen, and Drs. D. Samols, W. Merrick, and D. S. Kerr for their critical reading of the manuscript. This work was supported by Public Health Service Grant DK20478 and Metabolism Training Grant AM07319 (L.H. and I.D.W.).

1. Reed, L. J. (1974) *Acc. Chem. Res.* **7**, 40–46.
2. Yeaman, S. J. (1986) *Trends. Biochem. Sci.* **11**, 293–296.
3. Rahmatullah, M. & Roche, T. E. (1987) *J. Biol. Chem.* **262**, 10265–10271.
4. Dahl, H.-H. M., Hunt, S. M., Hutchison, W. M. & Brown, G. K. (1987) *J. Biol. Chem.* **262**, 7398–7403.
5. DeMeirleir, L., MacKay, N., Lam, H. W. A. M. & Robinson, B. H. (1988) *J. Biol. Chem.* **263**, 1991–1995.
6. Koike, K., Ohta, S., Urata, Y., Kagawa, Y. & Koike, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 41–45.
7. Blass, J. P. (1983) in *Inborn Errors of Pyruvate Metabolism*, eds. Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L. & Brown, M. S. (McGraw-Hill, New York), pp. 193–203.
8. Robinson, B. H., MacMillan, H., Petrova-Benedict, R. & Sherwood, W. G. (1987) *J. Pediatr.* **111**, 525–533.
9. Butterworth, R. F. (1985) in *Cerebral Energy Metabolism and Metabolic Encephalopathy*, ed. McCandless, D. W. (Plenum, New York), pp. 121–141.
10. Wexler, I. D., Kerr, D. S., Ho, L., Lusk, M. M., Pepin, R. A., Javed, A. A., Mole, J. E., Jesse, B. W., Thekkumkara, T. J., Pons, G. & Patel, M. S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7336–7340.
11. Wicking, C. A., Scholem, R. D., Hunt, S. M. & Brown, G. K. (1986) *Biochem. J.* **239**, 89–96.
12. Brown, G. K., Scholem, R. D., Hunt, S. M., Harrison, J. R. & Pollard, A. C. (1987) *J. Inherited Metab. Dis.* **10**, 359–366.
13. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
14. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13**, 31–40.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Chirgwin, J. W., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
17. Kawasaki, E. S., Clark, S. S., Coyne, M. Y., Smith, S. D., Champlin, R., Witte, O. N. & McCormick, F. P. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5698–5702.
18. Mead, D. A., Szczesna-Skorupa, E. & Kemper, B. (1986) *Protein Eng.* **1**, 67–74.
19. Weaver, R. F. & Weissmann, C. (1979) *Nucleic Acids Res.* **7**, 1175–1193.
20. Lawson, R., Aitken, A. & Yeaman, S. J. (1983) *Biochem. Soc. Trans.* **11**, 298–299.
21. Sugden, P. H., Kerbey, A. L., Randle, P. J., Waller, C. A. & Reid, K. B. M. (1979) *Biochem. J.* **181**, 419–426.
22. DeMarcucci, O. G. L., Gibb, G. M., Dick, J. & Lindsay, J. G. (1988) *Biochem. J.* **251**, 817–823.
23. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
24. Kerr, D. S., Berry, S. A., Lusk, M. M., Ho, L. & Patel, M. S. (1988) *Pediatr. Res.* **22**, 312–318.
25. Endo, H., Hasegawa, K., Narisawa, K., Tada, K., Kagawa, Y. & Ohta, S. (1989) *Am. J. Hum. Genet.* **44**, 358–364.
26. Wong, C., Dowling, C. E., Saiki, R. K., Higuchi, R. G., Erlich, H. A. & Kazazian, H. H. (1987) *Nature (London)* **330**, 384–386.
27. Moller, D. E. & Flier, J. S. (1988) *N. Engl. J. Med.* **319**, 1526–1529.