Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 1422–1426, March 1988
Biochemistry

Cloning and cDNA sequence of the dihydrolipoamide dehydrogenase component of human α-ketoacid dehydrogenase complexes

(λgt11 expression library/protein evolution/mitochondrial import leader sequence)

GABRIEL PONS*, CINDY RAEFSKY-ESTRIN, DONNA J. CAROTHERS, RONALD A. PEPIN†, ALI A. JAVED‡, BARRY W. JESSE§, MAHRUKH K. GANAPATHI, DAVID SAMOLS, AND MULCHAND S. PATEL¶

Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, OH 44106

Communicated by Frederick C. Robbins, November 2, 1987 (received for review September 9, 1987)

ABSTRACT **cDNA clones comprising the entire coding region for human dihydrolipoamide dehydrogenase (dihydrolipoamide:NAD$^+$ oxidoreductase, EC 1.8.1.4) have been isolated from a human liver cDNA library. The cDNA sequence of the largest clone consisted of 2082 base pairs and contained a 1527-base open reading frame that encodes a precursor dihydrolipoamide dehydrogenase of 509 amino acid residues. The first 35-amino acid residues of the open reading frame probably correspond to a typical mitochondrial import leader sequence. The predicted amino acid sequence of the mature protein, starting at the residue number 36 of the open reading frame, is almost identical (>98% homology) with the known partial amino acid sequence of the pig heart dihydrolipoamide dehydrogenase. The cDNA clone also contains a 3′ untranslated region of 505 bases with an unusual polyadenylylation signal (TATAAA) and a short poly(A) track. By blot-hybridization analysis with the cDNA as probe, two mRNAs, 2.2 and 2.4 kilobases in size, have been detected in human tissues and fibroblasts, whereas only one mRNA (2.4 kilobases) was detected in rat tissues.**

The α-ketoacid dehydrogenase complexes—the pyruvate dehydrogenase multienzyme complex, the α-ketoglutarate dehydrogenase multienzyme complex, and the branched-chain α-ketoacid dehydrogenase multienzyme complex—catalyze the oxidative decarboxylation of pyruvate, α-ketoglutarate, and the branched-chain α-ketoacids, respectively (1–4). These three complexes occupy key positions in energy metabolism and are involved (*i*) in connecting glycolysis with the Krebs cycle, (*ii*) in the Krebs cycle itself, and (*iii*) in the regulation of the oxidation of branched-chain amino acids, respectively. Each complex consists of at least three catalytic components. The decarboxylase component, which has been designated E_1, is specific for each complex, and it most likely represents the rate-limiting step in the overall reaction. The dihydrolipoamide acyltransferase component, designated E_2, also substrate specific, is involved in the transfer of the acyl group to CoA. The third component of each complex, dihydrolipoamide dehydrogenase (dihydrolipoamide:NAD$^+$ oxidoreductase, EC 1.8.1.4), designated E_3, is considered to be common to all three complexes. E_3, a homodimer (subunit M_r = 55,000), contains a noncovalently bound molecule of FAD per subunit and catalyzes the reoxidation of the disulfhydryl form of the lipoyl residue bound to the E_2 molecules of the complexes:

$$E_2\text{-Lip-(SH)}_2 + NAD^+ \overset{E_3}{\rightleftarrows} E_2\text{-Lip(S)}_2 + NADH + H^+.$$

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.

Three lines of evidence indicate that the E_3 present in these three complexes is identical; they are reconstitution experiments (5), immunological cross-reactivity (6–9), and genetic disorders of E_3, which cause simultaneous increases in the blood levels of pyruvate, α-ketoglutarate, and branched-chain ketoacids (10). The glycine cleavage system also contains a dihydrolipoamide dehydrogenase component known as L protein (11). However, this system has not been purified as an intact complex. We have recently suggested that this E_3 (i.e., L protein) is immunologically different from the E_3 common to the three α-ketoacid dehydrogenase complexes (12).

Approximately half of the amino acid sequence of pig heart E_3 has been previously determined by sequencing overlapping peptide fragments (13). The availability of a cDNA for mammalian E_3 would be valuable to complete the primary amino acid sequence of mammalian E_3. It also would be useful for comparison with the primary structure of other related flavoproteins and allow the isolation of the E_3 gene to study its structure–function relationships. In this study we present the cDNA sequence‖ of a 2.1-kilobase (kb) E_3 clone comprising a short 5′ untranslated region, a region covering the complete predicted amino acid sequence of the precursor E_3, and a 3′ untranslated region of 505 bases. A preliminary report of this work has been presented (14).

MATERIALS AND METHODS

Production of Antisera. Two different E_3 preparations were used: (*i*) E_3 separated from highly purified bovine kidney pyruvate dehydrogenase complex (a gift from Lester J. Reed, University of Texas, Austin) and (*ii*) E_3 purified from porcine heart (Sigma) without first isolating any one of the complexes. The polyclonal antisera (12, 15) were shown to be specific for E_3 based on: (*i*) immunoprecipitation of E_3 from rat liver mitochondrial extracts (12) and (*ii*) identification of a single protein band (corresponding to purified E_3) in human fibroblast extracts by immunoblot analysis (16). The antibodies were affinity-purified by adsorption to a column of Sepharose 4B to which purified pig heart E_3 had been coupled by following the manufacturer's recommendations.

Abbreviation: E_3, dihydrolipoamide dehydrogenase.

*Present address: Department of Biochemistry, University of Barcelona, Barcelona-36, Spain.
†Present address: Interferon Sciences, Inc., New Brunswick, NJ 08901.
‡Present address: Department of Cell Physiology, Boston Biomedical Research Institute, Boston, MA 02114.
§Present address: Department of Animal Sciences, Cook College, Rutgers University, New Brunswick, NJ 08873.
¶To whom reprint requests should be addressed.

‖This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03620).

The antibody was further passed through a second column to remove possible antibodies against *Escherichia coli* and coliphage proteins (17).

Isolation of E_3 cDNA Clones. A human liver cDNA library constructed in the expression vector λgt11, consisting of 2×10^6 independent recombinant phage (provided by T. Chandra and S. L. C. Woo, Baylor College of Medicine, Houston) (18), was plated on *E. coli* Y1090. Approximately 300,000 recombinant plaques were immunologically screened with a 1:50 dilution of the affinity-purified E_3 antiserum raised against bovine kidney E_3. Phage plaques corresponding to E_3 clones were detected by using a goat anti-rabbit IgG horseradish peroxidase-conjugated antibody and a horseradish peroxidase developing reagent, both from Bio-Rad. The cDNA inserts were isolated from the phage λ DNA by the plate lysate method (19) and, after digestion with *Eco*RI, were subcloned into the phage M13mp19 (20).

Nucleic Acid Methods and DNA Sequence Analysis. A 20-mer oligodeoxynucleotide mixture (128 possible combinations), corresponding to an amino acid sequence (Glu-Met-Ile-Asn-Glu-Ala-Ala) near the carboxyl terminus of pig heart E_3 (13), was synthesized by using an Applied Biosystems (Foster City, CA) automated synthesizer (model 380A). This sequence (with the change of valine for isoleucine) corresponds to the amino acid residues 427 to 433 in the human E_3 sequence presented in Fig. 2. Radiolabeling of the purified mixture was carried out in the presence of [γ-^{32}P]ATP and T4 polynucleotide kinase (19).

RNA was isolated from human tissues and fibroblasts and from rat tissues by the guanidine thiocyanate method (21). Poly(A)-containing RNA was prepared by oligo(dT)-cellulose chromatography (22). RNA blots were performed as described by Hod *et al.* (23) by using phage M13 primer-extended single-stranded probes labeled according to Davis *et al.* (24). Southern blotting was performed using GeneScreen membranes (New England Nuclear), and the blots were hybridized according to the manufacturer's recommendations.

The DNA sequence was analyzed by the M13 dideoxy sequencing method (25) with adenosine 5′-[α-(^{35}S)thio]triphosphate and a M13 sequencing kit from Bethesda Research Laboratories. M13 clones for sequencing were generated by subcloning restriction enzyme digests into M13mp18 or M13mp19. Restriction enzymes, T4 ligase, and DNA polymerase (Klenow fragment) were purchased from Boehringer Mannheim and used in accordance with the manufacturer's instructions.

RESULTS

Isolation of E_3 cDNA Clones. Screening of a human liver cDNA library with an antiserum against bovine kidney E_3 resulted in the isolation of two clones (λE_3-1 and λE_3-2) that appeared positive on rescreening. Moreover, these two clones reacted positively with a second antibody raised against a preparation of porcine heart E_3 and gave negative results when preimmune sera were used. The cDNA insert sizes of these two clones, obtained by *Eco*RI digestion, were approximately 1.1 and 2.1 kb and they cross-hybridized with each other (results not shown). λE_3-2 showed an internal *Eco*RI site almost in the middle of the clone producing two fragments of about 1.1 and 1 kb in size. Only λE_3-2 showed a strong hybridization with the oligodeoxynucleotide mixture (results not shown). The relationship between these two clones was established by DNA sequence analysis and is shown diagramatically in Fig. 1.

cDNA Sequences and Predicted Amino Acid Sequence of E_3. Restriction mapping and sequence analysis of clone λE_3-1 were shown to be identical to the 5′ portion of clone λE_3-2. The DNA sequencing strategy and restriction endonuclease map of these cDNA clones are shown in Fig. 1, and the composite cDNA sequence obtained is presented in Fig. 2. The internal *Eco*RI site of clone λE_3-2 was mapped at position 1057 and marks the end of clone λE_3-1.

The primary amino acid sequence of human precursor E_3 was deduced from the cDNA sequence. There is an open reading frame of 1527 bp located in the sequence between nucleotides 50 and 1577 (Fig. 2). This corresponds to a protein of 509 amino acids and a calculated molecular mass of 54,214 daltons. The first 35 amino acids of the open reading frame are absent from the amino terminus of pig heart E_3 and presumably represent the leader sequence of the protein. The calculated molecular mass of the mature polypeptide is 50,216 daltons. When the molecular mass of a FAD molecule (Da = 703) is added, the total molecular mass comes to 50,919 daltons. This value is close to those obtained by sodium dodecyl sulfate electrophoresis, giving a subunit molecular mass of about 55,000 daltons (6, 7, 26). The calculated molecular mass of the leader sequence (4069) is larger than those obtained for rat heart (3000; ref. 7) and ox heart (2000; ref. 26), although these data are based on differences between the precursor and the mature forms of E_3. Only about 50% of the primary amino acid sequence of pig heart E_3 has been determined (13, 27). The comparison of the sequences of the first 88 residues at the amino terminus and the last 162 residues at the carboxyl terminus of the mature E_3 shows about 98% homology between human liver

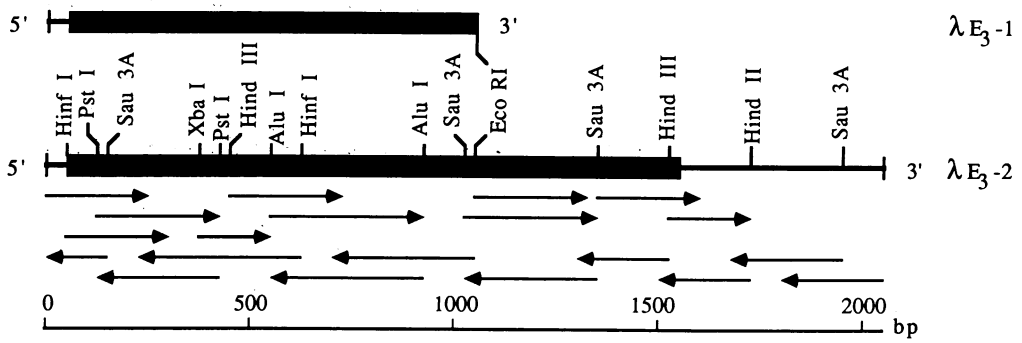

FIG. 1. Relationship between the two E_3 cDNA clones and sequencing strategy. The λE_3-1 clone corresponds to the fragment between the 5′ end of the λE_3-2 clone and the internal *Eco*RI site. The restriction endonuclease map of the λE_3-2 clone and sequencing strategy are shown. For *Alu* I and *Hin*fI, only the restriction sites used for sequencing are shown. The solid box represents the coding region, starting 51 nucleotides from the 5′ end. The direction and extension of the sequence obtained from different restriction fragments used to compile the cDNA sequence are indicated by horizontal arrows. The double-stranded sequence was obtained for every region of the cDNA except a region of 94 bases between *Alu* I (at 950) and *Sau*3A (at 1044) and a region from *Hin*dII to the end of the 3′ noncoding region, which were sequenced at least four times. bp, Base pairs.

```
                                GCTCCCAGCGGAGGTGAAAGTATTGGCGGAAAGGAAAATACAGCGGAAAA      50

ATG CAG AGC TGG AGT CGT GTG TAC TGC TCC TTG GCC AAG AGA GGC CAT TTC AAT CGA ATA     110
Met Gln Ser Trp Ser Arg Val Tyr Cys Ser Leu Ala Lys Arg Gly His Phe Asn Arg Ile     -16
-35

TCT CAT GGC CTA CAG GGA CTT TCT GCA GTG CCT CTG AGA ACT TAC GCA GAT CAG CCG ATT     170
Ser His Gly Leu Gln Gly Leu Ser Ala Val Pro Leu Arg Thr Tyr Ala Asp Gln Pro Ile       5

GAT GCT GAT GTA ACA GTT ATA GGT TCT GGT CCT GGA GGA TAT GTT GCT GCT ATT AAA GCT     230
Asp Ala Asp Val Thr Val Ile Gly Ser Gly Pro Gly Gly Tyr Val Ala Ala Ile Lys Ala      25

GCC CAG TTA GGC TTC AAG ACA GTC TGC ATT GAG AAA AAT GAA ACA CTT GGT GGA ACA TGC     290
Ala Gln Leu Gly Phe Lys Thr Val Cys Ile Glu Lys Asn Glu Thr Leu Gly Gly Thr Cys      45

TTG AAT GTT GGT TGT ATT CCT TCT AAG GCT TTA TTG AAC AAC TCT CAT TAT TAC CAT ATG     350
Leu Asn Val Gly Cys Ile Pro Ser Lys Ala Leu Leu Asn Asn Ser His Tyr Tyr His Met      65

GCC CAT GGA AAA GAT TTT GCA TCT AGA GGA ATT GAA ATG TCC GAA GTT CGC TTG AAT TTA     410
Ala His Gly Lys Asp Phe Ala Ser Arg Gly Ile Glu Met Ser Glu Val Arg Leu Asn Leu      85

GAC AAG ATG ATG GAG CAG AAG AGT ACT GCA GTA AAA GCT TTA ACA GGT GGA ATT GCC CAC     470
Asp Lys Met Met Glu Gln Lys Ser Thr Ala Val Lys Ala Leu Thr Gly Gly Ile Ala His     105

TTA TTC AAA CAG AAT AAG GTT GTT CAT GTC AAT GGA TAT AGA AAG ATA ACT GGC AAA AAT     530
Leu Phe Lys Gln Asn Lys Val Val His Val Asn Gly Tyr Arg Lys Ile Thr Gly Lys Asn     125

CAA GTC ACT GCT ACG AAA GCT GAT GGC GGC ACT CAG GTT ATT GAT ACA AAG AAC ATT CTT     590
Gln Val Thr Ala Thr Lys Ala Asp Gly Gly Thr Gln Val Ile Asp Thr Lys Asn Ile Leu     145

ATA GCC ACG GGT TCA GAA GTT ACT CCT TTT CCT GGA ATC ACG ATA GAT GAA GAT ACA ATA     650
Ile Ala Thr Gly Ser Glu Val Thr Pro Phe Pro Gly Ile Thr Ile Asp Glu Asp Thr Ile     165

GTG TCA TCT ACA GGT GCT TTA TCT TTA AAA AAA GTT CCA GAA AAG ATG GTT GTT ATT GGT     710
Val Ser Ser Thr Gly Ala Leu Ser Leu Lys Lys Val Pro Glu Lys Met Val Val Ile Gly     185

GCA GGA GTA ATA GGT GTA GAA TTG GGT TCA GTT TGG CAA AGA CTT GGT GCA GAT GTG ACA     770
Ala Gly Val Ile Gly Val Glu Leu Gly Ser Val Trp Gln Arg Leu Gly Ala Asp Val Thr     205

GCA GTT GAA TTT TTA GGT CAT GTA GGT GGA GTT GGA ATT GAT ATG GAG ATA TCT AAA AAC     830
Ala Val Glu Phe Leu Gly His Val Gly Gly Val Gly Ile Asp Met Glu Ile Ser Lys Asn     225

TTT CAA CGC ATC CTT CAA AAA CAG GGG TTT AAA TTT AAA TTG AAT ACA AAG GTT ACT GGT     890
Phe Gln Arg Ile Leu Gln Lys Gln Gly Phe Lys Phe Lys Leu Asn Thr Lys Val Thr Gly     245

GCT ACC AAG AAG TCA GAT GGA AAA ATT GAT GTT TCT ATT GAA GCT GCT TCT GGT GGT AAA     950
Ala Thr Lys Lys Ser Asp Gly Lys Ile Asp Val Ser Ile Glu Ala Ala Ser Gly Gly Lys     265

GCT GAA GTT ATC ACT TGT GAT GTA CTC TTG GTT TGC ATT GGC CGA CGA CCC TTT ACT AAG    1010
Ala Glu Val Ile Thr Cys Asp Val Leu Leu Val Cys Ile Gly Arg Arg Pro Phe Thr Lys     285

AAT TTG GGA CTA GAA GAG CTG GGA ATT GAA CTA GAT CCC AGA GGT AGA ATT CCA GTC AAT    1070
Asn Leu Gly Leu Glu Glu Leu Gly Ile Glu Leu Asp Pro Arg Gly Arg Ile Pro Val Asn     305

ACC AGA TTT CAA ACT AAA ATT CCA AAT ATC TAT GCC ATT GGT GAT GTA GTT GCT GGT CCA    1130
Thr Arg Phe Gln Thr Lys Ile Pro Asn Ile Tyr Ala Ile Gly Asp Val Val Ala Gly Pro     325

ATG CTG GCT CAC AAA GCA GAG GAT GAA GGC ATT ATC TGT GTT GAA GGA ATG GCT GGT GGT    1190
Met Leu Ala His Lys Ala Glu Asp Glu Gly Ile Ile Cys Val Glu Gly Met Ala Gly Gly     345

GCT GTG CAC ATT GAC TAC AAT TGT GTG CCA TCA GTG ATT TAC ACA CAC CCT GAA GTT GCT    1250
Ala Val His Ile Asp Tyr Asn Cys Val Pro Ser Val Ile Tyr Thr His Pro Glu Val Ala     365

TGG GTT GGC AAA TCA GAA GAG CAG TTG AAA GAA GAG GGT ATT GAG TAC AAA GTT GGG AAA    1310
Trp Val Gly Lys Ser Glu Glu Gln Leu Lys Glu Glu Gly Ile Glu Tyr Lys Val Gly Lys     385

TTC CCA TTT GCT GCT AAC AGC AGA GCT AAG ACA AAT GCT GAC ACA GAT GGC ATG GTG AAG    1370
Phe Pro Phe Ala Ala Asn Ser Arg Ala Lys Thr Asn Ala Asp Thr Asp Gly Met Val Lys     405

ATC CTT GGG CAG AAA TCG ACA GAC AGA GTA CTG GGA GCA CAT ATT CTT GGA CCA GGT GCT    1430
Ile Leu Gly Gln Lys Ser Thr Asp Arg Val Leu Gly Ala His Ile Leu Gly Pro Gly Ala     425

GGA GAA ATG GTA AAT GAA GCT GCT CTT GCT TTG GAA TAT GGA GCA TCC TGT GAA GAT ATA    1490
Gly Glu Met Val Asn Glu Ala Ala Leu Ala Leu Glu Tyr Gly Ala Ser Cys Glu Asp Ile     445

GCT AGA GTC TGT CAT GCA CAT CCG ACC TTA TCA GAA GCT TTT AGA GAA GCA AAT CTT GCT    1550
Ala Arg Val Cys His Ala His Pro Thr Leu Ser Glu Ala Phe Arg Glu Ala Asn Leu Ala     465

GCG TCA TTT GGC AAA TCA ATC AAC TTT TGA ATTAGAAGATTATATATTTTTTTTTCTGAAATTTCCTGG    1619
Ala Ser Phe Gly Lys Ser Ile Asn Phe ---                                             474

GAGCTTTTGTAGAAGTCACATTCCTGAACAGGATATTCTCACAGCTCCAAGAATTTCTAGGACTGAATTATGAAACTTT    1698

TGGAAGGTATTTAATAGGTTTGGACAAAATGGAATACTCTTATATCTATATTTTACATAAATTTAGTATTTTTGTTTCA    1777

GTGCACTAATATGTAAGACAAAAAGCTACTTATTGTAGCATCCTGGAATATCTCCGTCAACTCATATTTTCATGCTGTT    1856

CATGAAAGATTCAATGCCCCTGAATTTAAATAGCTTTTTTCTCTGATACAGAAAAGTTGAATTTTACATGGCTGGAGCT    1935

AGAATTTGATATGTGAACAGTTGTGTTTGAAGCACAGTGATCAAGTTATTTTTAATTTGGTTTTCACATTGGAAACAAG    2014

TCAGTCATTCAGATATGATTCAAATGTCTATAAACCGAACTGATGTAAGTAAAAAAAAAAAAAAAAAA               2082
```

FIG. 2. The nucleotide sequence of λE_3-2 clone and predicted amino acid sequence of human liver E_3. Nucleotides are numbered 5′ to 3′ beginning with the first base sequenced from the clone λE_3-2. Amino acids are numbered with positive numbers beginning with the first residue [alanine (GCA)] that matches the amino terminus of pig heart E_3 (13). Amino acid residues of the leader peptide (underlined) are indicated by negative numbers. The presumed polyadenylylation signal near the 3′ end of the clone is also underlined.

and pig heart E_3. The region corresponding to the active cysteines (amino acids 45–57; Fig. 2) is identical in pig and human E_3. The seven amino acid changes in human E_3 compared to the partial sequence of pig E_3 are: human Asp-86 → Glu (pig), human Lys-285 → Gln (pig), human Asp-297 → Arg (pig), human Tyr-316 → Ala (pig), human Leu-421 → Ile (pig), human Val-429 → Ile (pig), and human Ser-471 → Ala (pig), and in several cases these changes are conservative.

Characterization of Human E_3 mRNA. To examine the abundance and tissue distribution of E_3 mRNA, blot-hybridization analysis was performed on RNA extracted from human liver, kidney, brain, and heart. The results of such analysis using M13 single-stranded DNA probes from

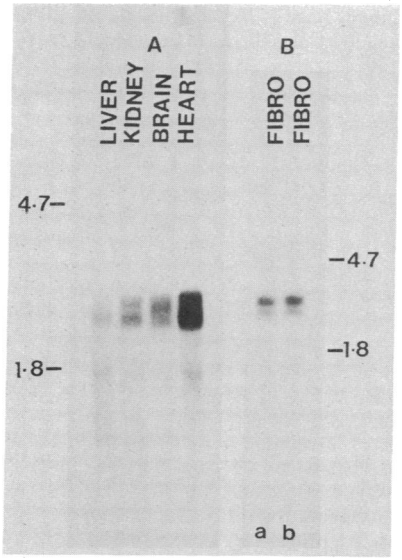

FIG. 3. Blot–hybridization analysis of RNA from human tissues and fibroblasts. Total RNA was extracted from human tissues obtained within 2–4 hr after death and kept frozen. (*A*) Twenty micrograms of total RNA from kidney, brain, and heart and 40 μg from liver were fractionated by electrophoresis on a formaldehyde-agarose gel, transferred to a GeneScreen membrane, and probed with a phage M13 single-stranded DNA probe made from clone λE_3-2. (*B*) Ten micrograms (lane a) and 15 μg (lane b) of poly(A)-containing RNA from human fibroblasts were fractionated by electrophoresis and transferred in the same conditions as described above. The size of the RNA was determined by comparing its electrophoretic mobility with those of ribosomal RNAs.

clone λE_3-2 are shown in Fig. 3*A*. In all human tissues, two major RNA hybridizing bands approximately 2.2 and 2.4 kb in length appeared, but the abundance varied considerably. Heart was the tissue with the highest content of E_3 mRNA. This pattern of E_3 mRNA abundance was constant in several blots using M13 probes either corresponding to the 5′ or the 3′ end of the λE_3-2 clone (data not shown). To rule out a possible artifact caused by the use of total RNA and RNA from tissues obtained at post-mortem, poly(A)-containing RNA from human skin fibroblasts was subjected to RNA blot analysis, which also confirmed the presence of two mRNA species (Fig. 3*B*). To find out whether this RNA pattern of E_3 was present in a different species, we performed a blot-hybridization analysis with poly(A)-containing RNA isolated from rat tissues. The results indicate the presence of a single band of about 2.4 kb in liver, kidney, and brain (Fig. 4).

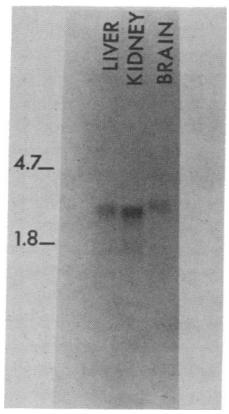

FIG. 4. Blot-hybridization analysis of RNA from rat tissues. Five micrograms of poly(A)-containing RNA was analyzed as described in Fig. 3.

DISCUSSION

The data presented here unequivocally establish the identity of the cDNA clones λE_3-1 and λE_3-2 as corresponding to the mRNA of the human E_3. This conclusion is based on the following observations: (*i*) polypeptides produced by these two λgt11 human liver cDNA clones were recognized by antibodies raised against highly purified E_3, (*ii*) the oligodeoxynucleotide mixture corresponding to an amino acid sequence near the carboxyl terminus of pig heart E_3 hybridized with clone λE_3-2, and (*iii*) the predicted amino acid sequence obtained from the nucleotide sequence of λE_3-2 clone (Fig. 2) is virtually identical with the partial known amino acid sequence of pig heart E_3 (13, 27).

Although there is no apparent homology between the leader sequences of different mitochondrial proteins (28), the analysis of the putative E_3 leader sequence (negatively numbered amino acid residues in Fig. 2) shows a number of features shared by other mitochondrial matrix signal sequences (29, 30). It contains no acidic amino acids and also lacks long stretches of uncharged amino acids. Instead, it has punctuated basic amino acid residues alternating with hydrophobic ones. These structural features, as previously suggested (30, 31), might allow the presequence to fold as an amphiphilic α-helix, independent of the rest of the mature protein, and hence able to interact with the mitochondrial membrane. Secondary structure analysis predictions and hydrophobic calculations based on the first 13 amino acid residues of the E_3 putative leader sequence (data not shown) support the formation of the amphiphilic α-helix structure.

The nucleotide sequence of λE_3-2 clone includes 50 bp of 5′ untranslated region. The sequence in this region contains a stop codon (TGA) in the same reading frame as the E_3 coding region. This is an opal stop codon that cannot be suppressed by the amber suppressor F contained in the host *E. coli* strain Y 1090. It is probable that the *E. coli* translation system reinitiates the translation at the ATG initiation codon of the E_3 coding region rather than producing a β-galactosidase fusion product. The nucleotide sequence upstream of this ATG codon is rich in purine bases and bears some resemblance to the Shine–Dalgarno consensus sequence.

The λE_3-2 cDNA contains 505 bp at the 3′ untranslated region and an unusual polyadenylylation signal (TATAAA) located 16 nucleotides upstream from a short stretch of poly(A) present in the clone. The polyadenylylation signal TATAAA, although described for some mRNAs (32), is not found very often. The difference between the size of the λE_3-2 clone (2.1 kb) and the hybridizing bands in the RNA blots (2.2 and 2.4 kb) suggests that the 3′ end of the λE_3-2 clone represents the 3′ end of the smaller mRNA (2.2 kb) seen on the RNA blots because it is unusual to find a stretch of 18 adenosine residues in the 3′ end preceded by a putative polyadenylylation signal that is not simply a long poly(A) track. The presence of two mRNA species hybridizing with the cDNA clone, detected by blot-hybridization analysis, could be explained by the presence of a second polyadenylylation signal located downstream of the 3′ end of the λE_3-2 clone. Other explanations are possible, and a clear answer will require further investigations.

E_3 belongs to the enzyme family called pyridine nucleotide disulfide oxidoreductases. These enzymes catalyze the electron transfer between pyridine nucleotides and disulfide compounds. Members of this family, in addition to E_3, are glutathione reductase (EC 1.6.4.2), mercury(II) reductase (EC 1.16.1.1), and thioredoxin reductase (EC 1.6.4.5). All of them contain a sequence with two active cysteines involved

in the active center. E_3, glutathione reductase, and mercury(II) reductase show a strong homology in this region, which is not apparent for thioredoxin reductase (13, 27, 33–35). A comparison of the amino acid sequence of human E_3 with the sequence of *E. coli* E_3 shows ≈44% homology and another ≈15% of the sequence shows conservative amino acid changes (data not shown). This strong homology is extended throughout the polypeptide, although the middle portion of the sequence seems to show the weakest homology. This portion of E_3 could be involved in the interaction with the other component(s) of the α-ketoacid dehydrogenase complexes that are different in structure and symmetry between *E. coli* and mammals. Glutathione reductase, in accordance with previous comparisons (26), also shows a significant homology with *E. coli* E_3 (28% homology) (35), pig heart E_3 (partial) (40% homology) (27), and human liver E_3 (33% homology) (present study). It is interesting to note that most of the homologies between glutathione reductase and both *E. coli* and human E_3 are coincident with the ones between both E_3 sequences. These homologies offer strong evidence for an evolutionary relationship between E_3 and glutathione reductase.

We thank Dr. Lester J. Reed (University of Texas, Austin) for the E_3 component of bovine kidney pyruvate dehydrogenase complex; Drs. T. Chandra and S. L. C. Woo (Baylor College of Medicine, Houston) for the λgt11 human liver cDNA library; Dr. Douglas Kerr for providing RNA from human tissues; and Dr. Richard W. Hanson for fruitful discussions and for critical reading of this manuscript. This work was supported by Public Health Service Grant AM 20478, Metabolism Training Grant AM 07319 (to B.W.J.), National Research Award NS 07376 (to M.K.G.), and a Fulbright–Ministerio de Educacion y Ciencia Fellowship Award (to G.P.).

1. Pettit, F. H. & Reed, L. J. (1982) *Methods Enzymol.* **89,** 376–386.
2. Koike, M. & Koike, K. (1976) *Adv. Biophys.* **9,** 187–227.
3. Pettit, F. H., Yeaman, S. J. & Reed, L. J. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4881–4885.
4. Cook, K. G., Bradford, A. P. & Yeaman, S. J. (1985) *Biochem. J.* **225,** 731–735.
5. Sakurai, Y., Fukuyoshi, Y., Hamada, M., Hayakawa, T. & Koike, M. (1970) *J. Biol. Chem.* **245,** 4453–4462.
6. Matuda, S. & Aaheki, T. (1982) *J. Biochem. (Tokyo)* **91,** 553–561.
7. Matuda, S., Shirahama, T., Saheki, T., Miura, S. & Mori, M. (1983) *Biochim. Biophys. Acta* **741,** 86–93.
8. Heffelfinger, S. C., Sewell, E. T. & Danner, D. J. (1983) *Biochem. J.* **213,** 339–344.
9. DeMarcucci, O. L., Hunter, A. & Lindsay, J. G. (1985) *Biochem. J.* **226,** 509–517.
10. Blass, J. P. (1983) in *The Metabolic Basis of Inherited Disease*, eds. Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L. & Brown, M. S. (McGraw-Hill, New York), pp. 193–203.
11. Kikuchi, G., Hiraga, K. & Yoshida, T. (1980) *Biochem. Soc. Trans.* **8,** 504–506.
12. Carothers, D. J., Raefsky-Estrin, C., Pons, G. & Patel, M. S. (1987) *Arch. Biochem. Biophys.* **256,** 597–605.
13. Arscott, L. D., Berman, M. & Williams, Ch. H., Jr. (1984) in *Proceedings of the Eighth International Symposium on Flavins and Flavoproteins*, eds. Bray, R. C., Engel, P. C. & Mayhew, S. G. (de Gruyter, Berlin), pp. 153–156.
14. Pons, G., Raefsky, C., Carothers, D. J., Pepin, R. A., Javed, A. A., Jesse, B. W., Ganapathi, M. K., Samols, D. & Patel, M. S. (1987) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **46,** 2214 (abstr.).
15. Vaitukaits, J. L. (1981) *Methods Enzymol.* **73,** 46–52.
16. Ho, L., Hu, C. W. C., Packman, S. & Patel, M. S. (1986) *J. Clin. Invest.* **78,** 844–847.
17. Kemp, D. J. & Cowman, A. F. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 4520–4525.
18. Kwok, S. C. M., Ledley, F. D., Dilella, A. G., Robson, K. J. H. & Woo, S. L. C. (1985) *Biochemistry* **24,** 556–561.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
21. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
22. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
23. Hod, Y., Morris, S. M. & Hanson, R. W. (1984) *J. Biol. Chem.* **259,** 15603–15608.
24. Davis, L. G., Dibner, M. D. & Battey, J. F. (1986) in *Basic Methods in Molecular Biology* (Elsevier, New York), pp. 276–284.
25. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
26. DeMarcucci, O. G. L., Hunter, A. & Lindsay, J. G. (1986) *Biochem. Soc. Trans.* **14,** 870.
27. Williams, Ch. H., Jr., Arscott, D. L. & Schulz, G. E. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2199–2201.
28. Colman, A. & Robinson, C. (1986) *Cell* **46,** 321–322.
29. Hurt, E. C. & Van Loon, A. P. G. M. (1986) *Trends Biochem. Sci.* **11,** 204–207.
30. Heijne, G. (1986) *EMBO J.* **5,** 1335–1342.
31. Hurt, E. C. & Schatz, G. (1987) *Nature (London)* **325,** 499–503.
32. Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41,** 349–359.
33. Packman, L. C. & Perham, R. N. (1982) *FEBS Lett.* **139,** 155–158.
34. Stephens, P. E., Lewis, H. M., Darlison, M. G. & Guest, J. R. (1983) *Eur. J. Biochem.* **135,** 519–527.
35. Rice, D. W., Schulz, G. E. & Guest, J. R. (1984) *J. Mol. Biol.* **174,** 483–496.