Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7546–7550, October 1988
Biochemistry

Cloning and nucleotide sequence of the gene for dihydrolipoamide acetyltransferase from *Saccharomyces cerevisiae*

(multidomain structure/sequence homology/pyruvate dehydrogenase complex)

XIAO-DA NIU*†, KAREN S. BROWNING*, ROBERT H. BEHAL*†, AND LESTER J. REED*†

*Clayton Foundation Biochemical Institute and †Department of Chemistry, The University of Texas at Austin, Austin, TX 78712

Contributed by Lester J. Reed, July 29, 1988

ABSTRACT **A 537-base cDNA encoding a portion of *Saccharomyces cerevisiae* dihydrolipoamide acetyltransferase (acetyl-CoA:dihydrolipoamide *S*-acetyltransferase, EC 2.3.1.12) was isolated from a λgt11 yeast cDNA library by immunoscreening. This cDNA was subcloned and used as a probe to screen a λgt11 yeast genomic DNA library. Two overlapping clones were used to determine the complete sequence of the acetyltransferase gene. The composite sequence has an open reading frame of 1446 nucleotides encoding a presequence of 28 amino acids and a mature protein of 454 amino acids (M_r = 48,546). The deduced amino acid sequence contains the experimentally determined amino acid sequences of the amino terminus and two internal peptide fragments of the acetyltransferase. Hybridization analysis of yeast genomic DNA showed that the gene has a single copy. A 915-base segment of the acetyltransferase gene hybridized to a yeast mRNA of ≈1.6 kilobases. Analysis of the deduced amino acid sequence of the dihydrolipoamide acetyltransferase revealed a multidomain structure similar to those reported for the corresponding acetyltransferases from *Escherichia coli* and rat liver, and extensive sequence similarity among the three enzymes. However, the yeast enzyme contains only one lipoyl domain, in contrast to three lipoyl domains reported for the *E. coli* enzyme and apparently two for the rat liver enzyme.**

Pyruvate dehydrogenase complexes from prokaryotic and eukaryotic sources are composed of multiple copies of three enzymes: pyruvate dehydrogenase (E_1), dihydrolipoamide acetyltransferase (E_2; acetyl-CoA dihydrolipoamide *S*-acetyltransferase, EC 2.3.1.12), and dihydrolipoamide dehydrogenase (E_3). The oligomeric E_2 component comprises a symmetrical core around which are arranged multiple copies of E_1 and E_3 (1). Analyses of the E_2 components of the *Escherichia coli* pyruvate and α-ketoglutarate dehydrogenase complexes by controlled proteolysis (2–4), molecular genetics (5–7), and ^{1}H-NMR spectroscopy (8–11) revealed that these enzymes have a multidomain structure. The amino-terminal segment of the polypeptide chain contains the lipoyl domain(s), followed by an E_3-binding domain, and then the catalytic inner-core domain. The domains are linked to each other by segments that are rich in proline and alanine residues. These interdomain linker segments are thought to provide flexibility to the lipoyl domains, facilitating active-site coupling within these multienzyme complexes (2, 9, 10).

In eukaryotic cells, the pyruvate dehydrogenase complex is located in mitochondria, within the inner membrane/matrix compartment. To gain further understanding of the structure, function, and regulation of eukaryotic pyruvate dehydrogenase complexes, we have undertaken isolation and expression of the genes encoding E_1, E_2, and E_3 from *Saccharomyces cerevisiae*. In this report, we present the cloning and sequencing of the gene for yeast E_2 and make comparisons with the corresponding enzymes from rat liver and *E. coli*.‡

MATERIALS AND METHODS

Materials. Restriction endonucleases and DNA-modifying enzymes were purchased from Bethesda Research Laboratories, New England Biolabs, and Promega Biotec (Madison, WI). The DNA sequencing reagents were purchased from United States Biochemical. Horseradish peroxidase-labeled goat anti-rabbit IgG was obtained from Kirkegaard and Perry Laboratories (Gaithersburg, MD). [γ-^{32}P]ATP, [α-^{32}P]dCTP, and [α-[^{35}S]thio]dATP were purchased from New England Nuclear. The λgt11 yeast cDNA library (12) was obtained from Hans Trachsel and Michael Altmann (Universität Bern, Bern, Switzerland), and the λgt11 yeast genomic DNA library was provided by Richard A. Young and Ronald W. Davis (Stanford University, Palo Alto, CA). Immobilon polyvinylidene difluoride membrane was obtained from Millipore. Specialty agaroses were obtained from FMC (Rockland, ME). The plasmid vector Bluescript was obtained from Stratagene (San Diego, CA). *E. coli* strains Y1090 and TB-1 were used to grow the λ phage library and plasmids, respectively.

Preparation of Antiserum. Highly purified pyruvate dehydrogenase complex from bakers' yeast was subjected to $NaDodSO_4$/polyacrylamide gel electrophoresis (13). The gel was stained with Coomassie brilliant blue and then destained, and the E_2 band was excised with a razor blade and soaked in deionized water. The gel pieces were frozen in liquid nitrogen, crushed in a mortar and pestle, and stored at −20°C (14). Crushed gel containing about 200 μg of E_2 was suspended in 1.5 ml of sterile 0.14 M NaCl. The suspension was then mixed with 2 ml of Freund's complete adjuvant, and the emulsion was injected subcutaneously into a female New Zealand White rabbit. Booster injections were given at 20-day intervals with Freund's incomplete adjuvant. Blood was collected 10 days after the fourth booster injection, and the serum was separated and stored at −70°C. Immunoblot analysis showed that the antiserum was specific for E_2 (15). Antibodies reacting with *E. coli* and phage proteins were removed from the serum by incubation with *E. coli* and phage lysates immobilized on nitrocellulose (16).

Amino-Terminal and Peptide Sequence Analysis. Highly purified pyruvate dehydrogenase complex was subjected to $NaDodSO_4$/polyacrylamide gel electrophoresis in a slab gel (12.5% acrylamide) as described (17). The gel was soaked in 2 M sodium acetate, and the protein bands were visualized with side illumination. The E_2 band was excised with a razor

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.

‡The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J04096).

blade, crushed, and extracted overnight with 100 mM NH_4HCO_3. The extract was dialyzed against 100 mM NH_4HCO_3/20% methanol and then against 50 mM NH_4HCO_3. The solvent was removed on a vacuum centrifuge, and the residue was subjected to automated sequence analysis with an Applied Biosystems model 470A gas-phase sequenator equipped with a model 120A on-line phenylthiohydantoin amino acid analyzer.

To obtain internal peptides of E_2 for sequence analysis, a gel slice containing about 40 μg of E_2 was applied to a second $NaDodSO_4$/polyacrylamide slab gel in the presence of *Staphylococcus aureus* V8 protease (0.2 μg) for 30 min as described by Cleveland *et al.* (18). After electrophoresis, the peptides were electroblotted onto a polyvinylidene difluoride membrane (19). The membrane was stained with Coomassie brilliant blue, destained, rinsed in deionized water, and air-dried. The peptide bands were cut out with a razor blade and subjected to automated sequence analysis.

Preparation of Oligonucleotides. Oligonucleotide primers for sequencing were synthesized on an Applied Biosystems model 381A DNA synthesizer and were used without further purification.

Screening of cDNA and Genomic DNA Libraries. The λgt11 yeast cDNA library was screened with rabbit antiserum against yeast E_2 essentially as described by Huynh *et al.* (16). The λgt11 yeast cDNA and genomic DNA libraries were also screened by plaque hybridization using a DNA probe made from a positive clone identified by screening with antibody against E_2 (20). Positive clones were plaque-purified and phage DNA was prepared from plate lysates (20). In brief, the top agarose was removed from three plates (150 × 15 mm) with confluent lysis and soaked in 20 ml of SM (50 mM Tris·HCl, pH 7.5/100 mM NaCl/10 mM $MgSO_4$/0.01% gelatin) overnight with shaking at 4°C. The debris was removed by centrifugation; and the supernatant fluid was mixed with 10 ml of a slurry of DEAE-cellulose in SM on a rotatory shaker for 30 min. The DEAE-cellulose was removed by centrifugation, and 0.2 volume of 3 M NaCl and 0.7 volume of 2-propanol were added to the supernatant fluid. The precipitate was collected by centrifugation and resuspended in 2 ml of TE (10 mM Tris·HCl, pH 7.6/1 mM EDTA). A stock solution was added to the suspension to give 0.5% $NaDodSO_4$, 10 mM NaCl, 2 mM EDTA, and 0.05 mg of proteinase K per ml (final concentrations), and the mixture was incubated at 37°C for 1 hr. The mixture was extracted twice with phenol/chloroform (1:1), and the phage DNA was precipitated with ethanol and dissolved in 20 μl of TE.

Subcloning of DNA for Sequencing. The λ phage DNA containing the insert was digested with *Eco*RI or when the insert contained internal *Eco*RI sites, with a combination of *Kpn* I and *Sst* I. The fragments were resolved in a 0.8% low-melting agarose gel. The bands containing the inserts were excised, melted at 65°C, and extracted with phenol/chloroform, and the DNA was precipitated with ethanol. The fragments were ligated into the plasmid vector Bluescript, which had been previously digested with *Eco*RI or with *Kpn* I and *Sst* I and treated with bacterial alkaline phosphatase. Plasmids containing inserts were prepared as described (21).

DNA Sequencing. The double-stranded plasmid DNA was sequenced with Sequenase (United States Biochemical) according to the supplier's instructions except that the plasmid was first denatured in 0.2 M NaOH, precipitated with ethanol, and hybridized to an oligonucleotide primer for 15 min at 37°C.

RNA Blot Analysis. A 10-μg sample of poly(A)$^+$ RNA from *S. cerevisiae* strain 20B-12 was denatured with glyoxal, fractionated in a 0.8% agarose gel, and blotted onto nitrocellulose as described (22). The probe, a 915-base DNA fragment designated E2-G2, was labeled with [α-^{32}P]dCTP by random priming with a mixture of oligonucleotides. Hybridization conditions were as described (23).

Southern Blot Analysis of Total Yeast Genomic DNA. Yeast genomic DNA was prepared as described (24). Ten-microgram DNA samples were digested with various restriction endonucleases overnight at 37°C. The digests were resolved in a 0.8% agarose gel and blotted onto nitrocellulose by capillary transfer in 10× SSC (1× SSC is 0.15 M NaCl/15 mM sodium citrate, pH 7.0). Blots were prehybridized for 10 hr at 42°C in 6× SSC containing 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin, 0.1% $NaDodSO_4$, and 0.1 mg of sonicated and denatured salmon testes DNA per ml. The blots were hybridized for 12 hr at 65°C in the same solution with the 915-base DNA probe E2-G2 radiolabeled as described above. The blots were washed twice in 2× SSC/0.1% $NaDodSO_4$ for a total of 20 min and once in 0.1× SSC/0.1% $NaDodSO_4$ for 2 hr at 68°C. Autoradiography was performed at −70°C with Kodak XAR-5 film and an intensifying screen.

Chromosome Analysis. A dried gel containing the yeast chromosomes resolved by pulsed-field gel electrophoresis was purchased from Clontech Laboratories (Palo Alto, CA). The gel was hybridized with the 915-base DNA probe E2-G2 according to the manufacturer's instructions.

Computer Analysis. Computer analysis was performed with the Beckman Microgenie program, version 4.

RESULTS

Amino Acid Sequence Analysis. The amino-terminal sequence of E_2 isolated from the yeast pyruvate dehydrogenase complex was determined to be (Ala?)-Ser-Tyr-Pro-Glu-His-Xaa-Ile-Ile-Gly-Met-Xaa-Ala-Leu-Ser-, where Xaa indicates residues that could not be identified. A sample of E_2 was digested with *S. aureus* V8 protease, and the resulting peptides were separated by $NaDodSO_4$/polyacrylamide gel electrophoresis and then electroblotted onto Immobilon polyvinylidene difluoride membrane. The amino-terminal sequences of two fragments were determined by automated Edman degradation to be Thr-Lys-Thr-Ser-Ala-Xaa-Glu-Ala-Lys-Lys-Ser-Asp-Val- and Xaa-Xaa-Gln-Glu-Ala-(Pro?)-Ala-Glu-Xaa-Thr-(Lys or Gly).

Isolation of the Gene for E_2. Two positive clones were initially identified by screening a yeast cDNA λgt11 library with antibody against yeast E_2. Both inserts, designated E21-1 and E21-2, were subcloned and their nucleotide sequences were determined. E21-1 contained an insert of 537

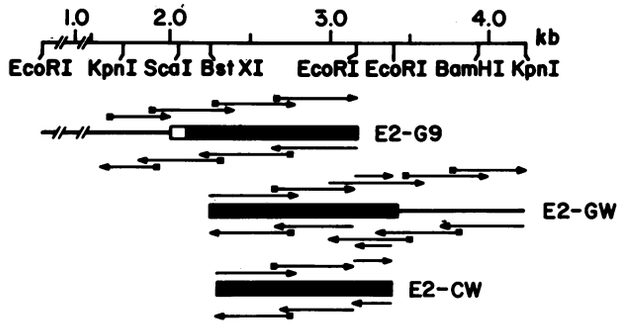

FIG. 1. Restriction endonuclease map and sequencing strategy for yeast E_2 gene. Two genomic DNA inserts (E2-G9 and E2-GW) and one cDNA insert (E2-CW) were sequenced. Solid bars represent the coding regions for the mature enzyme. Open bar represents the coding region for the putative presequence. Arrows show the direction and extent of the nucleotide sequencing. Arrows beginning with solid boxes represent sequences determined with synthetic primers based on previously obtained sequences. Other arrows represent sequences determined with universal or reverse sequencing primers. kb, Kilobases.

```
AGGAAGAGGCTGAGAACCCTAGATAATTGGTACCGGTGATCACCTCCAGTTTGCACATAC 1740
TATATATCTTCATGTAATACTACATATGTCATAACTTCAATAACTCATACAAGTCATCTG 1800
CTACTATGCTTTGCCGTTTCTACTATCCGTGAAGCGTGACCCGCCCTTCTCACCCCCCGT 1860
GTTACGCAAAGAACAACAAGAGGTCGAGACAACCAACCGGAACGACTAGAAGAGAGGCTG 1920
TTACGTTTCCTGTTTGTTCACGCTCAACATCTTCCATTGCTGTAAACTGCATCCAGTAAA 1980
ATTAATAGTTATTAGCGTACTATGTCTGCCTTTGTCAGGGTGGTTCCAAGAATATCCAGA 2040
                    M  S  A  F  V  R  V  V  P  R  I  S  R   -16
               -28
AGTTCAGTACTCACCAGATCATTGACACTGCAATTGACATGCTACGCATCGTACCCAGAG 2100
 S  S  V  L  T  R  S  L  R  L  Q  L  R  C  Y  A  S  Y  P  E     5
                                               1
CACACCATTATTGGTATGCCGGCACTGTCTCCTACGATGACGCAAGGTAATCTTGCTGCT 2160
 H  T  I  I  G  M  P  A  L  S  P  T  M  T  Q  G  N  L  A  A    25

TGGACTAAGAAGGAAGGTGACCAATTGTCTCCCGGTGAAGTTATTGCCGAAATAGAAACA 2220
 W  T  K  K  E  G  D  Q  L  S  P  G  E  V  I  A  E  I  E  T    45

GACAAGGCTCAAATGGACTTTGAGTTCCAAGAAGATGGTTACTTAGCCAAGATTCTAGTT 2280
 D  K  A  Q  M  D  F  E  F  Q  E  D  G  Y  L  A  K  I  L  V    65

CCTGAAGGTACAAAGGACATTCCTGTCAACAAGCCTATTGCCGTCTATGTGGAGGACAAA 2340
 P  E  G  T  K  D  I  P  V  N  K  P  I  A  V  Y  V  E  D  K    85

GCTGATGTGCCAGCTTTTAAGGACTTTAAGCTGGAGGATTCAGGTTCTGATTCAAAGACC 2400
 A  D  V  P  A  F  K  D  F  K  L  E  D  S  G  S  D  S  K  T   105

AGTACGAAGGCTCAGCCTGCCGAACCACAGGCAGAAAAGAAACAAGAAGCGCCAGCTGAA 2460
 S  T  K  A  Q  P  A  E  P  Q  A  E  K  K  Q  E  A  P  A  E   125

GAGACCAAGACTTCTGCACCTGAAGCTAAGAAATCTGACGTTGCTGCTCCTCAAGGTAGG 2520
 E  T  K  T  S  A  P  E  A  K  K  S  D  V  A  A  P  Q  G  R   145

ATTTTTGCCTCTCCACTTGCCAAGACTATCGCCTTGGAAAAGGGTATTTCTTTGAAGGAT 2580
 I  F  A  S  P  L  A  K  T  I  A  L  E  K  G  I  S  L  K  D   165

GTTCACGGCACTGGACCCCGCCGTAGAATTACCAAGGCTGACATTGAGTCATATCTAGAA 2640
 V  H  G  T  G  P  R  G  R  I  T  K  A  D  I  E  S  Y  L  E   185

AAGTCGTCTAAGCAGTCTTCTCAAACCAGTGGTGCTGCCGCCGCCACTCCTGCCGCCGCT 2700
 K  S  S  K  Q  S  S  Q  T  S  G  A  A  A  A  T  P  A  A  A   205

ACCTCAAGCACTACTGCTGGCTCTGCTCCATCGCCTTCTTCTACAGCATCATATGAGGAT 2760
 T  S  S  T  T  A  G  S  A  P  S  P  S  S  T  A  S  Y  E  D   225

GTTCCAATTTCAACCATGAGAAGCATCATTGCGAGAACGTTTATTGCAATCTACTCAAGGC 2820
 V  P  I  S  T  M  R  S  I  I  G  E  R  L  L  Q  S  T  Q  G   245

ATTCCATCATACATCGTTTCCTCCAAGATATCCATCTCCAAACTTTTGAAATTGAGACAG 2880
 I  P  S  Y  I  V  S  S  K  I  S  I  S  K  L  L  K  L  R  Q   265

TCCTTGAACGCTACAGCAAACGACAAGTACAAACTGTCCATTAATGACCTATTAGTAAAA 2940
 S  L  N  A  T  A  N  D  K  Y  K  L  S  I  N  D  L  L  V  K   285

GCCATCACTGTTGCGGCTAAGAGGGTGCCAGATGCCAATGCCTACTGGTTACCTAATGAG 3000
 A  I  T  V  A  A  K  R  V  P  D  A  N  A  Y  W  L  P  N  E   305

AACGTTATCCGTAAATTCAAGAATGTCGATGTCTCAGTCGCTGTTGCCACACCAACAGGA 3060
 N  V  I  R  K  F  K  N  V  D  V  S  V  A  V  A  T  P  T  G   325

TTATTGACACCAATTGTCAAGAATTGTGAGGCCAAGGGCTTGTCGCAAATCTCTAACGAA 3120
 L  L  T  P  I  V  K  N  C  E  A  K  G  L  S  Q  I  S  N  E   345

ATCAAGGAACTAGTCAAGCGTGCCAGAATAAACAAATTGGCACCAGAGGAATTCCAAGGT 3180
 I  K  E  L  V  K  R  A  R  I  N  K  L  A  P  E  E  F  Q  G   365

GGGACCATTTGCATATCCAATATGGGCATGAATAATGCTGTTAACATGTTTACTTCGATT 3240
 G  T  I  C  I  S  N  M  G  M  N  N  A  V  N  M  F  T  S  I   385

ATCAACCCACCACAGTCTACAATCTTGGCCATCGCTACTGTTGAAAGGGTCGCTGTGGAA 3300
 I  N  P  P  Q  S  T  I  L  A  I  A  T  V  E  R  V  A  V  E   405

GACGCCGCTGCTGAGAACGGATTCTCCTTTGATAACCAGGTTACCATAACAGGGACCTTT 3360
 D  A  A  A  E  N  G  F  S  F  D  N  Q  V  T  I  T  G  T  F   425

GATCATAGAACCATTGATGGCGCCAAAGGTGCAGAATTCATGAAGGAATTGAAAACTGTT 3420
 D  H  R  T  I  D  G  A  K  G  A  E  F  M  K  E  L  K  T  V   445

ATTGAAAATCCTTTGGAAATGCTATTGTGAGGTTAGAATGAAAATAAAATTCGCCAGTAA 3480
 I  E  N  P  L  E  M  L  L

ATGCGTATCTTGGAGAAAATATCAATTTGTTATTTTGTATATATAAAAAATTAAAAAAAA 3540
AAAAAAAAAAATACTCAAATACTCAAATACGTATACATTATACTGAAATACTTCGCTTGT 3600
GTTCGAATCTAAATCGTCTCTCTCTTCTGTCTCAAAATGATATAAATAAATCACGTCAAA 3660
CGGTGCGTCAACATCTTTGTTTTTTCACCTCATGTGGAATTGGTTGAAATTAACAAAGGA 3720
GTGGGATATAGGATATAAGGCATAAAGTTTAGGATACTACTTTCGTATATTATTGTTCGG 3780
TGAAAGAAACTAGTTCCCTCTTATCTCTCAATATTTGCCAAAATTAGCTTTTAACAAATA 3840
AACCATGAGTTTCCTACCATCTTTTATCTTAAGCCATGAATCTAAAGAACGTATTTCCAA 3900
AATTTTAACTTTGACTCATAATGTAGCACATTATGGCTGGATCCCATTTGTTTTGTATTT 3960
GGGCTGGGCACACACTTCTAATAGACCAAACTTTTTGAACTTACTGTCTCCGTTACCAAG 4020
TGTTTAAAAACATCAGTTCGATACAACTTGGGTGAGAGAGGAAGCCCATATTTCCAATTT 4080
TGAATTACTTGACCGTTACTGCAAATAATAAAAATTGACATGTACATAAAGTATATATAT 4140
AATAAAATATAAAATTCATTTTGCATATTGTTATTCTTTCACGCTTTGTTTTGATATGGT 4200
GAAAATTCGAAGTCTCTAGAATGTCCTTTAGGTACC
```

FIG. 2. Nucleotide sequence of the yeast E_2 gene and the deduced amino acid sequence (one-letter amino acid symbols). The experimentally determined amino acid sequences are underlined. A putative polyadenylylation signal is overlined.

bases and had an open reading frame that contained the experimentally determined amino acid sequences of two internal peptides of E_2. E21-1 was used as a probe to identify two additional clones, E2-GW and E2-CW, containing inserts of 1985 and 1065 bases, respectively. The latter insert was isolated from the same cDNA library and the former from a yeast genomic DNA library. Both inserts contained internal *Eco*RI sites, requiring that *Kpn* I and *Sst* I be used for subcloning. Sequencing results showed that E2-GW contains the entire nucleotide sequence of E21-1 and brought the open reading frame to a stop codon. To identify the missing portion of the 5′ end of the open reading frame, $\approx 2 \times 10^6$ recombinants from the yeast genomic DNA library were screened with a 915-base DNA fragment from E2-GW, designated E2-G2 (Fig. 1, nucleotides 2256–3171). By sizing of the DNA fragments after digestion with *Eco*RI, one of the resulting positive clones was identified as having the 5′ upstream sequence of the open reading frame. This insert, designated E2-G9, was subcloned and sequenced.

Sequence Analysis of cDNA and Genomic DNA Inserts. The restriction map and sequencing strategy for the cDNA and genomic DNA inserts are shown in Fig. 1. Two genomic DNA inserts, E2-G9 and E2-GW, with a 915-base overlap, contained the nucleotide sequence encoding the precursor of E_2 as well as 5′ and 3′ flanking regions. The nucleotide sequence and the deduced amino acid sequence of the protein are shown in Fig. 2. The genomic DNA sequenced comprised ≈4.2 kb. However, only the sequence of the E_2 gene and its flanking regions are presented in Fig. 2. The remainder of the

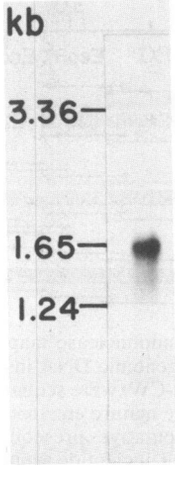

FIG. 3. RNA blot analysis of yeast poly(A)$^+$ RNA with E2-G2. The positions of satellite tobacco necrosis virus RNA (1.24 kb) and yeast 18S (1.65 kb) and 25S (3.36 kb) rRNA are indicated.

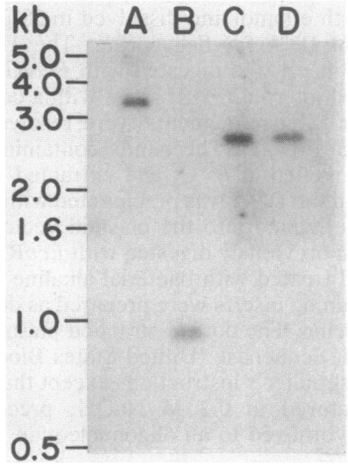

FIG. 4. Southern blot analysis of yeast genomic DNA. DNA samples (10 μg) were digested with restriction enzymes, blotted, and hybridized with E2-G2. Lane A, *Eco*RI; lane B, *Eco*RI and *Bst*XI; lane C, *Kpn* I; lane D, *Kpn* I and *Pst* I. Molecular size markers are as indicated.

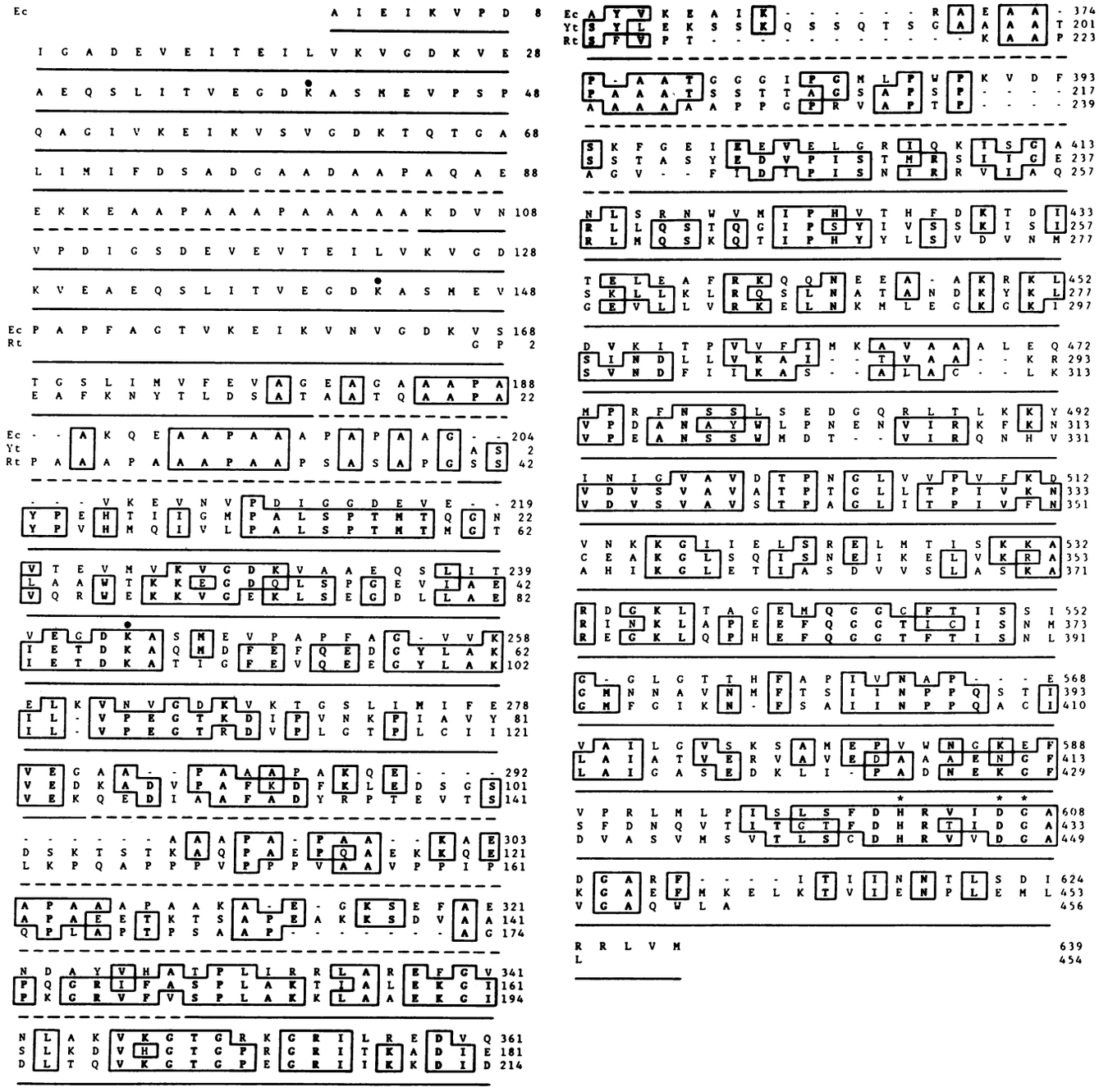

FIG. 5. Comparison of E_2 amino acid sequences. The deduced sequences of the *E. coli* (Ec), yeast (Yt), and putative rat liver (Rt) enzymes are aligned for maximal homology. Identical residues are boxed. The lipoyl domains (amino-terminal), E_3-binding domains, and catalytic inner-core domains (carboxyl-terminal) are represented by solid lines, and the putative interdomain linker segments are represented by dashed lines. The lipoyllysine residues in the lipoyl domains are indicated by solid circles, and residues in the putative active site are denoted by stars.

sequence contained one complete open reading frame and a portion of another. These sequences will be reported elsewhere. There is a 1-base mismatch at position 3279, where a thymine in the genomic insert is replaced with a cytosine in the cDNA insert. This mismatch is in the wobble position of the codon.

The open reading frame of 1446 nucleotides encodes a putative presequence of 28 amino acid residues and a mature protein of 454 amino acid residues, which has a calculated M_r = 48,546. The presequence is rich in basic and hydroxylated amino acid residues and is devoid of acidic amino acid residues. This feature, similar to the putative presequence of yeast E_3 (25), is characteristic of presequences of mitochondrial matrix/inner membrane proteins (26). The deduced amino acid sequence of the mature protein contains the experimentally determined sequences of the amino terminus and two internal peptides of the purified E_2 (underlined).

RNA Blot Analysis. RNA blot analysis of yeast poly(A)$^+$ RNA shows that the size of the mRNA is ≈1.6 kb (Fig. 3).

Southern Blot Analysis. To identify genomic DNA fragments carrying the gene of E_2, yeast DNA was digested with various restriction endonucleases and the resulting fragments were hybridized to the ^{32}P-labeled DNA probe E2-G2 (Fig. 4). A single band was hybridized in each digestion. The sizes of the fragments generated by *Bst*XI/*Eco*RI digestion (900 bases) and by *Kpn* I digestion (2500 bases) are in good

agreement with the restriction map deduced from the nucleotide sequence of the gene for E_2. These data indicate that there is a single copy of the gene for E_2. Additional analysis indicated that the E_2 gene is located on chromosome XIV (data not shown).

Comparison of E_2 Amino Acid Sequences. The deduced amino acid sequences of the E_2 enzymes from *E. coli*, yeast, and rat liver are compared in Fig. 5. The amino acid sequence of the putative rat liver acetyltransferase was deduced from the nucleotide sequence of a cDNA encoding the 70-kDa mitochondrial autoantigen of primary biliary cirrhosis (27–29). The sequence of the rat liver enzyme is incomplete (28, 29), apparently lacking sequences at the amino and carboxyl termini. The sequence comparisons of the acetyltransferases show that yeast E_2 has a multidomain structure similar to those identified previously for the E_2 components of the *E. coli* pyruvate and α-ketoglutarate dehydrogenase complexes. The yeast and *E. coli* acetyltransferases exhibit 31% sequence identity for 463 residues, after the introduction of gaps to optimize the alignment. The corresponding value for the yeast and rat liver acetyltransferases is 45% for 440 residues.

DISCUSSION

Amino acid sequence comparisons of the E_2 proteins from yeast, *E. coli*, and rat liver (Fig. 5) clearly demonstrate that the yeast enzyme has a highly segmented structure similar to that reported for the *E. coli* acetyltransferase and succinyltransferase (7, 11). The lipoyl domain, comprising the amino-terminal region, is followed by the E_3-binding domain and the catalytic inner-core domain. These putative domains are connected by interdomain linker segments. Sequence identity of 31% was found between the yeast and *E. coli* enzymes, and 45% between the yeast and rat liver enzymes. Conservation of sequence is particularly high around the lipoyllysine residues of the lipoyl domains, in the E_3-binding domains, and in the carboxyl-terminal segment of the catalytic inner-core domains. The putative catalytic site is thought to include the highly conserved His-Xaa-Xaa-Xaa-Asp-Gly segment near the carboxyl terminus (30).

In contrast to the *E. coli* acetyltransferase, which contains three lipoyl domains (5), and the putative rat liver acetyltransferase, which apparently has two lipoyl domains (29), the yeast acetyltransferase has only one lipoyl domain. In this respect, the yeast acetyltransferase resembles the *E. coli* dihydrolipoamide succinyltransferase (6, 31) and the dihydrolipoamide acetyltransferase component of the branched-chain α-keto acid dehydrogenase complex (32, 33).

Although the length and composition of the interdomain linker segment vary among the three acetyltransferases, these segments are similar in that they are rich in the conservatively substituted residues alanine, proline, serine, and threonine and in charged amino acid residues. The linker segments of the *E. coli* and rat liver acetyltransferases have a higher content of alanine and proline residues than the yeast enzyme, whereas the linker segments of the yeast enzyme have a higher content of threonine and serine residues. The interdomain linker segments are thought to be responsible, at least in part, for the relatively slow migration of E_2 in $NaDodSO_4$/polyacrylamide gel electrophoresis, resulting in abnormally high apparent molecular weight values (2, 34). Thus, the apparent molecular weight of the yeast acetyltransferase estimated by $NaDodSO_4$/polyacrylamide gel electrophoresis is ≈58,000 (35), whereas the molecular weight calculated from the deduced amino acid sequence is 48,456.

The availability of the structural gene for E_2 should facilitate investigation into mitochondrial import, self-assembly, and structure–function relationships of this oligomeric enzyme.

We thank Dr. David J. Uhlinger for the rabbit antibody to yeast E_2 and Drs. Robert W. Gracy and K. Ümit Yüksel (North Texas State University/Texas College of Osteopathic Medicine, Fort Worth) and Dr. Alexander Kurosky (University of Texas Medical Branch, Galveston) for amino acid sequence analyses. We thank Drs. Russell F. Doolittle and Dean Appling for critically reading the manuscript. This work was supported in part by Grant GM06590 from the U.S. Public Health Service.

1. Reed, L. J. (1974) *Acc. Chem. Res.* **7**, 40–46.
2. Bleile, D. M., Munk, P., Oliver, R. M. & Reed, L. J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4385–4389.
3. Packman, L. C. & Perham, R. M. (1986) *FEBS Lett.* **206**, 193–198.
4. Packman, L. C. & Perham, R. N. (1987) *Biochem. J.* **242**, 531–538.
5. Stephens, P. E., Darlison, M. G., Lewis, H. M. & Guest, J. R. (1983) *Eur. J. Biochem.* **133**, 481–489.
6. Spencer, M. E., Darlison, M. G., Stephens, P. E., Duckenfield, I. K. & Guest, J. R. (1984) *Eur. J. Biochem.* **141**, 361–374.
7. Miles, J. S. & Guest, J. R. (1988) *Biochem. Soc. Symp.* **54**, 45–65.
8. Perham, R. N., Duckworth, H. W. & Roberts, G. C. K. (1981) *Nature (London)* **292**, 474–477.
9. Perham, R. N. & Roberts, G. C. K. (1981) *Biochem. J.* **199**, 733–740.
10. Radford, S. E., Laue, E. D., Perham, R. N., Miles, J. S. & Guest, J. R. (1987) *Biochem. J.* **247**, 641–649.
11. Perham, R. N., Packman, L. C. & Radford, S. E. (1988) *Biochem. Soc. Symp.* **54**, 67–81.
12. Altmann, M., Handschin, C. & Trachsel, H. (1987) *Mol. Cell. Biol.* **7**, 998–1003.
13. Uhlinger, D. J., Yang, C.-Y. & Reed, L. J. (1986) *Biochemistry* **25**, 5673–5677.
14. DeMarcucci, O. & Lindsay, J. G. (1985) *Eur. J. Biochem.* **149**, 641–648.
15. Uhlinger, D. J. (1986) Dissertation (Univ. of Texas, Austin).
16. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Arlington, VA), Vol. 1, pp. 49–78.
17. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
18. Cleveland, D. W., Fischer, S. G., Kirschner, M. W. & Laemmli, U. K. (1977) *J. Biol. Chem.* **252**, 1102–1106.
19. Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038.
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
21. Greene, J. R. & Guarente, L. (1987) *Methods Enzymol.* **152**, 519–522.
22. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
23. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
24. Sherman, F., Fink, G. R. & Hicks, J. B. (1986) *Laboratory Course Manual for Methods in Yeast Genetics* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
25. Browning, K. S., Uhlinger, D. J. & Reed, L. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1831–1834.
26. von Heijne, G. (1986) *EMBO J.* **5**, 1335–1342.
27. Gershwin, M. E., Mackay, I. R., Sturgess, A. & Coppel, R. L. (1987) *J. Immunol.* **138**, 3525–3531.
28. Yeaman, S. J., Fussey, S. P. M., Danner, D. J., James, O. F. W., Mutimer, D. J. & Bassendine, M. R. (1988) *Lancet* **i**, 1067–1070.
29. Fussey, S. P. M., Guest, J. R., James, O. F. W., Bassendine, M. F. & Yeaman, S. J. (1988) *Proc. Natl. Acad. Sci.*, in press.
30. Guest, J. R. (1987) *FEMS Microbiol. Lett.* **44**, 417–422.
31. Collins, J. H. & Reed, L. J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4223–4227.
32. Lau, K. S., Griffin, T. A., Hu, C.-W. C. & Chuang, D. T., (1988) *Biochemistry* **27**, 1972–1981.
33. Hummel, K. B., Litwer, S., Bradford, A. P., Aitken, A., Danner, D. J. & Yeaman, S. J. (1988) *J. Biol. Chem.* **263**, 6165–6168.
34. Guest, J. R., Lewis, H. M., Graham, L. D., Packman, L. C. & Perham, R. N. (1985) *J. Mol. Biol.* **185**, 743–754.
35. Kresze, G.-B. & Ronft, H. (1981) *Eur. J. Biochem.* **119**, 581–587.