Volume 240, number 1,2, 45–48 FEB 06495 November 1988

Nucleotide sequence of a cDNA for the dihydrolipoamide acetyltransferase component of human pyruvate dehydrogenase complex

Thomas J. Thekkumkara, Lap Ho, Isaiah D. Wexler, Gabriel Pons, Te-Chung Liu and Mulchand S. Patel

Departments of Biochemistry and Pediatrics, Case Western Reserve, University School of Medicine, Cleveland, OH 44106, USA

Received 28 September 1988

Deoxynucleotide sequencing of a cDNA for the dihydrolipoamide acetyltransferase (PDC-E_2) component of human pyruvate dehydrogenase complex (PDC) revealed an open reading frame of 1848 base pairs corresponding to a leader sequence of 54 amino acids and a mature protein of 561 amino acids (59551 Da). Both an amino-terminal lipoyl-bearing domain and a carboxy-terminal catalytic domain are present in the deduced amino acid sequence. The lipoyl-bearing domain contains two repeating units of 127 amino acids, each harboring one lipoic acid-binding lysine. Thus, mammalian PDC-E_2 differs as to the number of lipoic acid-binding sites from other dihydrolipoamide acyltransferases in both prokaryotic and eukaryotic organisms.

cDNA; Dihydrolipoamide acetyltransferase; Nucleotide sequence; Amino acid sequence; (Human liver)

1. INTRODUCTION

Dihydrolipoamide acetyltransferase (PDC-E_2, EC 2.3.1.12) is one of the six components of the mammalian pyruvate dehydrogenase complex (PDC) [1,2]. PDC-E_2 is encoded by a nuclear gene and is synthesized as a precursor with a leader sequence which is subsequently cleaved to generate the mature form in the mitochondria [3]. PDC-E_2 is made up of two heterologous domains: the carboxyterminal inner core-forming (catalytic) domain and the amino-terminal lipoyl-bearing domain, which are linked by a trypsin-sensitive hinge region [4]. Lipoic acid is covalently linked to a lysyl-residue of the lipoyl-bearing flexible domain thereby allowing the functional group of the lipoyl-lysine to interact with active sites residing in other components of the complex. Here, we report the deduced amino acid sequence of human PDC-E_2 including the core-forming and lipoyl-bearing domains. The deduced amino acid sequence establishes the presence of two lipoyl-binding sites in the lipoyl-bearing domain of human PDC-E_2. An analysis of the structural relationship of human PDC-E_2 as compared to the E_2 components of other α-ketoacid dehydrogenase complexes is also presented.

Correspondence address: M.S. Patel, Department of Biochemistry, Case Western Reserve, University School of Medicine, Cleveland, OH 44106, USA

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession no. Y00978

2. MATERIALS AND METHODS

2.1. *Sequencing strategy*

The isolation of seven λ recombinant phage containing cDNAs for human PDC-E_2 was previously reported [5]. These recombinants were digested with *Eco*RI or other restriction endonucleases and the resultant cDNA fragments were cloned into M13mp19. Overlapping directional deletion clones were generated starting with single-stranded M13mp19 template, RD20 primer (New England Biolabs), and T_4 DNA polymerase

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

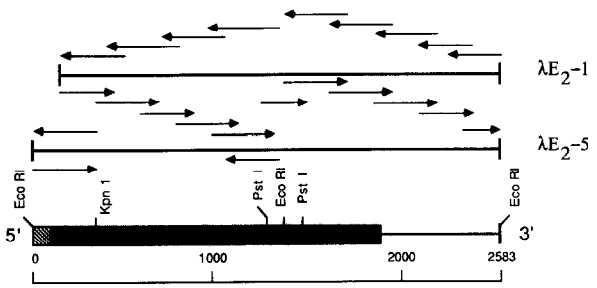

Fig.1. Partial restriction endonuclease map of PDC-E_2 cDNA clones derived from λE_2-1 and λE_2-5 and their sequencing strategy. Arrows indicate the position and direction of deletion fragments that were sequenced. The solid box and the hatched box represent the coding region of the mature PDC-E_2 and its leader sequence, respectively.

[6]. Deoxynucleotide sequencing of overlapping single-stranded M13mp19 clones was performed by the dideoxy chain-termination method [7] employing a modified T_7 DNA polymerase (Sequenase, US Biochemical).

2.2. *Northern analysis*

Total RNA from human heart and rat kidney was isolated and RNA blotting was performed as in [8], using a 1.4 kb *Eco*RI cDNA fragment of λE_2-1 labeled with [α-^{32}P]dCTP by the random priming method [9].

3. RESULTS AND DISCUSSION

A partial restriction endonuclease map of PDC-E_2 cDNA (2583 bp) and the DNA sequencing strategy are shown in fig.1. The deoxynucleotide sequence and deduced primary amino acid sequence from these clones are shown in fig.2. PDC-E_2 cDNA contained an open reading frame of 1848 base pairs which corresponded to a protein of 615 amino acids. The identity of these two clones (λE_2-1 and λE_2-5) as PDC-E_2 cDNA is established by the match of the deduced PDC-E_2 amino acid

Fig.2. Nucleotide sequence of PDC-E_2 cDNA and its deduced amino acid sequence. Nucleotides are numbered 5′ to 3′. Amino acid residues (one-letter code) of the mature form and its leader sequence are separately numbered with the mature peptide beginning at +1. X indicates the stop codon. The two lipoyl-binding sites are underlined (solid line) and the lipoyl-binding lysine within each site is marked by an asterisk (*). The two repeating units in the lipoyl-binding domain are bracketed. The E_3 binding site (amino acid residues 272–303) is marked by a broken underline and a highly conserved stretch of amino acids near the carboxy-terminal is identified by a double broken underline. Several potential polyadenylation signals are overlined.

```
CGAGTGACCTCGCGATCTGGCCCGGCTCCCGCTCGTCGCAACAGCGTGACTACAGGGTAT   60
 R  V  T  S  R  S  G  P  A  P  A  R  R  N  S  V  T  T  G  Y  (-35)
-54
GGCGGGGTCCGGGCACTGTGCGGCTGGACCCCCAGTTCTGGGGCCACGCCGCGGAACCGC  120
 G  G  V  R  A  L  C  G  W  T  P  S  S  G  A  T  P  R  N  R  (-15)

TTACTGCTGCAGCTTTTGGGGTCGCCCGGCCGCCGCTATTACAGTCTTCCCCCGCATCAG  180
 L  L  L  Q  L  L  G  S  P  G  R  R  Y  Y [S  L  P  P  H  Q  (06)
                                          -1 +1
AAGGTTCCATTGCCTTCTCTTTCCCCCACAATGCAGGCAGGCACCATAGCCCGTTGGAAA  240
 K  V  P  L  P  S  L  S  P  T  M  Q  A  G  T  I  A  R  W  K  (26)

AAAAAAGAGGGGGACAAAATCAATGAAGGTGACCTAATTGCAGAGGTTGAAACTGATAAA  300
 K  K  E  G  D  K  I  N  E  G  D  L  I  A  E  V  E  T  D  K*  (46)

GCCACTGTTGGATTTGAGAGCCTGGAGGAGTGTTATATGGCAAAGATACTTGTTGCTGAA  360
 A  T  V  G  F  E  S  L  E  E  C  Y  M  A  K  I  L  V  A  E  (66)

GGTACCAGGGATGTTCCCATCGGAGCGATCATCTGTATCACAGTTGGCAAGCCTGAGGAT  420
 G  T  R  D  V  P  I  G  A  I  I  C  I  T  V  G  K  P  E  D  (86)

ATTGAGGCCTTTAAAAATTATACACTGGATTCCTCAGCAGCACCTACCCCACAAGCGGCC  480
 I  E  A  F  K  N  Y  T  L  D  S  S  A  A  P  T  P  Q  A  A  (106)

CCAGCACCAACCCCTGCTGCCACTGCTTCGCCACCTACACCTTCTGCTCAGGCTCCTGGT  540
 P  A  P  T  P  A  A  T  A  S  P  P  T  P  S  A  Q  A  P  G  (126)

AGCTCATATCCCCCTCACATGCAGGTACTTCTTCCTGCCCTCTCTCCCACCATGACCATG  600
 S][S  Y  P  P  H  M  Q  V  L  L  P  A  L  S  P  T  M  T  M  (146)

GGCACAGTTCAGAGATGGGAAAAAAAAGTGGGTGAGAAGCTAAGTGAAGGAGACTTACTG  660
 G  T  V  Q  R  W  E  K  K  V  G  E  K  L  S  E  G  D  L  L  (166)

GCAGAGATAGAAACTGACAAAGCCACTATAGGTTTTGAAGTACAGGAAGAAGGTTATCTG  720
 A  E  I  E  T  D  K*  A  T  I  G  F  E  V  Q  E  E  G  Y  L  (186)

GCAAAAATCCTGGTCCCTGAAGGCACAAGAGATGTCCCTCTAGGAACCCCACTCTGTATC  780
 A  K  I  L  V  P  E  G  T  R  D  V  P  L  G  T  P  L  C  I  (206)

ATTGTAGAAAAAGAGGCAGATATATCAGCATTTGCTGACTATAGGCCAACCGAAGTAACA  840
 I  V  E  K  E  A  D  I  S  A  F  A  D  Y  R  P  T  E  V  T  (226)

GATTTAAAACCACAAGTGCCACCACCTACCCCACCCCCGGTGGCCGCTGTTCCTCCAACT  900
 D  L  K  P  Q  V  P  P  P  T  P  P  P  V  A  A  V  P  P  T  (246)

CCCCAGCCTTTAGCTCCTACACCTTCAGCACCCTGCCCAGCTACTCCTGCTGGACCAAAG  960
 P  Q  P  L  A  P  T  P]  S  A  P  C  P  A  T  P  A  G  P  K  (266)

GGAAGGGTGTTTGTTAGCCCTCTTGCAAAGAAGTTGGCAGTAGAGAAAGGGATTGATCTT 1020
 G  R  V  F  V  S  P  L  A  K  K  L  A  V  E  K  G  I  D  L  (286)
                    ----------------------------------------

ACACAAGTAAAAGGGACAGGACCAGATGGTAGAATCACCAAGAAGGATATCGACTCTTTT 1080
 T  Q  V  K  G  T  G  P  D  G  R  I  T  K  K  D  I  D  S  F  (306)
 ----------------------------------------

GTGCCTAGTAAAGTTGCTCCTGCTCCGGCAGCTGTTGTGCCTCCCACAGGTCCTGGAATG 1140
 V  P  S  K  V  A  P  A  P  A  A  V  V  P  P  T  G  P  G  M  (326)

GCACCAGTTCCTACAGGTGTCTTCACAGATATCCCAATCAGCAACATTCGTCGGGTTATT 1200
 A  P  V  P  T  G  V  F  T  D  I  P  I  S  N  I  R  R  V  I  (346)

GCACAGCGATTAATGCAATCAAAGCAAACCATACCTCATTATTACCTTTCTATCGATGTA 1260
 A  Q  R  L  M  Q  S  K  Q  T  I  P  H  Y  Y  L  S  I  D  V  (366)

AATATGGGAGAAGTTTTGTTGGTACGGAAAGAACTTAATAAGATATTAGAAGGGAGAAGC 1320
 N  M  G  E  V  L  L  V  R  K  E  L  N  K  I  L  E  G  R  S  (386)

AAAATTTCTGTCAATGACTTCATCATAAAAGCTTCAGCTTTGGCATGTTTAAAAGTTCCC 1380
 K  I  S  V  N  D  F  I  I  K  A  S  A  L  A  C  L  K  V  P  (406)

GAAGCAAATTCTTCTTGGATGGACACAGTTATAAGACAAAATCATGTTGTTGATGTCAGT 1440
 E  A  N  S  S  W  M  D  T  V  I  R  Q  N  H  V  V  D  V  S  (426)

GTTGCGGTCAGTACTCCTGCAGGACTCATCACACCTATTGTGTTTAATGCACATATAAAA 1500
 V  A  V  S  T  P  A  G  L  I  T  P  I  V  F  N  A  H  I  K  (446)

GGAGTGGAAACCATTGCTAATGATGTTGTTTCTTTAGCAACCAAAGCAAGAGAGGGTAAA 1560
 G  V  E  T  I  A  N  D  V  V  S  L  A  T  K  A  R  E  G  K  (466)

CTACAGCCACATGAATTCCAGGGTGGCACTTTTACGATCTCCAATTTAGGAATGTTTGGA 1620
 L  Q  P  H  E  F  Q  G  G  T  F  T  I  S  N  L  G  M  F  G  (486)

ATTAAGAATTTCTCTGCTATTATTAACCCACCTCAAGCATGTATTTTGGCAATTGGTGCT 1680
 I  K  N  F  S  A  I  I  N  P  P  Q  A  C  I  L  A  I  G  A  (506)

TCAGAGGATAAACTGGTCCCTGCAGATAATGAAAAAGGGTTTGATGTGGCTAGCATGATG 1740
 S  E  D  K  L  V  P  A  D  N  E  K  G  F  D  V  A  S  M  M  (526)

TCTGTTACACTCAGTTGTGATCACCGGGTGGTGGATGGAGCAGTTGGAGCCCAGTGGCTT 1800
 S  V  T  L  S  C  D  H  R  V  V  D  G  A  V  G  A  Q  W  L  (546)
                    ==========================

GCTGAGTTTAGAAAGTACCTTGAAAAACCTATCACTATGTTGTTGTAACTAACTCAAGAA 1860
 A  E  F  R  K  Y  L  E  K  P  I  T  M  L  L  X              (561)

TTTCTAAACTCTCCCAGGTCACACTGATTCATTCTTAACAAGATATTTATATGTTATTAA 1920
ACAGGTGGTTGCTTTTTATTTTAACCAGTTATTTTTATTATTGAGTCTGCTCAGATAAGT 1980
TATTTATAATGGGCATTACTGAATTTTTAAAATGCCGATTACACCCAAATATTGTGCACA 2040
TTTAATAATCAGACACCAGATTTTTAGCTCTGTACTCCTAATTAAGGGACATGTATGTGG 2100
CCTTGCCTAGCCCTTTGGTGATAAGTACTTCCTCTAGGAAATGTACGATAGGTAGAATTG 2160
TGGTTCCCTAAAGACAAGTACATAAAGGTGACCCTGATGAAACCTTGAAGTTCTGAAATT 2220
TAACTGCCTAAAATGTTCTCCTTAGATGTGAGAGAAAAGAGAAATCAGAAAAATTAATTCT 2280
CTTGGGGGAAGGGCTTGAATTGAAGCTTTACTTTAGAATTTAGCCCTGGTTTGAAATTTT 2340
CCATTACATGATCTTGGTTTATCATCGATGGGAAGGGTAGAAAACTTCAAGGAAAATAAG 2400
TGAAATTTTAAAAGTCAGCATTTTCTTAGACCTCTTCAGCTGATTGTTTATTTTTCTATG 2460
AATTCTGTACTCAAAATTACACATTTGTTTAAATAAATATCCACACAAATTCTCAGTTAC 2520
ATCAAGTAGCTGGTTTATATTTAGATTATCTCAAGTAGCGGCGAATAACCATGTGTAGGA 2580
ATT                                                           2583
```

sequence with (i) a known sequence of 10 amino acids surrounding the lipoyl-binding lysine residue of bovine kidney PDC-E_2 [10], and (ii) 22 of 23 residues of the amino acid sequence at the amino-terminal of bovine heart PDC-E_2. [The amino-terminal amino acid sequence (SLPPHEKVPL-PSLSPTMQAGTIA) generated from trypsin digestion of the lipoyl-bearing domain of bovine heart PDC-E_2 (Thomas E. Roche, Kansas State University; personal communication).] The calculated molecular mass of the deduced amino acid sequence of the mature PDC-E_2 (561 amino acids) is 59551 Da. The molecular mass of the leader sequence of PDC-E_2 is estimated to be approx. 6000 Da. This is consistent with the previously reported mass based on mobility of the precursor PDC-E_2 on SDS-PAGE [3].

The deduced amino acid sequence of human PDC-E_2 shows the presence of the amino-terminal lipoyl-bearing domain and the carboxy-terminal catalytic domain, both of which have been demonstrated to be present in PDC-E_2 based on the proteolytic cleavage of mammalian PDC-E_2 [4] and the sequencing of *E. coli* PDC-E_2 [11]. Our data show that human PDC-E_2 contains two lipoyl-binding sites [amino acid residues 35–52 (site I) and 162–179 (site II) in figs 2,3] in the lipoyl-bearing domain. The second lipoyl-binding site (site II) shows approx. 83% homology (15 out of 18 residues) with the first (site I) and the differences that exist between the two lipoyl-binding sites are conservative. The amino acid sequence of site II of human PDC-E_2 is also identical to a lipoyl-binding site present in a deduced protein sequence of rat liver PDC-E_2 cDNA [12,13].

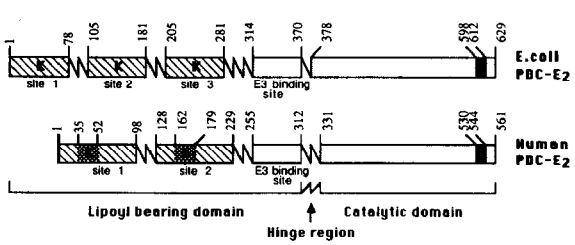

Fig.3. Comparison of the domain structures of *E. coli* and human liver PDC-E_2. The limits of these domains are approximate. K, the lipoyl-binding lysine; (▨) lipoyl-bearing repeating units; (▩) the highly conserved region surrounding the lipoyl-binding lysine residue; (□) E_3-binding site; (▩) catalytic domain; (■) the conserved region in the catalytic domain; and (ᴍ) the alanine-proline rich region.

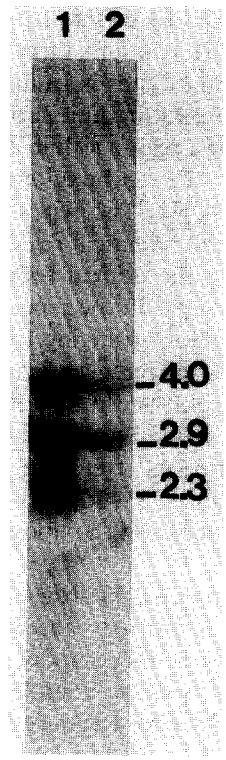

Fig.4. Northern blot analysis of total RNA isolated from human heart (25 μg; lane 1) and rat kidney (50 μg; lane 2). The size of the different mRNA species was determined by comparing their electrophoretic mobility to that of ribosomal RNA.

However, the deduced amino acid sequence of PDC-E_2 from rat liver did not extend far enough into the amino-terminus of the protein to include the lipoyl-binding site I.

An earlier study suggested that there was only one functional lipoic acid-binding site per molecule of PDC-E_2 based on radiolabeled acetylation of mammalian PDC [4]. Recently, Bradford et al. [14] were able to identify two different but closely related short peptides containing lipoyl-lysine residues. These investigators proposed that there are either two or more lipoyl-binding sites per mammalian PDC-E_2 polypeptide or possible microheterogeneity of bovine PDC-E_2 [14]. Our results (fig.2) demonstrate the presence of two lipoyl-binding sites per human PDC-E_2 polypeptide. This is in contrast to the E_2 component of mammalian branched-chain α-ketoacid dehydrogenase complex which contains only one lipoyl-binding site per polypeptide [13,15].

Mammalian PDC-E_2 differs from *E. coli* PDC-E_2 in that the latter contains three repeating units of approx. 100 amino acids in its lipoyl-bearing domain [10]. Each repeating unit surrounds a lipoyl-binding lysine and a region rich in alanine and proline residues at the carboxy-terminal of each repeating unit [10,11]. The repeating units in *E. coli* PDC-E_2 contain identical regions of 18 amino acids surrounding the critical lysine residue. The remaining amino acids in the repeating units show considerable homology (ranging from 69 to 80%), with many of the differences being conservative substitutions. In contrast, human PDC-E_2 includes two similar repeating units of 127 amino acids each containing a conserved stretch of 38 amino acids (amino acid residues 35–72 and 162–199; 79% homology) encompassing the lipoyl-binding lysine residue (figs 2,3).

A dihydrolipoamide dehydrogenase (E_3) binding site has been identified based on sequence homology among the E_2 components of the three α-ketoacid dehydrogenase complexes [13]. A sequence of 32 amino acids (from residues 272 to 303 in figs 2,3) which is highly homologous to the E_3-binding site [13], is also present in human PDC-E_2. Overall, there is approx. 90% homology in the E_3-binding site between human and rat liver PDC-E_2 [12] and 53% homology between human and *E. coli* PDC-E_2 [11]. Further comparison of the amino acid sequences of the catalytic domains of PDC-E_2 from human liver and *E. coli* revealed [11] a segment of 14 amino acids (residues 530–543 in figs 2,3) near the carboxy-terminal with approx. 79% homology:

Human liver PDC-E_2: ...L S C D H R V V D G A V G A ...
E. coli PDC-E_2: ...L S F D H R V I D G A D G A ...

Although the nucleotide sequence encoding this region in rat liver PDC-E_2 cDNA [12] is largely preserved, the deduced amino acid sequence of this segment of rat liver PDC-E_2 is apparently altered due to a shift in its reading frame. A histidine residue is conserved in both sequences and this amino acid may serve as the active site residue for catalysis.

Northern blot analysis of RNA isolated from human heart and rat kidney showed the presence of three hybridizing species (fig.4). This could be either due to the presence of multiple polyadenylation signals in the 3′ untranslated region of PDC-E_2, alternative splicing sites, or multiple transcription initiation sites.

Acknowledgements: This work was supported by National Institutes of Health Grant AM 20478 and Metabolism Training Grant AM 07319 (L.H. and I.D.W.). We are indebted to Dr Thomas E. Roche for allowing us to include his unpublished partial aminoterminal sequence of bovine heart PDC-E_2.

REFERENCES

[1] Reed, L.J. (1974) Acc. Chem. Res. 7, 40–46.
[2] Yeaman, S.J. (1986) Trends Biochem. Sci. 11, 293–296.
[3] De Marcucci, O.G.L., Gibb, G.M., Dick, J. and Lindsay, J.G. (1988) Biochem. J. 251, 817–823.
[4] Bleile, D.M., Munk, P., Oliver, R.M. and Reed, L.J. (1979) J. Biol. Chem. 256, 514–519.
[5] Thekkumkara, T.J., Jesse, B.W., Ho, L., Raefsky, C., Pepin, R.A., Javed, A.A., Pons, G. and Patel, M.S. (1987) Biochem. Biophys. Res. Commun. 145, 903–907.
[6] Dale, R.M., McClure, B.A. and Houchins, J.P. (1985) Plasmid 13, 13–41.
[7] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[8] Hod, Y., Morris, S.M. and Hanson, R.W. (1984) J. Biol. Chem. 259, 15603–15608.
[9] Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.
[10] Spencer, M.E., Darlison, M.G., Stephens, P.E., Duckenfield, I.K. and Guest, J.R. (1984) Eur. J. Biochem. 141, 361–374.
[11] Stephens, P.E., Darlison, M.G., Lewis, H.M. and Guest, J.R. (1983) Eur. J. Biochem. 133, 481–489.
[12] Gershwin, M.E., Mackay, I.R., Sturgess, A. and Coppel, R.L. (1987) J. Immunol. 138, 3525–3531.
[13] Hummel, K.B., Litwer, S., Bradford, A.P., Aitken, A., Danner, D.J. and Yeaman, S.J. (1988) J. Biol. Chem. 263, 6165–6168.
[14] Bradford, A.P., Howell, S., Aitken, A., James, L.A. and Yeaman, S.J. (1987) Biochem. J. 245, 919–922.
[15] Lau, K.S., Griffin, T.A., Hu, C.-W.C. and Chuang, D.T. (1988) Biochemistry 27, 1972–1981.