Structural Organization of the Gene for the E1α Subunit of the Human Pyruvate Dehydrogenase Complex*

(Received for publication, December 15, 1988)

Constantina Maragos, Wendy M. Hutchison, Kiyoshi Hayasaka†, Garry K. Brown§, and Hans-Henrik M. Dahl¶

From the Murdoch Institute for Research into Birth Defects, The Royal Children's Hospital, Melbourne, Australia 3052 and the § Department of Pediatrics, University of Melbourne, Melbourne, Australia 3052

The pyruvate dehydrogenase (PDH)† complex is one of the major enzyme systems involved in the control of aerobic energy metabolism. This complex, with a M₉ of about 7 × 10⁶, converts pyruvate to acetyl-CoA within the mitochondrial matrix. The catalytic activity of the complex is maintained by three enzyme components: pyruvate decarboxylase (E1, EC 1.2.4.1), dihydrolipoamide acyltransferase (E2, EC 2.3.1.12), and dihydrolipoamide dehydrogenase (E3, EC 1.8.1.4) (1). The E1 enzyme component is a tetramer composed of two α and two β subunits. The main mechanism for regulation of PDH complex activity is a cycle of phosphorylation and dephosphorylation (2). Phosphorylation occurs on 3 serine residues on the E1α subunit by a specific PDH kinase. However, phosphorylation of only 1 of these residues is required for inactivation of the whole complex (3). The removal of phosphates by a specific phosphatase restores activity. Genetic defects in the PDH complex are reported to be the most common cause of primary lactic acidosis in humans (4).

* This work was supported in part by the National Health and Medical Research Council of Australia. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Present address: Department of Pediatrics, School of Medicine, Tohoku University, Sendai, Japan.

‡ To whom correspondence and reprint requests should be addressed.

The abbreviations used are: PDH, pyruvate dehydrogenase; GRE, glucocorticoid-responsive element; IR, inverted repeat; DR, direct repeat; kb, kilobase pair(s); bp, base pair(s).

In the majority of cases, the basic defect appears to be in the E1α subunit. We have previously reported the isolation of cDNA clones of the entire length of the mRNA for the PDH E1α subunit. In addition, the gene for the functional E1α subunit has recently been localized to the X chromosome (6). Here we report the gene structure of the human PDH E1α subunit. This will facilitate further studies on the expression and regulation of this gene and will provide necessary information for analysis of mutations in patients with PDH E1α deficiency.

EXPERIMENTAL PROCEDURES

Materials—Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase, S1 nuclease, oligo(dT)-cellulose, M13mp18, M13mp19, and pUC9 vector DNAs were purchased from B. Eng. Mannheim. The M13 sequencing kit (Sequenase) was obtained from United States Biochemical Corp. Bluescribe M13- vector was obtained from Stratagene. GeneScreen Plus and nitrocellulose hybridization membranes were obtained from Du Pont-New England Nuclear and Schleicher and Schuell, respectively. SeaKem HGT agarose was from FMC Bioproducts. Oligo-labeling kit was purchased from BRESA. 

[a-³²P]dCTP (specific activity, 3000 Ci/mmol) for oligo-labeling, [γ-³²P]ATP (3000 Ci/mmol) for end-labeling, and [α-³²P]dATP (2000 Ci/mmol) for sequencing were produced from Amersham Corp. Synthetic oligonucleotide primers were obtained from Dr. N. Hoogenraad, Dept. of Biochemistry, La Trobe University, Melbourne, Australia.

Isolation and Screening of Cosmid Genomic Clones—The construction of a human leukocyte genomic DNA library using the cosmid vector pCVOO1 has been described previously (7). Approximately 90,000 recombinant colonies were transferred to nitrocellulose membranes and hybridized to a [³²P]oligo-labeled human PDH E1α subunit cDNA probe (clone PDH1) (5). DNA was extracted from positive colonies by alkaline lysis (8) and characterized by restriction enzyme sites were determined more precisely by a series of partial digestions of linearized cosmid DNA (cleaved at the SalI vector site) followed by electrophoresis on 0.7% agarose gels and hybridization to [³²P]oligo-labeled PstI (2.3 kb) and SalI/ClaI (2.5 kb) fragments which were isolated from the vector pCVOO1. These fragments flank both ends of the polylinker cloning site of pCVOO1. Restriction sites were positioned with increasing order of fragment sizes from the probe. In addition, restriction enzyme sites were determined more precisely by a series of single and double digests (8).

DNA fragments containing the PDH E1α gene were determined by blotting the restriction enzyme-digested DNA onto GeneScreen Plus membranes followed by hybridization to a [³²P]oligo-labeled PDH E1α subunit cDNA probe. A HindIII fragment (approximately 3.2 kb in size) containing the promoter region of the gene was also identified by hybridization to a [γ-³²P]ATP end-labeled synthetic oligonucleotide probe specific to the 5'-untranslated end of the PDH E1α subunit cDNA. Exon- and promoter-containing fragments were subsequently subcloned into either Bluescribe M13- or pUC9 vectors for further analysis.

Restriction enzyme sites of exon-containing fragments were mapped by a series of single and double digests. Exons were located
by hybridization to [3H]poly-labeled exon regions (determined by sequencing genomic DNA as described below) or using synthetic oligonucleotide probes (determined from the PDH Ela subunit cDNA sequence). In addition, exon locations were confirmed on the basis of unique restriction endonuclease sites deduced from the cDNA sequence.

DNA Sequence Analysis—To determine exon regions, cosmid DNA was digested with Sau3Al restriction endonuclease and "shotgun" sequencing reactions were separated by electrophoresis on denaturing 5% polyacrylamide/urea gels. Gels were dried with heat under vacuum and autoradiographed. In all, approximately two-thousand clones were screened for the presence of exon segments using PDH Ela subunit cDNA or specific cDNA restriction endonuclease fragments as probes. Single-stranded DNA isolated from positive clones was sequenced using the Sequenase kit. The products of the sequencing reactions were separated by electrophoresis on denaturing 5% polyacrylamide/urea gels. Gels were dried with heat under vacuum and autoradiographed. Alternatively, for exons containing unique restriction endonuclease sites (deduced from the cDNA sequence), the appropriate enzymes were used to generate DNA fragments in the cosmid DNA so that, after cloning the purified fragments into M13mp18 or -mp19, the exons were adjacent to the priming site in the vector for sequence analysis. In addition, some exons were sequenced directly from double-stranded cosmid DNA using exon-specific synthetic oligonucleotide primers specific to both ends of the exon. In total, ten 18-base primers were used.

The nucleotide sequence of the promoter region was determined by inserting double-stranded genomic DNA into pUC9 and sequencing both DNA strands using universal sequencing primers and three 17-mer synthetic oligonucleotide primers. After determining approximately 200 bases using the first primer (specific to the 5'-untranslated end of the PDH Ela cDNA), a new 17-base primer (specific to the 5'-end of the previous nucleotide sequence) was used to sequence approximately another 200 bases and so on. Using this approach, 767 bases of the promoter region were sequenced.

**S1 Nuclease Mapping**—Total RNA was extracted from human heart tissue (10) and poly(A) RNA purified on oligo(dT)-cellulose columns (11). The probe for S1 nuclease analysis was a 96-bp HindII genomic DNA fragment, spanning nucleotides -21 to +78, labeled with [3P] using T4 polynucleotide kinase (8). S1 nuclease analysis was essentially as described (12). Hybridization temperature was 68°C for 3 h.

**RESULTS**

Isolation and Characterization of the Gene for the Human PDH Ela Subunit—A human leukocyte genomic DNA library constructed using the cosmid vector pCV001 was screened with the human PDH Ela subunit cDNA probe. Two overlapping positive clones, cPDH 4a and cPDH 3, spanning approximately 50 kb of tandem genomic DNA were isolated and mapped (Fig. 1). Southern blot analysis of digested cosmid DNA probed with the full length PDH Ela cDNA probe indicated that the major portion of the gene was localized within two EcoRI/BamHI restriction endonuclease fragments of 5.2 and 4.0 kb. Both fragments are contained within the region of overlap of the two cosmid clones. The cDNA probe also hybridized to an 8.8- to 8.9-kb EcoRI fragment located outside the overlap region on cPDH 4a, indicating that the gene extends into this fragment. Comparison of Southern blot analysis of genomic DNA and cDNA, digested with a variety of restriction enzymes and hybridized to the PDH Ela cDNA probe, suggested that the entire gene is contained completely within these fragments. The promoter region of the gene is present within the 8.8- to 8.9-kb EcoRI fragment and on a 3.2-kb HindIII fragment as shown by hybridization to a synthetic oligonucleotide probe specific to the 5'-untranslated end of the PDH Ela subunit cDNA (data not shown).

**Detailed Structure of the Gene for the Human PDH Ela Subunit**—Fig. 1 and Table I show the detailed structure of the PDH Ela gene. Exon-intron boundaries were determined by sequencing the appropriate regions of cosmid DNA selected as described under "Experimental Procedures" and aligning them with the PDH Ela subunit cDNA sequence. Exon positions were subsequently mapped by alignment of restriction enzyme cleavage sites and hybridization of exon-specific fragments or exon-specific synthetic oligonucleotide probes to Southern blotted digested cosmid DNA. The gene for the human PDH Ela subunit contains 11 exons spanning approximately 17 kb of genomic DNA. As shown in Table I, the 11 exons range in size from 61 bp (exon 2) to 355 bp (exon 11; including the 5'-untranslated region) and the 10 introns range from approximately 200 bp (introns 2 and 10) to 5.7 kb (intron 1, including the 5'-untranslated region). All splice junction sequences flanking the introns conform to the consensus splice junction sequences and the GT/AG splice rule (13, 14). DNA sequence analysis of the two cosmid clones also enabled us to resolve differences between the published cDNA sequences (5, 18, 19). We have amended our previously published sequence as shown in Table II. The most important aspect of this correction is that the reading frame is in agreement with that of DeMeuleire et al. (19). In addition, the sequence analysis has confirmed the guanines at positions 178, 516, 823, and 1156, the cytosine at position 824, and the adenosine at position 1171.

**Preliminary S1 nuclease** (Fig. 2) and primer extension experiments, to determine the transcription initiation site, suggest that the 5' end of the PDH Ela mRNA is the adenosine residue, 105 bp upstream from the translation initiation codon. Exon 1 is therefore 158 bp, contains the 5'-untranslated region upstream of the initiation methionine codon, and codes for the first 19 amino acids (Table I). Further characterization of the transcription initiation site is in progress.

The promoter region of the gene for the human PDH Ela subunit is shown in Fig. 3. A "TATA box"-like sequence (TTATTA) is present at position -94 to -98. The "CAAT box"-like sequence (CCATTA) is present at position -117 to -113. Two possible binding sites for the cellular transcription factor Sp1 (15) are present at positions -89 to -84 (GGCCGG) and -175 to -170 (GGGCCC antisense). A sequence identical with the glucocorticoid responsive element (GRE, ACCA metals) is absent from the N-terminal region of the gene. Whether the CACCC box-like sequences in the gene for the human PDH Ela subunit are active has been described previously for other genes including the human metallothionein IIA (26) and the rat tryptophan oxygenase gene (27). When situated immediately upstream of the GRE, the CACCC box element has been implicated in the cooperative induction of transcription with the GRE (27). Whether the CACCC box-like sequences in the gene for the human PDH Ela subunit downstream from the GRE are functional or not remains to be determined. Two possible CAMP-responsive elements (CGCGCGCC) specifically resembling the activator protein 2 (AP-2) binding site (17) are present at -55 to -48 and -242 to -238. Several sets of inverted and direct repeats are also present.
The Human Pyruvate Dehydrogenase E1α Gene Structure

### TABLE I

| Exon | Size (bp) | Intron (kb) |
|------|-----------|-------------|
| 1    | 158       | 1           |
| 2    | 61        | 2           |
| 3    | 174       | 3           |
| 4    | 126       | 4           |
| 5    | 92        | 5           |
| 6    | 94        | 6           |
| 7    | 156       | 7           |
| 8    | 72        | 8           |
| 9    | 68        | 9           |
| 10   | 109       | 10          |
| 11   | 359       |             |

Exon-intron splice junction sequences of the gene for the human PDH E1α subunit

| Exon   | Size (bp) | Intron (kb) | Exon* | Intron* | Exon |
|--------|-----------|-------------|-------|---------|------|
| 1      | 158       | 1           | AAG CCG Pro 19 | gttgacc....cttttaag GCA AGC Ala 20 |
| 2      | 61        | 2           | ATT AAG Lys 29 | gtaagagg....catttcag AAA TGT Lys 40 |
| 3      | 174       | 3           | GCT CAG Gln 97 | gttgagtg....taucacag GAA GCT Gln 98 |
| 4      | 126       | 4           | ACA G Gly 140 | gtttgtctg....atttccag GA CAG Gly 140 |
| 5      | 92        | 5           | GCG CAG Gln 170 | gtaagcc....astgttag GTG CCC Val 171 |
| 6      | 94        | 6           | AAC CAG Gln 201 | gtaattat....acttccag GGC CAG Gly 202 |
| 7      | 156       | 7           | CTG AGA Arg 253 | gtaagac....attaccag GTG GAT Val 254 |
| 8      | 72        | 8           | GGG AAG Lys 277 | gtaagcct....gttttag GGG CCC Gly 278 |
| 9      | 68        | 9           | GTC AG Ser 300 | gtacgcct....ccccatag T TAC Ser 300 |
| 10     | 109       | 10          | CTA AAG Lys 336 | gtaccgc....atttttag GAA ATT Glu 337 |

### DISCUSSION

The functional gene for the human PDH E1α subunit has been mapped to the p22.1 to 22.2 region of the X chromosome (6). This gene is approximately 17 kb in length and contains 11 exons as characterized from two cosmid clones.

DNA sequence analysis of exons in the two cosmid clones provided a means of comparing exons with the previously published PDH E1α cDNA sequences (5, 18, 19). The sequence reported by Koike et al. (18) differs from ours (5) in that it lacks a 93-bp segment which would occupy position 621 to 714 of our sequence. This could be due to a sequencing error. However, the deleted segment corresponds exactly to the complete sequence of exon 6. The PDH E1α cDNA clone of Koike's group originated from a human foreskin fibroblast cDNA library, in contrast to our PDH E1α cDNA clone which originated from a human liver cDNA library. Whether the deleted segment in the cDNA sequence of Koike's group represents differential or incorrect splicing of a primary transcript is not known at this time. Another possibility is that the PDH E1α cDNA sequence reported by Koike's group originated from a "processed" pseudogene. Examples of processed human genes include metallothionein (21), β-tubulin (22), immunoglobulin (23), and phosphoglycerate kinase (24). We have recently reported a second locus on chromosome 4 which hybridizes with a weaker signal to the PDH E1α subunit cDNA probe (6). This locus may be a pseudogene or a gene for a closely related protein. It has not yet been determined whether sequences at the chromosome 4 locus are transcribed. Previous Northern blot analysis with the PDH E1α cDNA probe of the RNA from human tissues (skeletal muscle, cardiac muscle, liver, kidney, and brain) and human skin fibroblast cells demonstrated two different mRNA species (5). However, these two mRNAs which differed only in the 3'-untranslated region were both shown to be deficient in a patient with severe PDH deficiency (20), thus making it unlikely that they are derived from different gene loci. Two mRNA species were similarly detected by DeMeirleir et al. (19) in human skin fibroblast cells. In contrast, Koike's group detected only one mRNA species which may be characteristic of the HeLa cell line used to extract poly(A)* RNA.

The transcription initiation site is located in a very GC-
rich region. Appropriately located in the putative promoter region of the gene for the human PDH Ela subunit are sequences resembling the TATA box (13), CAAT box (13), and Sp1 binding elements (15). These features are characteristic of “housekeeping” genes (25). The promoter region also contains inverted and direct DNA repeats. However, the significance of these repeats, if any, is unknown. Such inverted and palindromic repeats are capable of forming secondary structures that may be involved in the control and/or modification of gene expression. The precise role of glucocorticoid responsive element (GRE), and CAMP receptor elements which are also present is not known at this time. In preliminary experiments, dexamethasone alone had no significant effect on PDH Ela mRNA levels from human skin fibroblast cells after treatment with this steroid.

Having characterized the normal gene for the human PDH Ela subunit, further studies on the expression and regulation of this gene are underway. These studies will provide necessary information for the analysis of mutations and their effects in patients with PDH Ela deficiency.

**Fig. 2.** S1 nuclease analysis of the 5’-end of the human PDH Ela mRNA. Lane 1 displays S1 mapping with the 96-bp HinfI genomic probe. Dideoxynucleotide sequencing reactions (lanes A, C, G, and T) of the probe are also shown. The position of the 5’-end of the mRNA is indicated by the arrow. The band above is the reannealed probe.

**Fig. 3.** Nucleotide sequence analysis of the promoter region of the gene for the human PDH Ela subunit. The promoter region of the gene is numbered −767 to −1. +1 to +13 denotes part of the first exon. The boxed areas denote the following elements: TATA box, CAAT box, Sp1 binding sites (Sp1), consensus sequence for glucocorticoid responsive element (GRE), and CAMP receptor binding sites (AP-2), respectively. The brackets with the intervening broken line marked IR-1 to -6 denote inverted repeats. Paired arrows DR-1 and -2 denote direct repeats.

**Acknowledgments**—We wish to thank Dr. K. H. Choo of the Murdoch Institute, Melbourne, for providing the human leukocyte genomic DNA library and Dr. N. Hoogenraad of the Department of Biochemistry, La Trobe University, Melbourne, for synthesizing the oligonucleotide probes. We thank Professor D. M. Danks, R. M. Brown, and Dr. H. J. McArdle for their helpful discussions. The assistance of K. Bromley for the preparation of the figures and J. Humphreys for typing this manuscript is also acknowledged.

**REFERENCES**

1. Reed, L. J., Pettit, F. H., and Yeaman, S. J. (1978) in *Microenvironments and Metabolic Compartmentation* (Sere, P. A., and Estabrook, R. W., eds) pp. 305–321, Academic Press, Orlando, FL.
2. Linn, T. C., Pettit, F. H., and Reed, L. J. (1969) *Proc. Natl. Acad. Sci. U. S. A.* 62, 234–241.
3. Yeaman, S. J., Hutcheson, E. T., Roche, T. E., Pettit, F. H., Brown, J. R., Reed, L. J., Watson, P. C., and Dixon, G. H. (1978) *Biochemistry* 17, 2364–2370.
4. Robinson, B. H., Taylor, J., and Sherwood, W. G. (1980) *Pediatr. Res.* 14, 956–962.
5. Dahl, H.-H. M., Hunt, S. M., Hutchison, W. M., and Brown, G. K. (1987) *J. Biol. Chem.* 262, 7398–7403.
6. Brown, R. M., Dahl, H.-H. M., and Brown, G. K. (1989) *Genomics* 4, 174–181.
7. Choo, K. H., Filby, G., Greco, S., Lau, Y. F., and Kan, Y. W. (1986) *Gene* (Amst.) 46, 277–286.
8. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
9. Messing, J. (1983) *Methods Enzymol.* 101, 20–78.
10. Wake, S. A., and Mercer, J. F. B. (1985) Biochem. J. 228, 425-432
11. Avid, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408-1412
12. Favaloro, J., Treisman, R., and Kamen, R. (1980) Methods Enzymol. 65, 718-749
13. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-383
14. Mount, S. M. (1982) Nucleic Acids Res. 10, 459-472
15. Kadonaga, J. T., Jones, K. A., and Tjian, R. (1986) Trends Biochem. Sci. 11, 20-23
16. Jantzen, H. M., Strahle, U., Gloss, B., Stewart, F., Schmid, W., Boshart, M., Miksicek, R., and Schutz, G. (1987) Cell 49, 29-38
17. Roesler, W. J., Vandenbark, G. R., and Hanson, R. W. (1988) J. Biol. Chem. 263, 9063-9066
18. Koike, K., Ohta, S., Urata, Y., Kagawa, Y., and Koike, M. (1986) Proc. Natl. Acad. Sci. U. S. A. 86, 41-45
19. De Meirleir, L., MacKay, N., Lam Hon Wah, A. M., and Robinson, B. H. (1988) J. Biol. Chem. 263, 1991-1995
20. Brown, G. K., Scholten, R. D., Hunt, S. M., Harrison, J. R., and Pollard, A. C. (1987) J. Inherited Metab. Dis. 10, 359-366
21. Karin, M., and Richards, R. I. (1982) Nature 299, 797-802
22. Wilde, C. D., Crowther, C. E., Cripe, T. P., Lee, M. G.-S., and Cowan, N. M. (1982) Nature 297, 83-84
23. Hollis, G. P., Hieter, P. A., McBride, O. N., Swan, D., and Leder, P. (1982) Nature 296, 321-325
24. McCarrey, J. R., and Thomas, K. (1987) Nature 326, 501-505
25. Martini, G., Toniolo, D., Vulliamy, T., Luzzatto, L., Dono, R., Viglietto, G., Paonessa, G., D'Urso, M., and Persico, M. G. (1986) Embo J. 5, 1849-1855
26. Lee, W., Haslinger, A., Karin, M., and Tjian, R. (1987) Nature 325, 368-372
27. Schüle, R., Muller, M., Otsuka-Murakami, H., and Renkawitz, R. (1988) Nature 332, 87-90