*Proc. Natl. Acad. Sci. USA*
Vol. 87, pp. 5594–5597, August 1990
Biochemistry

# Molecular cloning and characterization of human pyruvate dehydrogenase β subunit gene

**(thiamin pyrophosphate-dependent enzyme/DNA sequence/intron–exon junctions/promoter)**

KICHIKO KOIKE*, YOSHISHIGE URATA, AND MASAHIKO KOIKE

Department of Pathological Biochemistry, Atomic Disease Institute, Nagasaki University School of Medicine, Nagasaki, 852, Japan

*Communicated by Lester J. Reed, May 7, 1990*

**ABSTRACT** **A genomic clone encompassing the entire gene for the human pyruvate dehydrogenase β subunit (PDHβ) has been isolated by screening a leukocyte genomic library with a nick-translated human foreskin fibroblast PDHβ cDNA probe. The 18-kilobase clone was characterized by restriction enzyme analysis, extensive DNA sequencing, and primer-extension analysis. The PDHβ structural gene is composed of 10 exons and 9 introns. All intron–exon splice junctions follow the GT/AG rule. The *Alu* family was found in introns 2 and 8. The 5′ flanking region of the PDHβ gene contains a "CAAT" consensus promoter sequence but no "TATA" sequence. Primer-extension analysis indicated that the PDHβ gene transcription start site is an adenine residue located 132 bases upstream from the initiation codon in exon 1.**

The mammalian mitochondrial pyruvate dehydrogenase (PDH) complex catalyzes an oxidative decarboxylation of pyruvate (1, 2). It is an unusually large multimeric structure composed of multiple copies of four types of polypeptide chains for the three constituent enzyme activities, pyruvate dehydrogenase (lipoamide) [pyruvate:lipoamide 2-oxidoreductase (decarboxylating and acceptor-acetylating), EC 1.2.4.1], dihydrolipoamide acetyltransferase (acetyl-CoA:dihydrolipoamide *S*-acetyltransferase, EC 2.3.1.12), and dihydrolipoamide dehydrogenase (dihydrolipoamide:$NAD^+$ oxidoreductase, EC 1.8.1.4). The dihydrolipoamide acetyltransferase component forms the structural core of the complex to which the PDH and dihydrolipoamide dehydrogenase components bind independently by noncovalent bonds.

Mammalian PDH is composed of two nonidentical subunits, α (PDHα; $M_r$, 41,000) and β (PDHβ; $M_r$, 36,000) and has a tetrameric structure, $\alpha_2\beta_2$ ($M_r$, 153,000) (3, 4). It utilizes thiamin pyrophosphate as cofactor, and the thiamin pyrophosphate-binding site is apparently located in PDHα (5). It appears that PDHβ possesses a dual function, one for constitution of the $\alpha_2\beta_2$ structure and the other for binding PDH to the structural core (dihydrolipoamide acetyltransferase) of the complex.

We have reported (6) the amino acid sequences of human PDHα and PDHβ deduced from the nucleotide sequences of human PDHα and PDHβ cDNAs, which were isolated from a cultured foreskin fibroblast cDNA library. The gene for the human PDHα subunit has been isolated and characterized (7, 19).

A genomic PDH deficiency in infants in Japan has been reported that results in progressive brain and neurological disease with persistent lactic and pyruvic acidemia and growth retardation (8). The majority of these patients showed low levels of PDH activity and some lacked PDHα, PDHβ, or both and showed abnormal bands by immunoblot analysis (9). Molecular cloning and characterization of the human PDH gene should facilitate studies on the structures of the PDHα and PDHβ genes and their expression. This knowledge may also provide essential information about the nature of genetic mutations leading to molecular disease and aid in carrier detection, prenatal diagnosis, and treatment.

This paper describes the molecular cloning of the entire human PDHβ gene and its characterization by restriction enzyme analysis, complete nucleotide sequencing,† and primer-extension analysis of the mRNA.

## MATERIALS AND METHODS

**Screening of Human Genomic DNA Library.** A human leukocyte library in the cloning vector λEMBL3 (Clontech) was screened with the full-length human cultured foreskin fibroblast PDHβ cDNA probe (6) labeled with $^{32}P$ by the nick-translation method. Screening, hybridization, and washing procedures were carried out by standard techniques (10).

**Clone Characterization.** DNA from the positive phage clone was characterized by standard restriction endonuclease mapping and Southern blot analysis (11). DNA sequencing of the whole and specific restriction fragments subcloned into the plasmid vector pUC19 was performed by the dideoxynucleotide chain-termination method (12).

**Primer-Extension Analysis.** Poly(A)$^+$ RNAs from HeLa cells, human liver (autopsy material), and pig liver were isolated by the guanidine isothiocyanate method (13) followed by oligo(dT)-cellulose chromatography (14). An oligonucleotide primer (5′-CTCCGCAACCAAGCCAG-3′) complementary to the nucleotide sequence encoding part of the PDHβ presequence (amino acid residues −26 to −21) was synthesized with a DNA synthesizer (model 380B, Applied Biosystems). Primer extension of PDHβ mRNA transcripts in poly(A)$^+$ RNA preparations were performed as described by Laird *et al.* (15), with the following modification. A 10-ng sample of this primer was added to 5 μg of poly(A)$^+$ RNA in 10 μl of hybridization buffer (50 mM Tris·HCl, pH 8.3/50 mM NaCl/9 mM $MgCl_2$) and heated at 100°C for 1 min. The reaction mixture was hybridized for 10 min at 65°C and then cooled slowly at 37°C over a period of about 2 hr. Additions were made to give the following mixture (final concentrations): 50 mM Tris·HCl, pH 8.3/20 mM NaCl/3.6 mM $MgCl_2$/5 mM dithiothreitol/0.5 mM dCTP/0.5 mM dGTP/0.5 mM dTTP/15 μM [α-$^{32}P$]dATP (1.85 MBq) in a total volume of 25 μl. Rous-associated virus 2 reverse transcriptase (10 units) was added, and the mixture was incubated at 37°C for an additional 15 min. After addition of 1.2 μl of 10 mM dATP chase solution, incubation was continued at 37°C for an additional 15 min. The reaction was terminated by addition of

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: PDH, pyruvate dehydrogenase; PDHα, PDH α subunit; PDHβ, PDH β subunit.

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the DDBJ and EMBL/GenBank data bases (accession no. D90086).

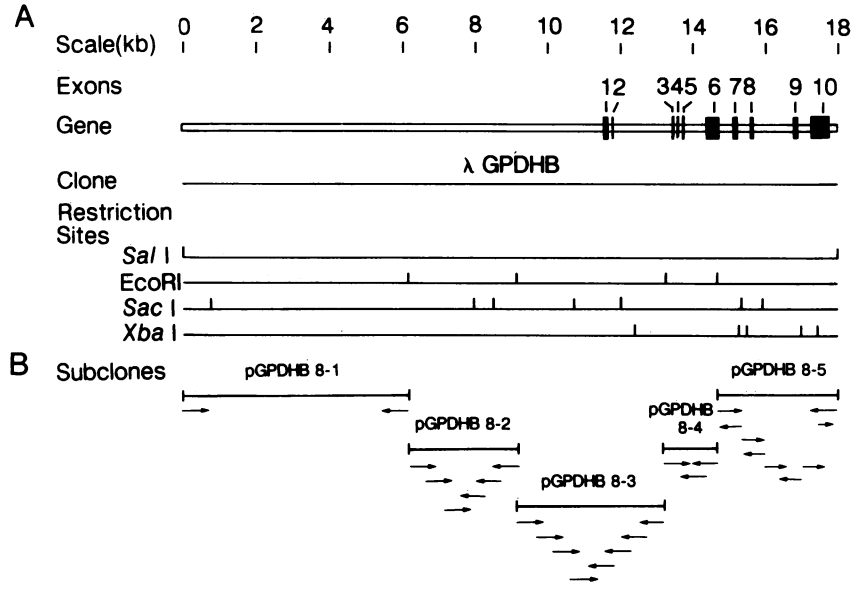


FIG. 1. Physical map and sequencing strategy for human PDHβ gene. (*A*) The structure of the gene is schematically represented by the bar at the top of the diagram. Exons are represented by solid boxes (numbered 1–10). The direction of transcription is from left to right. The scale (every 2 kb) is drawn above the gene. Recognition sites for restriction endonucleases *Sal* I, *Eco*RI, *Sac* I, and *Xba* I are given below the phage clone (λGPDHB8) represented by solid line. (*B*) The horizontal arrows indicate the direction and extent of individual sequencing experiments. *Eco*RI-digested fragments subcloned into pUC19 were fragmented by digestion with *Sac* I and *Xba* I (pGPDHB8-5) and by bidirectional deletion with BAL-31 nuclease (pGPDH-B8-2 to -8-4). The fragments were subcloned and sequenced.

EDTA to 20 mM. The extracted products were analyzed by electrophoresis in a 6% polyacrylamide/7 M urea gel.

## RESULTS AND DISCUSSION

**Isolation and Characterization of Human PDHβ Genomic Clone.** A human leukocyte genomic library in the cloning vector λEMBL3 was screened with the human $^{32}$P-labeled PDHβ cDNA (6) as probe. One positive bacteriophage was isolated from $1 \times 10^6$ plaque-forming units of recombinant λEMBL3 phage after tertiary screening. The DNA insert from the positive clone, designated as λGPDHB8, was purified and subcloned into the plasmid vector pUC19. A restriction endonuclease map of the 18-kilobase (kb) DNA insert from the positive clone that encompassed the human PDHβ gene is shown in Fig. 1*A*. The five *Eco*RI restriction fragments of the DNA insert were further examined by Southern blot analysis with the full-length PDHβ cDNA

Table 1. Exon–intron organization of the human PDHβ gene

| Exon | | Sequence at exon–intron junction | | Intron | |
|---|---|---|---|---|---|
| No. | Size, bp | 5′ splice donor | 3′ splice acceptor | No. | Size, bp |
| 1 | 174 | CGGGAG/gtaggc———— | ————tccctgccag/GTCTCC | 1 | 82 |
| | | ArgGlu | ValSer | | |
| | | 14 | 15 | | |
| 2 | 54 | CTGCAG/gtaaca———— | ————cctgtgacag/GTGACA | 2 | 1645 |
| | | LeuGln | ValThr | | |
| | | 32 | 33 | | |
| 3 | 108 | TACAAG/gtaact———— | ————ttaattgtag/GTTAGT | 3 | 83 |
| | | TyrLys | ValSer | | |
| | | 68 | 69 | | |
| 4 | 63 | TCAGAG/gtaagc———— | ————tactttatag/ATGGGC | 4 | 101 |
| | | SerGlu | MetGly | | |
| | | 89 | 90 | | |
| 5 | 36 | GCTATG/gtatgt———— | ————gattttacag/GCTGGG | 5 | 653 |
| | | AlaMet | AlaGly | | |
| | | 101 | 102 | | |
| 6 | 286 | AATCCAG/gtcagc———— | ————tgtttttcag/TGGTGGTG | 6 | 419 |
| | | AsnProV | alValVal | | |
| | | 197 | 198 | | |
| 7 | 111 | AGGCAAG/gtaaga———— | ————ctgttttcag/GAACACAT | 7 | 326 |
| | | ArgGlnG | lyThrHis | | |
| | | 234 | 235 | | |
| 8 | 92 | TGTGAG/gtgagt———— | ————acctgtttag/GTGATA | 8 | 1094 |
| | | CysGlu | ValIle | | |
| | | 264 | 265 | | |
| 9 | 142 | ATGGAAG/gtattt———— | ————cttctcctag/GTCCTGCG | 9 | 293 |
| | | MetGluG | lyProAla | | |
| | | 312 | 313 | | |
| 10 | 550 | ATTTAG and 3′ untranslated region | | | |
| | | Ilestop | | | |
| | | 359 | | | |

Adjacent exon (uppercase letters) and intron (lowercase letters) sequences are given for each junction. The numbers of amino acids bordering the splice junctions indicate the positions of the corresponding amino acids in the PDHβ gene, deduced from the correct PDHβ cDNA.

Table 2. Corrections in the previously published human PDHβ cDNA sequence

| Nucleotide position(s) | Change of nucleotide sequence | Effect on amino acid sequence |
|---|---|---|
| 24 | Deletion of T | Frame shift* |
| 40 | Insertion of G | Frame shift* |
| 90, 91 | CG to GC | Val to Leu† |
| 438 | G to A | No effect |
| 641 | Insertion of C | Frame shift‡ |
| 663 | Deletion of G | Frame shift‡ |
| 928 | Deletion of A | Frame shift§ |
| 936 | Insertion of G | Frame shift§ |

The previous human PDHβ cDNA sequence is in ref. 6.
*Five residues.
†Amino-terminal.
‡Eight residues.
§Three residues.

probe. The cDNA probe hybridized to three *Eco*RI fragments of 4.0, 1.5, and 3.5 kb, suggesting the localization of the major portion of the PDHβ gene in these fragments. The five *Eco*RI fragments were subcloned into pUC19 and designated as pGPDHB8-1 (6.0 kb), -8-2 (3.0 kb), -8-3 (4.0 kb), -8-4 (1.5 kb), and -8-5 (3.5 kb), as shown in Fig. 1*B*.

**Nucleotide Sequence and Organization of the Human PDHβ Gene.** To facilitate detailed restriction endonuclease mapping and nucleotide sequence analysis, fragments of the DNA insert were prepared by restriction endonuclease digestion or BAL-31 nuclease deletion (16) and subcloned into pUC19. Fragments of the two DNA inserts from pGPDHB8-3 and -8-4 were prepared by BAL-31 nuclease deletion and were subcloned. Fragments of pGPDHB8-5 were prepared by *Sac* I or *Xba* I restriction endonuclease digestion and were subcloned. These fragments and the five *Eco*RI fragments were subjected to extensive sequence analysis after strand denaturation, using the strategy outlined in Fig. 1*B*. Nucleotide sequences were determined by the dideoxynucleotide chain-termination method. The detailed structure of the PDHβ gene is shown in Fig. 1*A* and Table 1. The cloned DNA spans 18 kb. However, the major coding region apparently is localized to the 8872-base-pair (bp) segment starting with the 5′ end of pGPDHB8-3 (second *Eco*RI site) and terminating with the 3′ end of pGPDHB8-5. The gene is composed of 10 exons and 9 introns within its coding region. As shown in Table 1, the 10 exons range in size from 36 bp (exon 5) to 552 bp (exon 10), and the 9 introns ranged from 82 bp (intron 1) to 1645 bp

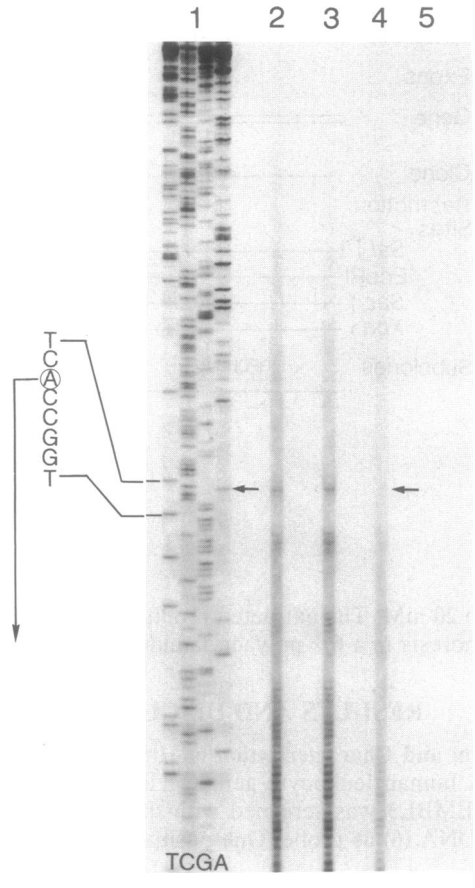


FIG. 3. Primer-extension analysis with human PDHβ mRNA. Sequencing ladders were made in parallel, using the same oligonucleotide primer and pGPDH8-3 template (see Fig. 1*B*) that contained exons 1 and 2 and the 5′ flanking region of the gene. Lanes: 1, pGPDHB8-3 DNA (2 μg) sequencing ladder electrophoresed in parallel; 2, HeLa cells poly(A)$^+$ RNA (5 μg); 3, human liver (5 μg); 4, pig liver poly(A)$^+$ RNAs (5 μg); 5, control tRNA (5 μg). The major transcription initiation site is indicated by an arrow. The nucleotide of the transcription start site is circled.

(intron 2). Intron–exon boundaries were established by comparison of nucleotide sequences of the PDHβ cDNA and the gene. All intron–exon splice junctions follow the GT/AG rule (17). Exon 1 contained transcription and translation initiation sites and exon 2 contained the sequence encoding the amino-

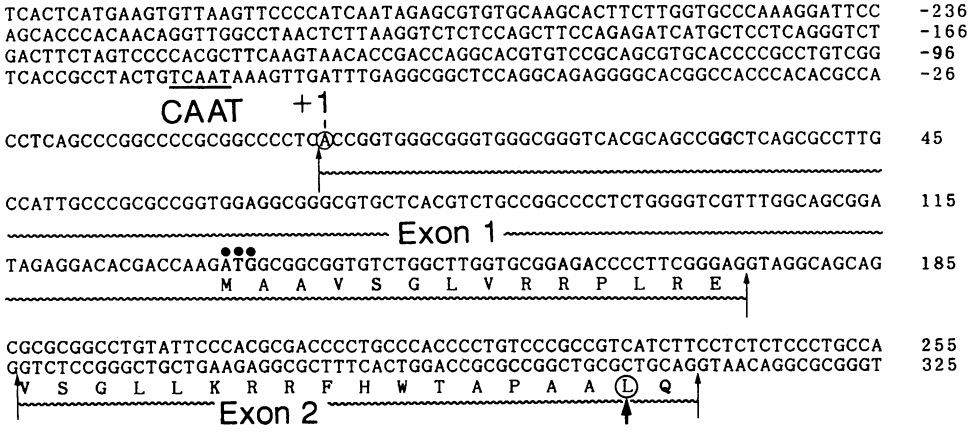


FIG. 2. Nucleotide sequence of the 5′ end of the major coding region of the human PDHβ gene. Nucleotides are numbered from the transcription initiation site with an adenine residue at position +1. The translation initiation site codon (ATG) is marked with solid circles. Exons are underlined with wavy lines and numbered. Amino acid sequence (standard one-letter symbols) is given below the coding sequence. The amino-terminal amino acid, leucine in exon 2, is marked with a circle and an arrow. The sequence underlined with a solid line is a putative CAAT box.

terminal region of the mature PDHβ. Two *Alu* repeats were found in the human PDHβ gene. One was located in intron 2 (294 bp long) and the other was in intron 8 (289 bp long).

There are some discrepancies between our published sequence of PDHβ cDNA (6) and the sequences of the 10 exons. Discrepancies ranging from single-base alterations (such as deletion, insertion, and substitution) to transposition resulting in reading-frame shifts of up to eight amino acids are summarized in Table 2. We believe the gene sequence is correct. The resulting changes in deduced amino acid sequence of human PDHβ are consistent with the results of amino acid sequence analysis of pig heart PDHβ.

**Identification of Human PDHβ Gene Transcription Start Site.** Analysis of the nucleotide sequence of the human PDHβ gene for conserved sequences of functional significance revealed a number of features in common with other mammalian genes. The 5′ flanking region of the PDHβ gene contains a typical TCAAT consensus promoter sequence, as shown in Fig. 2. We have reported (6) that the human PDHβ cDNA contains sequences that are compatible with HeLa cell PDHβ mRNA determined by RNA blot analysis. The 5′ end of the mRNA was determined by the primer-extension method, using a synthetic oligonucleotide primer (17-mer) complementary to bases in exon 1 (Fig. 2). The primer apparently hybridized to poly(A)$^+$ RNAs from HeLa cells, human liver, and pig liver and underwent extension by reverse transcriptase. As shown in Fig. 3, transcription was initiated at an adenine residue, which is 132 bp upstream from the ATG initiating translation codon in exon 1 (Fig. 2). This was preceded by a CAAT box at nucleotide positions −82 to −78. This clone will be a useful tool for further studies on regulation of gene expression.

**Clinical Application.** A large number of patients with genetic deficiency in the PDH component of the PDH complex have been reported (8, 9). However, except in one case (18), the nature of molecular defects in these patients remains to be elucidated. Useful restriction fragment length polymorphisms have not been detected in the PDHα and -β genes (9). Knowledge of the entire nucleotide sequence of the normal human PDHβ gene should facilitate analyses of genetic defects affecting this gene.

We thank Mr. M. Fujinaka for the generous gift of the human leukocyte genomic library (Clontech), Ms. S. Nakao for helpful assistance, and Ms. M. Shiinoki for skillful assistance in preparation of the manuscript. In addition, we would like to thank Dr. Lester J. Reed for helpful comments on the manuscript. This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas and for Scientific Research from the Ministry of Education, Science and Culture of Japan, and by a grant from the Vitamin B Research Committee.

1. Reed, L. J. (1974) *Acc. Chem. Res.* **7**, 40–46.
2. Koike, M. & Koike, K. (1976) *Adv. Biophys.* **9**, 187–227.
3. Hayakawa, T., Kanzaki, T., Kitamura, T., Fukuyoshi, Y., Sakurai, Y., Koike, K., Suematsu, T. & Koike, M. (1969) *J. Biol. Chem.* **244**, 3660–3670.
4. Hamada, M., Hiraoka, T., Koike, K., Ogasawara, K., Kanzaki, T. & Koike, M. (1976) *J. Biochem.* **79**, 1273–1285.
5. Hawkins, C. F., Borges, A. & Perham, R. N. (1989) *FEBS Lett.* **255**, 77–82.
6. Koike, K., Ohta, S., Urata, Y., Kagawa, Y. & Koike, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 41–45.
7. Maragos, C., Hutchison, W. M., Hayasaka, K., Brown, G. K. & Dahl, H.-H. M. (1989) *J. Biol. Chem.* **264**, 12294–12298.
8. Koike, K. (1988) in *Thiamin Pyrophosphate Biochemistry: Enzymatic Studies of the Genetic Defect of Pyruvate Dehydrogenase in Chronic Pyruvic and Lactic Acidemia*, eds., Schellenberger, A. & Schowen, R. L. (CRC Press, Boca Raton, FL), Vol. 2, pp. 105–113.
9. Koike, K. & Urata, Y. (1989) *Ann. N.Y. Acad. Sci.* **573**, 450–452.
10. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), 2nd Ed.
11. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
13. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
14. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
15. Laird, J. E., Jack, L., Hall, L., Boulton, A. P., Parker, D. & Craig, R. K. (1988) *Biochem. J.* **254**, 85–94.
16. Davis, L. G., Dibner, M. D. & Battey, J. F. (1986) in *Basic Method in Molecular Biology: Preparation of Insert for M13 Cloning by Successive Bal31 Exonuclease Deletion* (Elsevier, New York), pp. 244–288.
17. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
18. Endo, H., Hasegawa, K., Narisawa, K., Tada, K., Kagawa, Y. & Ohta, S. (1989) *Am. J. Hum. Genet.* **44**, 358–364.
19. Koike, K., Urata, Y., Matsuo, S. & Koike, M. (1990) *Gene*, in press.