

Coding sequence of the precursor of the β subunit of rat propionyl-CoA carboxylase

(mitochondrial enzyme/leader peptide structure/propionic acidemia)

JAN P. KRAUS*, FRANK FIRGAIRA*, JIŘÍ NOVOTNÝ†, FRANTIŠEK KALOUSEK*, KENNETH R. WILLIAMS‡, CYNTHIA WILLIAMSON*, TOSHIHIRO OHURA*, AND LEON E. ROSENBERG*

Departments of *Human Genetics and †Molecular Biophysics and Biochemistry, Yale University School of Medicine, 333 Cedar Street, P.O. Box 3333, New Haven, CT 06510; and ‡Massachusetts General Hospital and Harvard Medical School, Molecular and Cellular Research Laboratory, Boston, MA 02114

Contributed by Leon E. Rosenberg, July 14, 1986

ABSTRACT A cDNA encoding the cytoplasmic precursor of the β subunit of the mitochondrial enzyme propionyl-CoA carboxylase (EC 6.4.1.3) was cloned and sequenced. The DNA sequence of 2070 nucleotides is almost identical in size to the major hybridizing mRNA from rat liver (2000 ± 50 nucleotides), suggesting that the cloned DNA represents nearly all of the mRNA sequence. A polypeptide expressed *in vitro* from an mRNA transcript of this cDNA is indistinguishable in size from the β subunit precursor (58,500 Da). An open reading frame of 1623 nucleotides, flanked by stop codons, encodes a polypeptide of 541 amino acids; the predicted amino acid sequence was confirmed as that of the β subunit of propionyl-CoA carboxylase by matching it to the amino acid sequences of five peptides derived from pure mature rat enzyme. Although the exact length of the cleavable, NH_2 -terminal leader peptide has not been determined because the NH_2 -terminal residue of the mature subunit is blocked, the leader is most likely 40–42 amino acids in length and is highly positively charged. Computer-aided analysis of secondary structure suggests that the leader peptide consists of two α -helical segments, with the two most NH_2 -terminal arginine residues occupying opposite sites of the first helix; this helix has no apparent hydrophobic moment.

Propionyl-CoA carboxylase [PCCase; propanoyl-CoA: carbon-dioxide ligase (ADP-forming), EC 6.4.1.3], a biotin-dependent mitochondrial enzyme, is a key enzyme in the catabolic pathway for odd-chain fatty acids, isoleucine, threonine, methionine, and valine. In humans, propionic acidemia due to recessively inherited deficiency of PCCase activity often causes life-threatening ketosis and acidosis (1). The native enzyme is probably a dodecamer (2) composed of six biotin-containing α subunits ($M_r = 70,000$ – $72,000$) and six β subunits ($M_r = 54,000$ – $56,000$) (3). The α and β subunits are each synthesized initially as larger precursors on cytoplasmic polyribosomes. These precursors, containing NH_2 -terminal leader sequences, are subsequently transported across both mitochondrial membranes, cleaved to their mature size, and assembled (4). To facilitate investigation of PCCase biogenesis and the nature of mutations in the various complementation groups of inherited PCCase deficiency (5), we have isolated cDNA clones for the β subunit (β PCCase) (6). In this communication we present the entire amino acid sequence of the β PCCase precursor as deduced from the cDNA sequence. This primary structure is supported by amino acid sequences of five peptides isolated from homogeneous rat liver PCCase.

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.

EXPERIMENTAL PROCEDURES

cDNA Cloning. We have previously described the isolation and identification of β PCCase cDNA clones from a cDNA library prepared from a highly enriched, immunopurified β PCCase mRNA (6, 7). A restriction map of the clones has also been published (6).

***In Vitro* Transcription and Translation.** Plasmid DNA containing the T7 promoter joined with the β PCCase coding sequence was prepared for transcription by restriction cleavage at a unique *Pvu* II site downstream from the coding sequence. The product was extracted with phenol and precipitated with ethanol, and 1 μ g of linearized DNA was then used in a 50- μ l transcription reaction mixture containing nucleoside triphosphates at 0.5 mM each, 10 μ Ci (1 Ci = 37 GBq) of [α - 32 P]GTP, 40 mM Tris-HCl at pH 7.5, 6 mM MgCl_2 , 2 mM spermidine, 10 mM NaCl, 10 mM dithiothreitol, 70 units of RNasin (Promega Biotec, Madison, WI), and 125 units of T7 RNA polymerase. The reaction was carried out for 90 min at 40°C. The products were extracted with phenol and precipitated with ethanol from 2 M ammonium acetate to remove the unincorporated nucleotides. Cell-free translation in rabbit reticulocyte lysate followed by immunoprecipitation was carried out as previously described (7).

Blot Hybridization Analysis. Total RNA from rat and human liver or from human skin fibroblasts were prepared as described (8). Polyadenylated RNA was isolated by oligo(dT)-cellulose chromatography according to the supplier's instructions (Collaborative Research, Waltham, MA). Blot hybridization analysis was performed after electrophoresis of the RNAs through a 1% agarose gel containing formaldehyde (9).

DNA Sequencing. DNA was sequenced according to Maxam and Gilbert (10). Restriction endonuclease fragments were sequenced on both strands by radiolabeling at their 5' termini using T4 polynucleotide kinase (Bethesda Research Laboratories) and [γ - 32 P]ATP (Amersham) or at their 3' termini by using the Klenow fragment of *Escherichia coli* DNA polymerase I (Boehringer Mannheim) and the appropriate [α - 32 P]dNTP (Amersham).

Succinylation, Isolation, and Sequencing of Arginine Peptides from PCCase. Rat liver PCCase was purified to homogeneity by using affinity chromatography on avidin (11). To obtain arginine peptides, 30 nmol of PCCase was dialyzed against 5 M guanidine hydrochloride (pH 8.5) for 36 hr. The sample (20 ml) was then succinylated (12) with 150 mg of succinic anhydride while the pH was maintained at 8.5. When succinylation was complete (5 min), solid $\text{NH}_2\text{OH}\cdot\text{HCl}$ was added to a final concentration of 1 M, 5 M NaOH was used to adjust the pH to 10, and the mixture was stirred for 1 hr.

Abbreviations: PCCase, propionyl-CoA carboxylase; β PCCase, β subunit of PCCase; bp, base pair(s).

After dialysis against 0.25 mM NaHCO₃ and drying, the succinylated PCCase was dissolved in 1 ml of 0.2 M NaHCO₃ and digested with tosylphenylalanyl chloromethyl ketone (TPCK)-treated trypsin at 1:20 (wt/wt) ratio at 37°C for 8 hr. The resulting peptides were separated on a Vydac C₄ reverse-phase column (0.45 × 25 cm) equilibrated with 0.05% trifluoroacetic acid and then eluted as described previously with increasing concentrations of acetonitrile (13). Aliquots were taken from the peak fractions for hydrolysis in 6 M HCl (115°C, 16 hr), followed by analysis on a Beckman 121 M amino acid analyzer. Several selected peptides were either coupled to aminopolystyrene and then subjected to solid-phase sequencing on a Sequemat mini 15 sequenator or were dried under reduced pressure and then loaded onto an Applied Biosystems gas phase sequencer in 0.1 ml of trifluoroacetic acid. The resulting phenylthiohydantoin amino acids were identified by reverse-phase HPLC (13).

RESULTS AND DISCUSSION

β PCCase cDNA Clones. We had previously isolated and identified two overlapping β PCCase cDNA clones from a rat liver cDNA library highly enriched for β PCCase sequences (6). The unique portions [2070 base pairs (bp)] of these two clones were joined and inserted into a plasmid, pGEM2 (Promega Biotec, Madison, WI), bearing a T7 RNA polymerase promoter. The β PCCase cDNA was transcribed *in vitro*, and the transcript was translated by using a reticulocyte lysate-programmed system. Fig. 1 shows that the translated polypeptide (58.5 kDa) is indistinguishable in size from the β PCCase precursor immunoprecipitated from a translation mixture of total rat liver mRNA.

RNA Analysis. Blot hybridization analysis of rat and human liver and human skin fibroblast mRNA with the β PCCase cDNA probe just described yielded a major hybridizing species of 2000 ± 50 nucleotides in both rat and human mRNA (Fig. 2). In addition, minor bands were observed at 2600 and 4500 nucleotides in rat mRNA and at 450 and 4500 nucleotides in human mRNA. The 2600-nucleotide rat transcript may represent a minor mRNA species elongated at the 3' end. The existence of such an elongated species has been shown to account for the two mRNAs found for the α chain

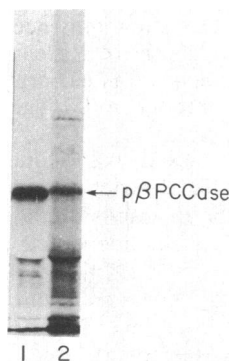


FIG. 1. NaDodSO₄/polyacrylamide gel analysis of β PCCase polypeptides translated from transcribed β PCCase cDNA and from total rat liver mRNA. A plasmid containing the T7 RNA polymerase promoter joined with the β PCCase cDNA sequence was transcribed *in vitro* and the product was used to direct a reticulocyte lysate-programmed translation mixture containing [³⁵S]methionine. After translation, an aliquot (1 μ l) of the mixture was applied to the gel (lane 1). The β PCCase band in lane 1 is, in fact, a doublet produced by initiation at both the initiator methionine codon and at the first internal methionine (residue 10). Total rat liver mRNA was translated in 100 μ l of translation mixture and the precursor of β PCCase ($p\beta$ PCCase) was immunoprecipitated with anti- β PCCase antiserum (lane 2). The radioactive bands were visualized by fluorography for 10 hr.

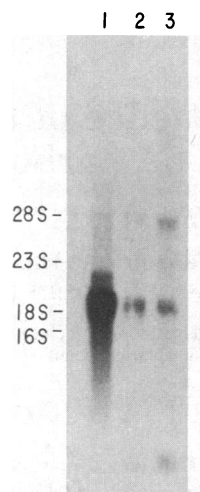


FIG. 2. Blot hybridization analysis of β PCCase mRNA from rat liver, human liver, and human skin fibroblasts. mRNA samples were denatured with formaldehyde and formamide and electrophoresed in a 1% agarose gel (9). The separated mRNAs were transferred to nitrocellulose, and the blot was probed with radiolabeled rat β PCCase cDNA. Lane 1, rat liver mRNA, 5 μ g; lane 2, human liver mRNA, 5 μ g; lane 3, human skin fibroblast mRNA, 10 μ g. The sizes of ribosomal RNA markers are as follows: 16S, 1541 nucleotides; 18S, 1740 nucleotides; 23S, 3000 nucleotides; and 28S, 4850 nucleotides.

of β -hexosaminidase (14). The 4500-nucleotide species probably represent nuclear precursors. The nature of the 450-nucleotide species, particularly prominent in human fibroblast mRNA, is unknown.

Nucleotide Sequence. The cloned β PCCase cDNA sequence is 2070 bp long (Fig. 3). Comparison of the estimated size of the β PCCase mRNA with that of the cloned cDNA suggests that very little mRNA information is missing from the cDNA. The open reading frame of 1623 bp is flanked by stop codons at positions 1 and 1630. The putative translation initiation codon beginning at position 7 within the sequence AAAAAAUGG differs somewhat from the optimal context for initiation, CCACCAUGG, derived from surveys of eukaryotic mRNAs (15). However, the two most highly conserved residues, an A at position -3 and a G at position +4, are present in the β PCCase mRNA. The 3' untranslated region of 441 bp does not contain a poly(A) tail. Codon usage in the β PCCase mRNA is summarized in Table 1. Of all the possible 61 codons, only 57 are utilized in the sequence. CUA (Leu), GUA (Val), and UCG (Ser) are not used. UUA is used only once in 39 leucine codons, and all 15 Gln residues are coded for by CAG rather than CAA.

Peptide Sequence Analysis of Mature β PCCase. The homogeneous rat PCCase was modified at lysine residues by succinylation, and the modified protein was cleaved with trypsin. Thus, the protein was cleaved only at arginine residues. The resulting peptides were separated by HPLC and their sequences were determined on a solid-phase sequenator or gas-phase sequencer. Out of eight sequenced peptides, five matched the β PCCase DNA sequence, while the other three were later found to match the human α PCCase chain (16). The NH₂ termini of both the α and the β PCCase mature subunits were found to be blocked, as evidenced by the lack of any identifiable sequence when 900 pmol of the succinylated subunits was loaded onto the gas-phase sequencer. With the exception of a steadily decreasing background of glycine, cycles 2–10 contained less than 10 pmol of any individual phenylthiohydantoin derivative.

Amino Acid Sequence of β PCCase and a Predicted Secondary Structure of Its Leader. The open reading frame of 1623 nucleotides in the cDNA sequence (Fig. 3) encodes a poly-

Table 1. Codon usage in β PCCase mRNA

AA	Codon	No.	AA	Codon	No.	AA	Codon	No.	AA	Codon	No.
F	UUU	13	S	UCU	4	Y	UAU	9	C	UGU	4
F	UUC	15	S	UCC	8	Y	UAC	3	C	UGC	7
L	UUA	1	S	UCA	6	*	UAA	0	*	UGA	1
L	UUG	4	S	UCG	0	*	UAG	0	W	UGG	2
L	CUU	3	P	CCU	12	H	CAU	7	R	CGU	4
L	CUC	7	P	CCC	7	H	CAC	6	R	CGC	5
L	CUA	0	P	CCA	3	Q	CAA	0	R	CGA	4
L	CUG	24	P	CCG	3	Q	CAG	15	R	CGG	6
I	AUU	10	T	ACU	6	N	AAU	11	S	AGU	5
I	AUC	24	T	ACC	8	N	AAC	11	S	AGC	9
I	AUA	4	T	ACA	10	K	AAA	6	R	AGA	8
M	AUG	12	T	ACG	3	K	AAG	17	R	AGG	7
V	GUU	10	A	GCU	14	D	GAU	8	G	GGU	9
V	GUC	17	A	GCC	25	D	GAC	22	G	GGC	16
V	GUA	0	A	GCA	11	E	GAA	9	G	GGA	15
V	GUG	20	A	GCG	6	E	GAG	18	G	GGG	8

AA, amino acid residue (standard one-letter code); No., number of times the given codon is used in β PCCase; *, stop codon.

features that are shared with the other dozen known mitochondrial leaders that target cytoplasmic precursors to the mitochondrial matrix, suggesting that a common structural motif is responsible for the targeting role of these sequences. Although there is little conservation of amino acid sequence among the leader segments, recent analyses have suggested

the presence of α -helical domains in these leader peptides (19, 20). von Heijne, in particular, has emphasized the possible importance of the amphipathic character of these α -helices (20). Profiles of secondary structure propensity and of hydrophobicity (21) of the β PCCase leader sequence (Fig. 4 *Upper*) strongly predict three to four turns of an α -helix

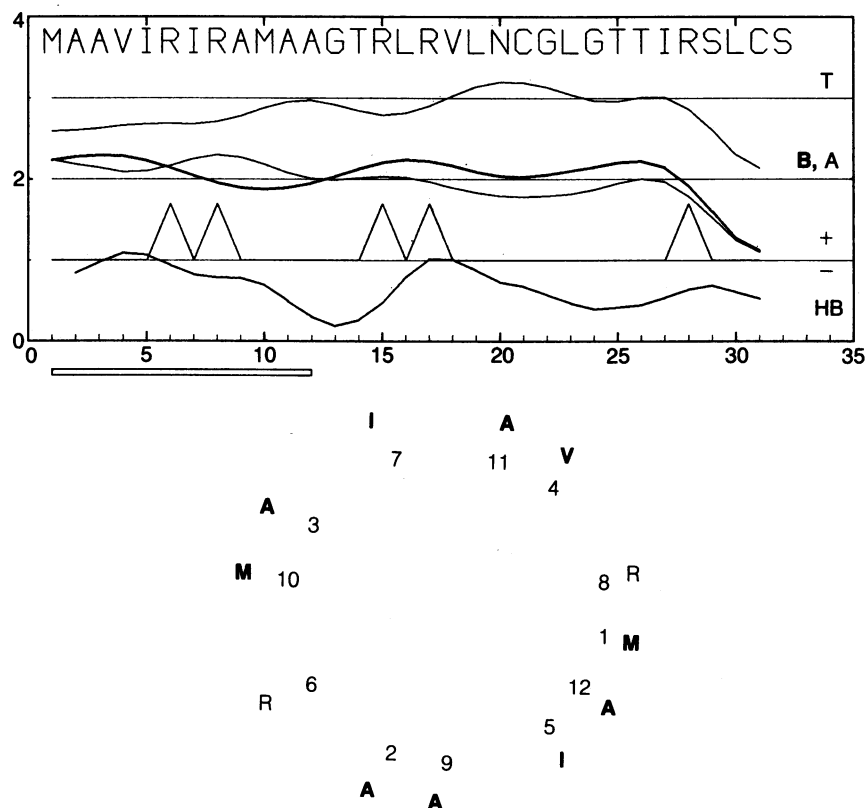


FIG. 4. Amino acid sequence profiles of the first 32 amino acid residues of the leader peptide of the β PCCase precursor and a helical-wheel analysis. (*Upper*) From top to bottom, the amino acid sequence (one-letter code) used to compute the profiles; reverse-turn propensity (T); α -helix (thin line) and β -sheet (heavy line) propensities (B, A); profile of charged residues (+ or -); and hydrophobicity (HB). The most NH_2 -terminal α -helical region predicted is represented by the horizontal box below the panel. (*Lower*) Helical-wheel analysis of amino acid residues 1-12. Residues are numbered consecutively from the NH_2 terminus of the helix. The residues are plotted circularly with a successive angular displacement of 100° . Such a plot corresponds to angular periodicity of residues in the ideal α -helix with 3.7 residues per helical turn (22). Note the perfect distribution of hydrophobic residues (boldface letters) in the helix, indicating that the whole helix with the exception of the two arginines is hydrophobic.

between positions 1 and 12 and also give a weaker, more ambiguous, prediction of α -helix for residues 17 to 28. The strong prediction for an NH₂-terminal α -helical segment is supported by use of the helical wheel (Fig. 4 Lower). Although the NH₂-terminal α -helix is exceptionally hydrophobic, its overall hydrophobic moment (23) is virtually nil because its two pairs of methionine and isoleucine residues, as well as its two charged arginine side chains, occupy opposite faces of the helix. Unlike the NH₂-terminal segment, the second α -helix (residues 17–28), if it exists, is likely to be amphipathic. Clearly, more experimental and theoretical work is needed to define more precisely the putative role of amphipathicity in mitochondrial transport.

The authors thank the following people: Anne-Marie Lamhonwah and Roy Gravel for sharing with us the unpublished partial cDNA sequences of both subunits of human PCCase; Peter Hodges for many helpful suggestions; Wayne Fenton for his help with computer-assisted analysis of nucleotide sequences; Manju Swaroop and Kathy Stone for their expert technical help with DNA and protein sequencing, respectively; and Connie Woznick for secretarial assistance. This work was supported by Grants AM09527, AM12579, and GM31359 from the National Institutes of Health.

1. Rosenberg, L. E. (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), 5th Ed., pp. 474–497.
2. Haase, F. C., Beegen, H. & Allen, S. H. G. (1984) *Eur. J. Biochem.* **140**, 147–151.
3. Kalousek, F., Darigo, M. D. & Rosenberg, L. E. (1980) *J. Biol. Chem.* **245**, 60–65.
4. Kraus, J. P., Kalousek, F. & Rosenberg, L. E. (1983) *J. Biol. Chem.* **258**, 7245–7248.
5. Gravel, R. A., Lam, F. K., Scully, K. J. & Hsia, Y. E. (1977) *Am. J. Hum. Genet.* **29**, 378–388.
6. Kraus, J. P., Williamson, C. L., Firgaira, F. A., Yang-Feng, T. L., Munke, M., Francke, U. & Rosenberg, L. E. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2047–2051.
7. Kraus, J. P. & Rosenberg, L. E. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4015–4019.
8. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1970) *Biochemistry* **18**, 5294–5299.
9. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 202–203.
10. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
11. Gravel, R. A., Lam, K. F., Mahuran, D. & Kronis, A. (1980) *Arch. Biochem. Biophys.* **201**, 669–673.
12. Williams, K. R., LoPresti, M. D., Setoguchi, M. & Konigsberg, W. H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4614–4617.
13. Williams, K. R., Hemmings, H. C., LoPresti, M. B., Konigsberg, W. H. & Greengard, P. (1986) *J. Biol. Chem.* **261**, 1890–1903.
14. Myerowitz, R., Piekarz, R., Neufeld, E. F., Shows, T. B. & Suzuki, K. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7830–7834.
15. Kozak, M. (1986) *Cell* **44**, 283–292.
16. Lamhonwah, A.-M., Barankiewicz, T. J., Willard, H. F., Mahuran, D. J., Quan, F. & Gravel, R. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4864–4868.
17. O'Donnell, I. J. & Inglis, A. S. (1974) *Aust. J. Biol. Sci.* **27**, 369–382.
18. Mueckler, M. M. & Pitot, H. C. (1985) *J. Biol. Chem.* **260**, 12993–12997.
19. Horwich, A. L., Kalousek, F., Fenton, W. A., Pollock, R. A. & Rosenberg, L. E. (1986) *Cell* **44**, 451–459.
20. von Heijne, G. (1986) *EMBO J.* **5**, 1335–1342.
21. Novotný, J. & Auffray, C. (1984) *Nucleic Acids Res.* **12**, 243–255.
22. Schiffer, M. & Edmundson, A. B. (1968) *Biophys. J.* **8**, 29–39.
23. Eisenberg, D., Weiss, R. M. & Terwilliger, T. S. (1982) *Nature (London)* **299**, 371–374.