Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 3024–3028, April 1994
Medical Sciences

Molecular cloning, sequencing, and functional expression of a cDNA encoding human coproporphyrinogen oxidase

PAVEL MARTASEK*, JEAN MICHEL CAMADRO†, MARIE-HÉLÈNE DELFAU-LARUE*, JEAN-BAPTISTE DUMAS‡, JEAN JACQUES MONTAGNE‡§, HUBERT DE VERNEUIL*, PIERRE LABBE†, AND BERNARD GRANDCHAMP*

*Laboratoire de Génétique Moléculaire, Institut National de la Santé et de la Recherche Médicale CFJ 8904 Faculté de Médecine Xavier Bichat, Université Paris 7, 75018 Paris, France; †Laboratoire de Biochimie des Porphyrines and §Laboratoire de Bioactivation des Peptides, Institut Jacques Monod, Université Paris 7, 75251 Paris Cedex 5, France; and ‡Laboratoire de Neurobiologie Cellulaire et Moléculaire, Centre National de la Recherche Scientifique, 91198, Gif sur Yvette Cedex, France

Communicated by Rudi Schmid, December 6, 1993

ABSTRACT **Coproporphyrinogen oxidase (EC 1.3.3.3) catalyzes the sixth step in the heme biosynthetic pathway, the oxidation of coproporphyrinogen III to protoporphyrinogen IX. The activity of this enzyme is deficient in the disease hereditary coproporphyria. The sequence of the cDNA and predicted amino acid sequence of the human coproporphyrinogen oxidase are presented. The human protein sequence contains a region completely homologous to that we obtained by sequencing an 11-amino acid peptide fragment from purified murine liver coproporphyrinogen oxidase. Results of Southern blotting were consistent with the presence of a single human coproporphyrinogen oxidase gene, and Northern blotting demonstrated one transcript of similar size in erythroid and nonerythroid cell lines. Expression of the cDNA coding for the putative mature human coproporphyrinogen oxidase in *Escherichia coli* resulted in a 17-fold increase in coproporphyrinogen activity over endogenous activity.**

Coproporphyrinogen oxidase (CPX; EC 1.3.3.3) is a soluble mitochondrial protein that is localized in the intermembrane space within mammalian cells (1, 2) and catalyzes the sixth step in heme biosynthesis, the conversion of the two propionate groups at positions 2 and 4 of coproporphyrinogen III to two vinyl groups, thus producing protoporphyrinogen IX (3). CPX has been characterized and purified to homogeneity from a variety of sources, including the yeast *Saccharomyces cerevisiae* (4), mouse liver (5), and bovine liver (6). The CPX genes from *S. cerevisiae* (7), *Salmonella typhimurium* (8), and soybean (9) have been cloned and sequenced. Recently, the sequence and expression of a mouse CPX cDNA have been published (10).

Heme carries out many vital functions in mammalian physiology, in that it serves as the prosthetic group of hemoproteins which mediate oxygen transport and storage, generation of cellular energy, formation of certain steroid hormones, some reduction reactions, and the detoxification of many drugs. In addition, heme is involved in the regulation of protein synthesis and modifies cell development (11, 12). The induction of heme biosynthesis is an early event in erythroid differentiation, and the control of CPX activity may be important in the regulation of this process (13).

The CPX activity is reduced to 50% of normal in tissues of patients with hereditary coproporphyria, an autosomal dominant disease characterized clinically by acute attacks of paresis, abdominal pain, and psychiatric disturbances. Biochemically, the disease is characterized by increased excretion of coproporphyrin III in feces and urine (14). Two different phenotypes of unrelated homozygous patients have been described, both with a reduction of CPX activity to about 10% of normal, but the molecular basis for the different clinical and biochemical characteristics of these two forms is unknown (15, 16).

To facilitate the characterization of the structure of CPX, to determine the molecular defects in hereditary coproporphyria, and to study the role of CPX in the regulation of the heme biosynthetic pathway, molecular cloning of a cDNA for human CPX was an important step.

To achieve this goal, murine CPX was purified; the amino-terminus of the intact protein and an internal peptide fragment were sequenced; and degenerate oligonucleotides were designed to amplify a specific portion of the murine cDNA by using reverse transcription (RT) and the polymerase chain reaction (PCR). By this approach a murine cDNA probe was obtained that was used to isolate a human cDNA. In this communication the primary sequence of human CPX¶ and the expression of a catalytically active protein in *Escherichia coli* are described.

MATERIALS AND METHODS

Purification, Activity, and Amino Acid Sequencing of CPX from Murine Liver. CPX was purified from 600 g of mouse (Swiss strain) livers according to the method of Bogard *et al.* (5). To remove ampholytes associated with the protein, the enzyme preparations with the highest purity were further applied to a Pharmacia Mono-P FPLC column equilibrated with 50 mM ammonium hydrogen carbonate buffer (pH 7.5) and the enzyme was eluted in a linear gradient of NaCl (0–500 mM in the equilibration buffer). The purified enzyme was dialyzed against 50 mM ammonium hydrogen carbonate buffer (pH 7.5) and freeze dried. The amino-terminal sequence of CPX was obtained from an aliquot of the protein (0.1 nmol) dissolved in 0.1% trifluoroacetic acid in water and subjected to automated Edman degradation on an Applied Biosystems A 470 gas-phase sequencer. A standard degradation program was used and phenylthiohydantoin derivatives of amino acids were identified by on-line HPLC and absorption at 270 nm (Applied Biosystems model 120A).

Internal peptides were obtained by adding two crystals of cyanogen bromide to an aliquot of the protein (1 nmol) dissolved in 70% (wt/vol) formic acid in water. After 18 hr of incubation at room temperature, the sample was diluted 1:1 with water and freeze-dried. Peptides were solubilized in Laemmli SDS/PAGE buffer, separated by 18% acrylamide/SDS gel electrophoresis, and electroblotted to Immobilon TM-P membranes (Millipore). The peptides were covalently trapped on the membrane surface by using the SequeNet attachment procedure as recommended by the manufacturer.

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: CPX, coproporphyrinogen oxidase; RT, reverse transcription; SLIC, single-strand ligation to single-stranded cDNA.
¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. Z28409).

After coupling, the most abundant peptide was subjected to automated solid-phase Edman degradation (Milligen/Biosearch 6600 solid-phase sequencer; Millipore), and the phenylthiohydantoin amino acid derivative was identified by on-line HPLC and absorption at 270 nm (Waters MS HPLC). The enzyme activity was measured by a fluorometric coupled assay (17).

Synthesis of a Murine Probe by PCR. Degenerate oligonucleotides (CO3, DEB2) were prepared on the basis of protein amino-terminus and peptide sequences according to codon usage (18). RNA was isolated from murine liver by using standard techniques (19). After RT, cDNAs were amplified *in vitro* in mixtures of 50 μl containing 70 mM Tris·HCl (pH 8.8), 1.5 mM $MgCl_2$, 0.1% Triton X-100, 200 μmol of each dNTP, 25 pmol of each primer, and 1 unit of *Taq* DNA polymerase (GIBCO/BRL). The cDNA was denaturated for 3 min at 95°C and amplified for 30 cycles (15 sec at 95°C, 30 sec at 43°C, 1 sec at 52°C, 30 sec at 72°C) in a DNA thermal cycler (Hybaid, Middlesex, U.K.). Resulting PCR products were separated by gel electrophoresis in 1.5% agarose and blotted onto a N+ Hybond nylon membrane (Amersham), and specificity was checked by hybridization with the ^{32}P-labeled oligonucleotide COL3, which corresponds to an internal sequence. The desired PCR product was subcloned in the plasmid vector pGEM4Z. The double-stranded templates were sequenced by the dideoxynucleotide chain-termination method, using T7 DNA polymerase and an automated fluorescent DNA sequencer (ALF, Pharmacia) (20).

Synthesis of a Human cDNA Probe. RNA was isolated from the human cell line K562 and reverse-transcribed as described above. The first-strand cDNA was amplified by PCR using a set of primers internal to the mouse sequence (5′-GGAGGACATGAAGACCAAGATGGA-3′ and 5′-TCTGTAGTTGAAATGCATGGTGGG-3′). The specificity of the human PCR product was controlled by hybridization with the mouse cDNA probe, and the amplified fragment was cloned in the plasmid vector pGEM7Zf+ (Promega) and sequenced as described for the mouse cDNA.

Screening of Human cDNA Libraries. cDNA libraries from a human lymphoblastoid cell line in pGEM Blue (21) and from human foreskin fibroblasts in the pcD2 vector (developed by H. Okayama, Osaka University) were screened (22). The libraries were plated at a density of 50,000 colonies per plate (176 cm^2) and transferred in duplicate to nitrocellulose membranes (Schleicher & Schuell). Colony hybridization was carried out, using the cloned human cDNA fragment as a probe. Inserts from positive clones were sequenced as above in combination with the use of oligonucleotide primers.

Cloning of the 5′ End of CPX by Oligodeoxyribonucleotide Ligation to Single-Stranded cDNA. To clone the 5′ end of the CPX gene, the single-strand ligation to single-stranded cDNA (SLIC) strategy was used as described by Dumas *et al.* (23). Briefly, following the preparation of poly(A)$^+$ RNA, the synthesis of the single-stranded cDNA was carried out, using the specific primer 5′-ATCAGCAGCTCCATCTTGGTCTTCAT-3′. Excess primer was removed by differential precipitation in ammonium acetate/2-propanol and RNAs were hydrolyzed in 0.3 M NaOH. Single-stranded ligation was carried out with the oligonucleotide 5′-CTGCATCTATCTAATGCTCCTCTCGCTACCTGCTCACTCTGCGTGACATC-3′, using T4 RNA ligase. This oligonucleotide was synthesized by using a modified phosphoramidite for the first coupling step. Thus the resulting oligonucleotide has a 3′-NH_2 end to prevent self-ligation. The single-stranded cDNA ligation product was used for two-step PCR using primers complementary to the ligated oligonucleotide (OLI-1, 5′-GATGTCACGCAGAGTGAGCAGGTAG-3′; OLI-2, 5′-AGAGTGAGCAGGTAGCGAGAGGAG-3′) and the 5′ region of the previously sequenced cDNA clone (COPCR1, 5′-TCATGTCGCCCGGCCTCCTTC-3′; COPCR-2, 5′-GGGGCCATGAAGCTGCTGC-3′). Desired PCR products were subcloned in the pUC18 vector and sequenced.

Northern Blotting. Total RNA was extracted by standard techniques (19). Poly(A)$^+$ RNA was prepared by using an oligo(dT) column. RNA was electrophoresed on a horizontal 1% agarose/formaldehyde gel, transferred to a nylon filter by capillary action, and hybridized with a 2.5-kb *Xho* I fragment containing a mature CPX coding sequence as a probe. The fragment was radiolabeled with [α-^{32}P]dCTP by random priming.

Southern Blotting. Human genomic DNA was obtained from peripheral lymphocytes, and 15 μg was digested with various restriction enzymes, separated on a 0.8% agarose gel, blotted to a Hybond N+ nylon membrane (Amersham), and hybridized with the same probe as that used for Northern blotting.

Immunodetection of CPX. Cell-free extracts were prepared by mixing the cells in 10% trichloracetic acid in the presence of glass beads (0.4- to 0.5-mm diameter) and neutralized by adding 1 M Tris. The cell lysate was centrifuged for 10 min at 5000 × *g*. The pellet was washed once with diethyl ether and the proteins were solubilized in the Laemmli SDS/PAGE sample buffer. SDS/PAGE was performed, followed by the electrophoretic transfer of the proteins to nitrocellulose sheets. Incubation with the antiserum and visualization with alkaline phosphatase-conjugated anti-rabbit IgG secondary antibodies were performed as recommended by the supplier (Promega).

Prokaryotic CPX Expression. The cDNAs coding for mature human CPX in the sense and antisense orientations were expressed in *E. coli* (24), using the pGEX-2T vector (Pharmacia). To make the construct, primers corresponding to nt 31–52 (5′-CGCTGATCAACTTCGCTGGGGAGGCCGGAG-3′) and nt 1073–1094 (5′-GCGTGATCAAGCTTCAAACCCCTGCACAGCCATTC-3′), respectively, were used to amplify a cDNA fragment. These primers contained *Bcl* I restriction sites which were used to subclone the amplified fragment in the *Bam*HI restriction site of the pGEX-2T plasmid vector. After transformation of *E. coli*, single colonies were isolated and inserts were sequenced. The assay of CPX in lysates of bacterial cultures was performed as previously described (16).

RESULTS

Purified murine CPX appeared as a single band with apparent molecular mass of 35 kDa on an SDS/polyacrylamide gel (Fig. 1). The amino-terminal part of CPX and the most abundant cyanogen bromide peptide (Fig. 1, lane C) were

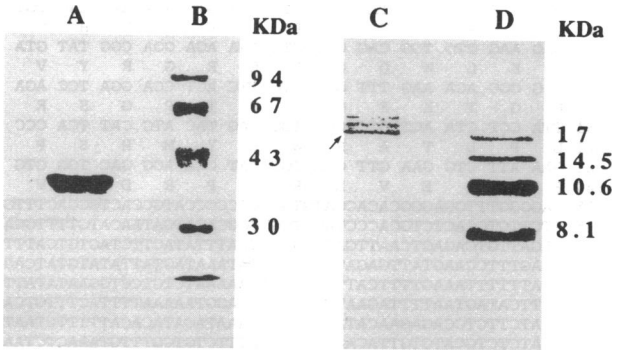


FIG. 1. SDS/PAGE of intact murine CPX and fragments derived by cyanogen bromide. Shown are Coomassie blue-stained gels. Lane A, 10 μg of purified mouse CPX (10% polyacrylamide gel); lane B, molecular mass markers; lane C, cyanogen bromide peptides derived from purified coproporphyrinogen oxidase (15% polyacrylamide gel); and lane D, molecular mass markers.

(S/K)P(T/C)PGRREEDGDELARRSDTFMS[D]TFMSTPPT-------- FNYRYFEVEEA

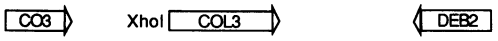
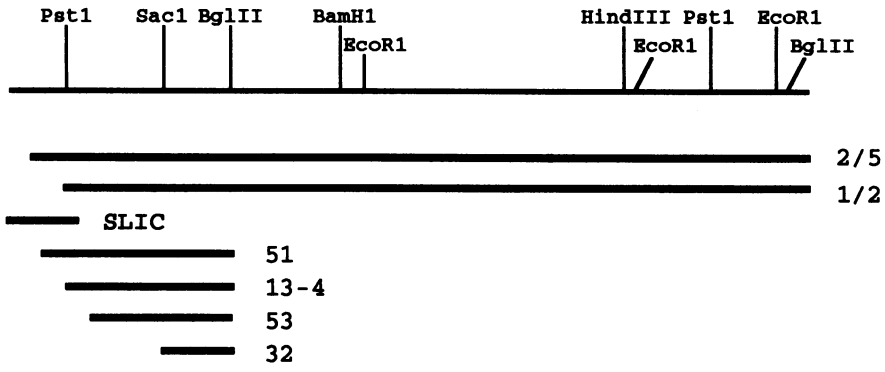


FIG. 2. Sequences of the amino-terminal portion of intact murine CPX and the cyanogen bromide peptide shown in Fig. 1. The open arrows indicate the location and orientation of oligonucleotides based on the amino acid sequence. CO3, 5'-GAGGAGGACGGN-GACGA-3'; COL3, 5'-GACACSTTCATGTCYACSCCSCC-3'; and DEB2, 5'-GCYTCYTCMACYTCAAAGTA-3' (N = any nucleotide, S = C or G, Y = T or C, and M = C or A).

partially sequenced, and degenerate oligonucleotides were designed as shown in Fig. 2. The comparison of the amino acid sequence of the cyanogen bromide peptide with the deduced amino acid sequence from the previously published yeast enzyme sequence (7) helped to localize a peptide approximately 150 amino acids downstream from the amino terminus of the mature protein. Indeed, RT-PCR from mouse liver RNA produced a 470-bp fragment when oligonucleotides CO3 and DEB2 were used as primers, and this fragment gave a hybridization signal with the internal probe COL3 (Fig. 2). The identity of this fragment was confirmed by comparison of its predicted amino acid sequence with the mouse internal and amino-terminal sequences. A human cDNA fragment was obtained by RT-PCR from mRNA of the cell line K562, using oligonucleotides chosen within the mouse cDNA sequence. This cDNA fragment was cloned

A



B

```
-42                                                GGGGCTGGCCACCGCCGCTTCGGGCATGTGCAGCGGGCGGAG   -1
  1 ATG TTG CCT AAG ACC TCG GGG ACG CGG GCC ACT TCG CTG GGG AGG CCG GAG GAG GAG GAG              60
  1 M   L   P   K   T   S   G   T   R   A   T   S   L   G   R   P   E   E   E   E               20
 61 GAT GAG CTG GCC CAC CGC TGC AGC AGC TTC ATG GCC CCG CCT GTG ACC GAC CTG GGC GAG              120
 21 D   E   L   A   H   R   C   S   S   F   M   A   P   P   V   T   D   L   G   E               40
121 CTG CGA AGG AGG CCG GGC GAC ATG AAG ACC AAG ATG GAG CTG CTG ATT CTG GAG ACC CAG              180
 41 L   R   R   R   P   G   D   M   K   T   K   M   E   L   L   I   L   E   T   Q               60
181 GCC CAG GTG TGC CAG GCT CTG GCA CAG GTA GAC GGG GGC GCC AAC TTT TCT GTG GAC CGG              240
 61 A   Q   V   C   Q   A   L   A   Q   V   D   G   G   A   N   F   S   V   D   R               80
241 TGG GAG AGG AAG GAA GGA GGT GGC GGC ATC AGC TGT GTA CTT CAA GAT GGG TGT GTT TTC              300
 81 W   E   R   K   E   G   G   G   G   I   S   C   V   L   Q   D   G   C   V   F              100
301 GAA AAG GCT GGG GTG AGC ATT TCT GTT GTT CAT GGA AAT CCT CAT GCT CCT ACT ATC CAT TTC AAC      360
101 E   K   A   G   V   S   I   S   V   V   H   G   N   L   S   E   E   A   A   K              120
361 CAA ATG AGA AGC AGA GGA AAA GTT CTG AAG ACT AAA GAT GGT AAA TTG CCA TTT TGT GCT              420
121 Q   M   R   S   R   G   K   V   L   K   T   K   D   G   K   L   P   F   C   A              140
421 ATG GGC GTG AGC TCT GTT ATC CAC CCC AAG AAT CCT CAT GCT CCT ACT ATC CAT TTC AAC              480
141 M   G   V   S   S   V   I   H   P   K   N   P   H   A   P   T   I   H   F   N              160
481 TAC AGA TAC TTT GAA GTA GAA GAA GCT GAT GGC CAC AAG CAG TGG TGG TTT GGT GGT GGA              540
161 Y   R   Y   F   E   V   E   E   A   D   G   H   K   Q   W   W   F   G   G   G              180
541 TGT GAC CTC ACT CCA ACA TAC TTG AAT CAA GAA GAC GCT GTC CAT TTT CAC AGA ACT CTG              600
181 C   D   L   T   P   T   Y   L   N   Q   E   D   A   V   H   F   H   R   T   L              200
601 AAG GAG GCT TGT GAC CAG CAT GGT CCA GAT CTC TAC CCC AAA TTT AAA AAA TGG TGT GAT              660
201 K   E   A   C   D   Q   H   G   P   D   L   Y   P   K   F   K   K   W   C   D              220
661 GAT TAC TTC TTT ATA GCC CAT CGT GGA GAG CGG CGG GGC ATT GGT GGT ATC TTT TTT GAT              720
221 D   Y   F   F   I   A   H   R   G   E   R   R   G   I   G   G   I   F   F   D              240
721 GAT CTT GAC TCT CCG TCC AAG GAG GAG GTG TTT CGC TTT GTA CAG AGC TGT GCC AGG GCT              780
241 D   L   D   S   P   S   K   E   E   V   F   R   F   V   Q   S   C   A   R   A              260
781 GTA GTT CCT TCT TAC ATT CCC CTT GTG AAA AAG CAC TGT GAT GAC TCA TTC ACC CCC CAG              840
261 V   V   P   S   Y   I   P   L   V   K   K   H   C   D   D   S   F   T   P   Q              280
841 GAG AAG CTG TGG CAG CAG CTC AGA AGA GGA CGG TAT GTA GAA TTT AAT CTG CTG TAT GAT              900
281 E   K   L   W   Q   Q   L   R   R   G   R   Y   V   E   F   N   L   L   Y   D              300
901 CGG GGC ACA AAG TTT GGC CTC TTC ACT CCA GGA TCC AGA ATT GAA AGT ATC TTG ATG TCT              960
301 R   G   T   K   F   G   L   F   T   P   G   S   R   I   E   S   I   L   M   S              320
961 TTA CCT CTA ACT GCC CGA TGG GAG TAC ATG CAT TCA CCC TCA GAG AAT TCC AAA GAA GCT             1020
321 L   P   L   T   A   R   W   E   Y   M   H   S   P   S   E   N   S   K   E   A              340
1021 GAA ATT CTG GAA GTT CTA CGC CAT CCA AGG GAC TGG GTG CGT TGA TGCAGGCAGAATGGCTGTG            1084
341 E   I   L   E   V   L   R   H   P   R   D   W   V   R   *
1085 CAGGGGTTTGGAGGGCACACGATGTGTGCTCCCCCATGCCACTGGCACTTTGCCACTGTGTGGCAGTTACCCGTGCCTTA
1165 GTCTTCTCCACTCTGCACCCTACCTCGTGGGCAGATGATAACATGTTTTGGATGCTGTCAGTGATGAATGGTGGGATGGC
1245 CAGATTGTCAGAGTCAATTGATTAAACCTCATTTATACTTCTAGTGTCATTTTATATGACTAGTTTACAAAATAGGACAT
1325 TGAGTTTCCAAGTATTGAGATAAGGGAATATAAATAGTATTATATGTATCAGGAAATCTCTCATCTTGTTTTTGTTTCAT
1405 GTATTTTTTAAAGTTTTCATTTGTGCCACAAAAATCTGTCGTGGAATATATTTTATTTTCATTAATTCAGTGAAGTTGAG
1485 ACTTCATAGTAATTTTAGAATGCAACTTGAAGGTAAAAATTTTACTTTGTCAATACTGAAGTCTCTGCTGTAATCCTTAT
1565 ATATCTTCTCCAGAGAACATAATATTGTCAAATAGATACACATTTTTCTAATAGGTATTTAGAAGCACTTGAAATATTCT
1645 TAATCTCTGCATGTGTTACAATTCAGTTATTTCTGTCGTTTGTAAACTCTAAAGTGACATTACTATTATTTTAGAGATGT
1725 CTAAGTTGTAATTTTGATTTTTGTGGAACCATTGTTTGTTAATGTTGGGATTTCTGCACTTTTGAATGTGAAAGCTTATA
1805 TCCCTGAATTCTGATACTTAAGAGTTTTCTATTTCTGACATCTCTGTGTGGAAGTTGAGACTAAGAATAATCCTAGGGAT
1885 GTCATGAATTTAGGCAATGTTTCTCATTTGGAAAATGAAATGAGAAATAATTTCCTTCTTTAAAGCAAGTATATATAGTA
1965 TGAGAAACTTGGAGGCATTTCATACACACGTTTCTTAGGAAAATGGACACATTGAAAATGTCCTCTTTTTTATATTAGAG
2045 ATTCTGCAGCTCTTTGCTCTTAAGAGCAAATCACAACAGGATTCTTAATGTATGATTTCTTTGTTCATATATTATGAATG
2125 TATTATTTTATTTGCTTCGTAATAAAGTTTATAAGAAAGAGCATCTCATACATATCATTATCGTGGAACACGTTGAATGT
2205 TTGTGATTCTGTGTGGCCTTTTTGGGGCTGGAAAATGTATGAATTCTTCAACTGTCTTACAAGAGATCTGCTAATAAATT
2285 TTATTATTAAATAAGAAAAAAAA
```

FIG. 3. (*A*) Restriction map of cDNA inserts and a schematic representation of cDNA sequences obtained from SLIC experiments. (*B*) Nucleotide sequence of human CPX cDNA and the deduced amino acid sequence. The peptides homologous to amino acid sequences of the corresponding murine amino-terminal part of the CPX and the cyanogen bromide-derived peptide are underlined. Consensus polyadenylylation signals AATAAA are located 23 and 154 nt 5' to the poly(A) tail. The 5' end of the longest cDNA insert isolated from the cDNA library starts at nt 30.

and sequenced for use as a probe. A foreskin fibroblast cDNA library in the pcD2 plasmid vector (22) was screened; from 100,000 colonies, two independent positive clones (2/5 and 1/2) were isolated. Plasmids from these clones were subjected to restriction endonuclease mapping (Fig. 3), and the longest insert was sequenced. An open reading frame of 1032 nucleotides which ended with a TGA stop codon was found. However, the first ATG found in this sequence encoded a methionine which, based on comparison with the mouse amino-terminal protein sequence, corresponded to the 21st amino acid of the mature protein. This indicated that the 5′ part of the human cDNA was absent from the isolated clones. Screening another human library from a lymphoblastoid cell line (200,000 clones) resulted only in the isolation of shorter cDNA clones. Therefore, we used a modification of the anchored PCR technique, SLIC (23), to obtain sequences situated 5′ of our cDNA clones. PCR products obtained from SLIC experiments were cloned, and 40 clones hybridizing to a probe situated at the 5′ end of the known cDNA sequence were further analyzed. Sequencing of these clones allowed us to deduce 71 additional nucleotides. The nucleotide sequence of the cDNA for human CPX was constructed from overlapping sequences of cDNA clones and SLIC products. The compiled sequence contained 2341 nt with a single open reading frame beginning with an ATG codon and ending with a TGA stop codon. The coding region was followed by 1237 nt of untranslated 3′ region with two putative polyadenylylation signals.

The nucleotide sequence coding for the putative mature human CPX was expressed in *E. coli* as a fusion protein with glutathione *S*-transferase. Lysates of bacterial cultures containing the expression plasmid displayed a 17-fold increase in CPX activity as compared with bacteria transformed with a plasmid containing the CPX sequences in the reverse orientation (Table 1).

The sizes of mature CPX proteins and those of CPX mRNAs were estimated in two human cell lines by Western and Northern blotting, respectively. The cell line K562 is an erythroid cell line, whereas HepG2 is of hepatic origin. With both cell lines a single mRNA of approximately 2500 bases was detected. This size was consistent with that of our longest cDNA clone. The mature enzyme from these two cell lines also displayed the same apparent molecular mass, ≈35 kDa (Fig. 4).

Southern blotting of human genomic DNA probed with the cDNA of clone 2/5 revealed a simple restriction pattern, most consistent with a single human CPX gene (Fig. 5).

DISCUSSION

The results of the present study revealed the primary structure of the human CPX cDNA and its predicted amino acid sequence. The cDNA clones for murine CPX were obtained by RT-PCR from mouse liver mRNA, using degenerate primers. The murine cDNA clones were identified as containing CPX sequences on the basis of the identity between the amino acid sequences from the purified protein and those deduced from the cDNA. The sequence of the murine cDNA allowed us to design oligonucleotide primers to amplify and clone a portion of human cDNA, which proved to be highly

Table 1. Expression of human CPX in *E. coli*

Construct	CPX activity,* pmol/hr per mg
pGEX-2T	551
pGEX-2T: CPX sense	9747
pGEX-2T: CPX antisense	572

*In pmol of protoporphyrinogen IX synthesized per hr per mg of protein at 37°C, expressed as mean of two duplicate experiments.

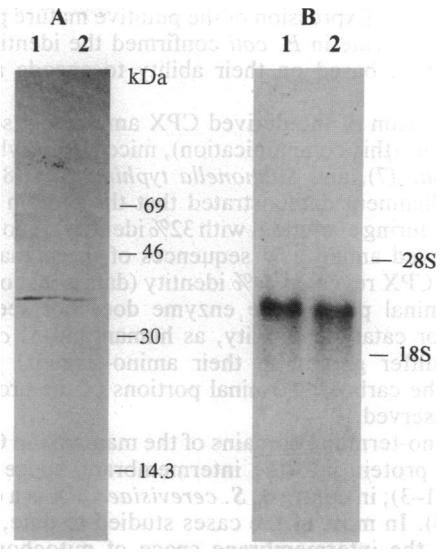


FIG. 4. Western and Northern blot analysis of CPX protein and mRNA. (*A*) Western blot using rabbit antiserum to murine liver CPX. Total protein homogenates from K562 (lane 1) and HepG2 (lane 2) cell lines were applied on an SDS/10% polyacrylamide gel, electrophoresed, electroblotted onto a nitrocellulose filter, incubated with the antiserum, and detected with alkaline phosphatase-conjugated anti-rabbit IgG secondary antibodies. (*B*) Northern blot analysis of CPX mRNA. Poly(A)$^+$ RNAs (5 μg) were isolated from K562 (lane 1) and HepG2 (lane 2) cell lines, electrophoresed into a formaldehyde/1% agarose gel, transferred to a nylon filter by capillary blotting, and probed with a radiolabeled cDNA from clone 2/5. Positions of marker rRNAs are indicated.

similar (84% identical) to its murine homologue. This human cDNA fragment was used as a probe to screen human cDNA libraries, permitting the isolation of cDNA clones spanning most of the region coding for the mature enzyme and the entire 3′ untranslated region. Upstream sequences were obtained after cloning the products from SLIC experiments. From this additional information a potential start codon was identified. On the basis of a comparison with the amino-terminal sequence of the murine enzyme, the initiation codon would lie 30 bases upstream of the sequence coding for the

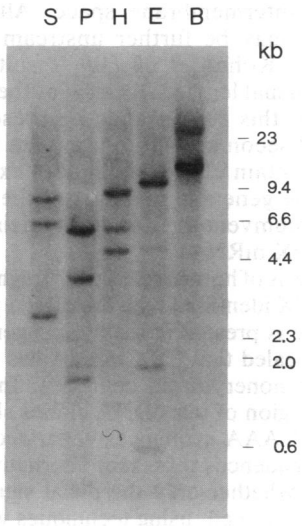


FIG. 5. Southern blot analysis of CPX. Fifteen micrograms of human genomic lymphocyte DNA was digested with the indicated restriction endonucleases and electrophoresed into a 0.8% agarose gel. The DNA was transferred to a nylon membrane by capillary action and hybridized with radiolabeled cDNA from clone 2/5. S, *Sac* I; P, *Pst* I; H, *Hin*dIII; E, *Eco*RI; B, *Bam*HI.

mature protein. Expression of the putative mature portion of the human enzyme in *E. coli* confirmed the identity of our cDNA clones, based on their ability to encode an active enzyme.

A comparison of the derived CPX amino acid sequences from humans (this communication), mice (10), soybeans (9), *S. cerevisiae* (7), and *Salmonella typhimurium* (8) by progressive alignment demonstrated that the protein is highly conserved during evolution, with 32% identity. A comparison of the derived amino acid sequences of the human and *S. cerevisiae* CPX revealed 49% identity (data not shown). The amino-terminal part of the enzyme does not seem to be required for catalytic activity, as human and *S. cerevisiae* enzymes differ greatly in their amino-terminal parts. In contrast, the carboxyl-terminal portions of the proteins are highly conserved.

The amino-terminal domains of the mammalian CPX may direct the protein into the intermembrane space of mitochondria (1–3); in contrast, *S. cerevisiae* CPX is a cytosolic enzyme (4). In most of the cases studied to date, proteins located in the intermembrane space of mitochondria are synthesized with relatively long presequences and undergo a two-step cleavage during mitochondrial import. The amino-terminal portions of these presequences are positively charged and the carboxyl-terminal portion contains hydrophobic residues that function as the intermembrane space targeting domain (25). The putative mitochondrial targeting domain deduced from the cDNA sequence of CPX would be only 10 amino acids in length. Alignment of the human and mouse nucleotide sequences indicated that the deduced initiating codon in the mouse sequence corresponds to the first ATG in our cDNA. Kohno *et al.* (10) proposed that the mouse presequence was 31 amino acids in length on the basis of a comparison with the amino-terminal sequence of the bovine enzyme. However, our protein microsequencing data of the murine enzyme indicate that the mouse presequence would contain only 10 amino acids. Whether the discrepancy between the bovine (10) and murine enzyme sequences is related to species differences or to partial proteolysis during purification of the bovine protein remains to be determined. Experiments specifically designed to study the import of the precursor form into mitochondria will demonstrate whether or not this unusually short presequence is sufficient to target the mature enzyme to the intermembrane space. Alternatively, the initiation codon may be further upstream than that proposed by us and Kohno *et al.* (10), resulting in a presequence of more usual length. Because of the extremely rich G+C content of this region and the presence of a very strong predicted secondary structure (data not shown), we were unable to obtain cDNA sequences extending further 5′. Cloning of the gene and mapping of the transcriptional start site will circumvent the problems arising during RT of the 5′ end of CPX mRNA.

Southern analysis of human restriction fragments containing sequences for CPX identified with the cDNA probe suggested that a single gene is present in the human genome. Northern blot analyses revealed that CPX mRNA has the same length in erythroid and nonerythroid cell lines. Inspection of the 3′-untranslated region of our cDNA clones showed the presence of two AATAAA sequences separated by 125 bases. Whether these sequences represent alternative polyadenylylation signals or whether only the distal signal is functional remains to be determined, using techniques with high resolution. The level of CPX transcripts seemed to be higher in K562 cells than in HepG2 cells, and the CPX protein was found to be proportionally increased. CPX mRNA was also found to increase during the erythroid differentiation of mouse erythroleukemic cells in culture (10). These data suggest that CPX expression is differentially regulated in erythroid and nonerythroid cells at a pretranslational level. Characterization of the gene will clarify the organization of promoter region and help to elucidate the underlying mechanisms.

The human CPX cDNA will permit future studies on the expression of the CPX gene, the regulation of the heme biosynthetic pathway, and the identification of the molecular defects in human hereditary coproporphyria.

We thank Dr. Rosine Labbe-Bois for her valuable help in our initial attempts to clone human CPX by functional complementation of yeast CPX disruptants and Dr. Allan Hance for critical review of the manuscript. This work was supported by Institut National de la Santé et de la Recherche Médicale Research Grant CJF 8904, by the Association Française Contre les Myopathies, by the Association Claude Bernard, by the Centre National de la Recherche Scientifique, and by the University of Paris 7.

1. Elder, G. H. & Evans, J. O. (1978) *Biochem. J.* **172,** 345–347.
2. Grandchamp, B., Phung, N. & Nordmann, Y. (1978) *Biochem. J.* **176,** 97–102.
3. Dailey, H. A. (1990) in *Biosynthesis of Heme and Chlorophylls*, ed. Dailey, H. A. (McGraw-Hill, New York), pp. 123–161.
4. Camadro, J. M., Chambon, H., Jolles, J. & Labbe, P. (1986) *Eur. J. Biochem.* **156,** 579–587.
5. Bogard, M., Camadro, J. M., Nordmann, Y. & Labbe, P. (1989) *Eur. J. Biochem.* **181,** 417–421.
6. Yoshinaga, T. & Sano, S. (1980) *J. Biol. Chem.* **255,** 4722–4726.
7. Zagorec, M., Buhler, J. M., Treich, I., Keng, T., Guarente, L. & Labbe-Bois, R. (1988) *J. Biol. Chem.* **263,** 9718–9724.
8. Xu, K. & Elliott, T. (1993) *J. Bacteriol.* **16,** 4990–4999.
9. Madsen, O., Sandal, L., Sandal, N. N. & Marcker, K. A. (1993) *Plant Mol. Biol.* **23,** 35–43.
10. Kohno, H., Furukawa, T., Yoshinaga, T., Tokunaga, R. & Taketani, S. (1993) *J. Biol. Chem.* **268,** 21350–21363.
11. Bottomley, S. & Müller-Eberhard, U. (1988) *Semin. Hematol.* **25,** 282–302.
12. Sassa, S. (1988) *Semin. Hematol.* **25,** 312–320.
13. Conder, L. H., Woodard, S. I. & Dailey, H. A. (1991) *Biochem. J.* **275,** 321–326.
14. Kappas, A., Sassa, S., Galbraith, R. A. & Nordmann, Y. (1989) in *The Metabolic Basis of Inherited Diseases*, eds. Scriver, C. R., Beaudet, A. L., Sly, S. W. & Valle, D. (Academic, New York), pp. 1305–1365.
15. Grandchamp, B., Phung, N. & Nordmann, Y. (1977) *Lancet* **ii,** 1348–1349.
16. Nordmann, Y., Grandchamp, B., de Verneuil, H. & Phung, L. (1983) *J. Clin. Invest.* **72,** 1139–1149.
17. Labbe, P., Camadro, J. M. & Chambon, H. (1985) *Anal. Biochem.* **149,** 248–260.
18. Lathe, R. (1985) *J. Mol. Biol.* **183,** 1–12.
19. Chirgwin, J. M., Pryzybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
20. Ansorge, W., Sproat, B. S., Stegemann, J. & Schwager, C. (1986) *J. Biochem. Biophys. Methods* **13,** 315–323.
21. Romana, M., Grandchamp, B., Dubart, A., Amselem, S., Chabret, C., Nordmann, Y., Goossens, M. & Romeo, P. H. (1991) *Eur. J. Clin. Invest.* **21,** 225.
22. Igarashi, M., Nagata, A., Jinno, S., Suto, K. & Okayama, H. (1991) *Nature (London)* **353,** 80–83.
23. Dumas, M. E. J. B., Delort, J. & Mallet, J. (1992) *Molecular Biology 15, PCR Protocols: Current Methods and Applications* (Humama, Totowa, NJ), pp. 365–385.
24. Smith, D. B. & Johnson, K. S. (1988) *Gene* **67,** 31–40.
25. Glick, B. S., Beasley, E. M. & Schatz, G. (1992) *Trends Biol. Sci.* **17,** 453–459.