Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7049–7053, October 1988
Biochemistry

Human uroporphyrinogen III synthase: Molecular cloning, nucleotide sequence, and expression of a full-length cDNA

(heme biosynthesis/congenital erythropoietic porphyria/prokaryotic expression)

SHIH-FENG TSAI, DAVID F. BISHOP, AND ROBERT J. DESNICK*

Division of Medical Genetics, Mount Sinai School of Medicine, New York, NY 10029

Communicated by Lawrence Bogorad, April 21, 1988 (received for review December 3, 1987)

ABSTRACT **Uroporphyrinogen III synthase [URO-synthase; hydroxymethylbilane hydro-lyase (cyclizing), EC 4.2.1.75], the fourth enzyme in the heme biosynthetic pathway, is responsible for conversion of the linear tetrapyrrole, hydroxymethylbilane, to the cyclic tetrapyrrole, uroporphyrinogen III. The deficient activity of URO-synthase is the enzymatic defect in the autosomal recessive disorder congenital erythropoietic porphyria. To facilitate the isolation of a full-length cDNA for human URO-synthase, the human erythrocyte enzyme was purified to homogeneity and 81 nonoverlapping amino acids were determined by microsequencing the N terminus and four tryptic peptides. Two synthetic oligonucleotide mixtures were used to screen 1.2 × 10^6 recombinants from a human adult liver cDNA library. Eight clones were positive with both oligonucleotide mixtures. Of these, dideoxy sequencing of the 1.3 kilobase insert from clone pUROS-2 revealed 5′ and 3′ untranslated sequences of 196 and 284 base pairs, respectively, and an open reading frame of 798 base pairs encoding a protein of 265 amino acids with a predicted molecular mass of 28,607 Da. The authenticity of this clone was established by colinearity of the predicted amino acid sequence with 81 microsequenced residues from the purified enzyme. In addition, high levels of enzymatic activity and immunoreactive protein were expressed when a blunt-ended 971-base-pair *Ava* II cDNA fragment containing the entire coding region was inserted into vectors for expression in *Escherichia coli*. The isolation and expression of this full-length cDNA for human URO-synthase should facilitate studies of the structure, organization, and chromosomal localization of this heme biosynthetic gene as well as the characterization of the molecular lesions causing congenital erythropoietic porphyria.**

Uroporphyrinogen III synthase [URO-synthase; hydroxymethylbilane hydro-lyase (cyclizing), EC 4.2.1.75] was identified in 1953 by Bogorad (1) who demonstrated that both URO-synthase and hydroxymethylbilane synthase [HMB-synthase; previously designated porphobilinogen deaminase; porphobilinogen ammonia-lyase (polymerizing), EC 4.3.1.8] were required for the conversion of the monopyrrole, porphobilinogen, to the cyclic tetrapyrrole, uroporphyrinogen III (2). During the next two decades, the precise function of the two enzymes in uroporphyrinogen III biosynthesis was the subject of active investigation and debate, particularly with regard to their reaction mechanisms and intermediates and their potential interaction in a cytosolic enzyme complex (for reviews, see refs. 3 and 4). It is now recognized that HMB-synthase catalyzes the head to tail condensation of four molecules of porphobilinogen to form the linear tetrapyrrole HMB (4, 5). In the presence of URO-synthase, HMB is rapidly converted to uroporphyrinogen III by an intramolecular rearrangement of the D-pyrrole group and ring closure (5–7). In the absence of URO-synthase, HMB nonenzymatically cyclizes to form the nonphysiologic uroporphyrinogen I isomer.

The deficient activity of URO-synthase is the enzymatic defect in congenital erythropoietic porphyria, an inborn error of heme biosynthesis that is inherited as an autosomal recessive trait (8). Enzymatic diagnosis of affected homozygotes and asymptomatic heterozygous carriers can be made by the demonstration of deficient or intermediate levels of URO-synthase activity, respectively (9).

To determine the molecular nature of the mutations in the URO-synthase gene that cause congenital erythropoietic porphyria and to further characterize the structural and functional properties of this heme biosynthetic enzyme, efforts were directed to isolate the full-length cDNA encoding human URO-synthase. In this communication, the isolation and complete nucleotide sequence of a full-length cDNA encoding human URO-synthase are described.† In addition, the use of the URO-synthase cDNA for the prokaryotic expression of large amounts of catalytically active enzyme protein is demonstrated.

EXPERIMENTAL PROCEDURES

Amino Acid Microsequencing. Homogeneous URO-synthase was purified from human erythrocytes (7). The post-HPLC peak 2 preparation, which retained enzymatic activity, was used for N-terminal microsequencing. The post-Sephadex G-100 preparation was subjected to preparative $NaDodSO_4$/PAGE and the 29.5-kDa band was electroeluted, digested with tosylphenylalanine chloromethyl ketone-treated trypsin (10), and the peptides were isolated by reversed-phase HPLC. The amino acid sequences of the amino-terminal and selected tryptic peptides (Fig. 1, T-5, T-9, T-15, and T-20) were determined by automated gas-phase microsequencing and HPLC identification of the phenylthiohydantoin amino acid derivatives (11).

Construction of Synthetic Oligonucleotides. Oligonucleotides for library screening and DNA sequencing were synthesized on a Biosearch Synthesizer by β-cyanoethylphosphoramidite chemistry. Two oligonucleotide mixtures (17-mers) were constructed to all codons for an amino-terminal and an internal tryptic peptide (T-9) sequence. Unique sequence oligonucleotides (17-mers) were synthesized for use as primers in sequencing reactions. Aliquots of all oligonucleotide solutions were 5′-end-labeled with [γ-^{32}P]-ATP and T4 polynucleotide kinase (12).

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: HMB, hydroxymethylbilane; nt, nucleotide(s); URO-synthase, uroporphyrinogen III synthase; IPTG, isopropyl β-D-thiogalactopyranoside.

*To whom reprint requests should be addressed at: Mount Sinai School of Medicine, Fifth Avenue at 100th Street, New York, NY 10029.

†The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03824).

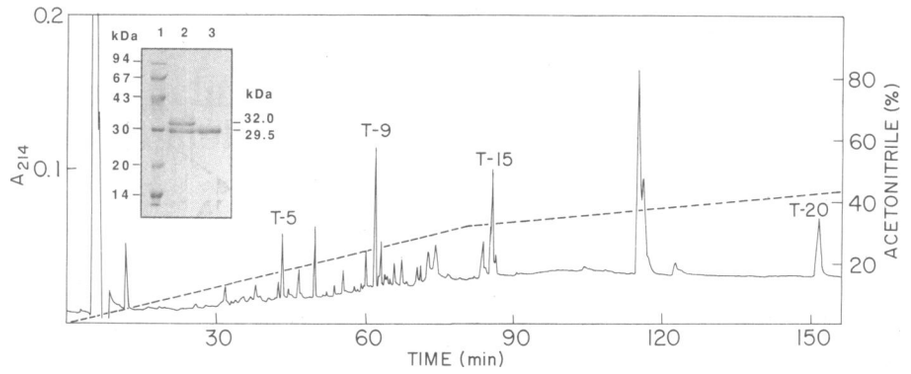

FIG. 1. Electroelution of purified URO-synthase and separation of tryptic peptides. (*Inset*) NaDodSO$_4$/PAGE of electroeluted URO-synthase. Lanes: 1, molecular mass standards; 2, post-Sephadex G-100 preparation of human URO-synthase (4 μg, 97,000-fold purified); 3, electroeluted protein (2 μg). (*Main*) Reversed-phase HPLC separation of tryptic peptides from human URO-synthase. The electroeluted 29.5-kDa protein was digested with tosylphenylalanine chloromethyl ketone-treated trypsin and chromatographed on a Vydac C4 reversed-phase column equilibrated with solution A (100% water containing 0.05% trifluoracetic acid). The peptides were separated on a linear gradient of 0–45% solution B (80% acetonitrile/20% water containing 0.05% trifluoroacetic acid) for 80 min followed by 45–55% solution B for 80 min at a flow rate of 0.7 ml/min. Protein concentration was monitored by absorbance at 214 nm.

Isolation and Characterization of cDNA Clones. A total of 1.2×10^6 recombinants from the pKT218 human adult liver cDNA library (13), kindly provided by Stuart Orkin (Harvard Medical School), were screened with both sets of radiolabeled oligonucleotide mixtures by colony hybridization (14). The filters were prehybridized at 50°C for 4 hr with 6× SSC/5× Denhardt's solution/denatured salmon sperm DNA (100 μg/ml) (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0; 1× Denhardt's solution = 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin). Hybridization of each radiolabeled oligonucleotide mixture (1×10^6 cpm/ml) was carried out at 50°C for 16 hr. Stringent washing was performed at 50°C with 6× SSC/0.1% NaDodSO$_4$ for 3 hr. Plasmid DNA from positive clones was prepared by alkaline lysis (15). After digestion with *Pst* I, the insert fragments were electrophoresed in a 1% agarose gel and then transferred to nitrocellulose filters (16). Conditions for hybridization and washing were the same as those used for screening. Positive clones were initially grouped by insert size; plasmid DNA from representative clones of each group was nick-translated (12) for cross-hybridization experiments.

DNA Sequencing. The *Pst* I insert from positive clone pUROS-2 was subcloned directly into M13mp18, or into M13mp18 and M13mp19 after digestion with *Sau*3AI or *Hin*dIII. All DNA sequencing reactions were carried out by primer extension (17) using either the M13 universal primer or URO-synthase-specific synthetic oligonucleotides.

Computer Analysis. Sequence similarity searches were carried out with the National Institutes of Health GenBank (18) nucleic acid data base (release 48.0) using the FASTN program (19) and the National Biomedical Research Foundation protein data base (release 13.0) using the IFIND/ALIGN programs (using default parameters) on the Bionet Network (IntelliGenetics). The SEQHP and SEQDP routines of the IDEAS program (20) were used to determine optimal alignments and homology significance; the number of random sequence comparisons was 500.

Prokaryotic Expression of Human URO-Synthase. For expression of the human URO-synthase cDNA, the blunt-ended 971-base-pair (bp) *Ava* II fragment (after Klenow treatment) containing the entire coding region (798 bp), was inserted into the Bluescript KS(−) (pBSKS; Stratagene) and the pKK223-3 (Pharmacia) vectors by standard techniques (12). The pBSKS plasmid DNA was digested with *Kpn* I and treated with T4 DNA polymerase to generate blunt ends. The recombinant construct, designated pBSKS-UROS, contained the URO-synthase coding region in-frame behind a portion of the β-galactosidase sequence in the vector, permitting the inducible expression of a fusion protein in transformed *Escherichia coli* JM109. Similarly, the blunt-ended *Ava* II URO-synthase fragment was inserted into the *Sma* I site or *Eco*RI cloning site (after generating blunt ends with Klenow enzyme) of pKK223-3. These constructs, designated pKK223-3S-UROS and pKK223-3E-UROS, respectively, were used to transform *E. coli* JM105 for expression of the native enzyme protein. For expression, sense and antisense clones were grown for 16 hr in LB medium, each culture (100 μl) was used to inoculate 5 ml of LB medium, and then cultures were incubated at 37°C to late logarithmic phase. Aliquots (≈2 ml) of each culture were removed and incubated for 3 hr in the presence or absence of 5 mM isopropyl β-D-thiogalactopyranoside (IPTG). Then, 1.5 ml of each culture was microfuged for 30 s and the pellets were washed with 0.9% NaCl. The cells were resuspended in 250 μl of URO-synthase lysis buffer (9) and disrupted (three 15-s cycles) at 0°C with a Branson cup sonifier. The lysate was microfuged for 5 min and the supernatants were assayed for URO-synthase activity (9). Immunoblotting of the expressed URO-synthase was performed with goat anti-human URO-synthase antibodies (7).

RESULTS

Amino Acid Sequencing and Oligonucleotide Synthesis. Microsequencing of the N terminus and four tryptic peptides

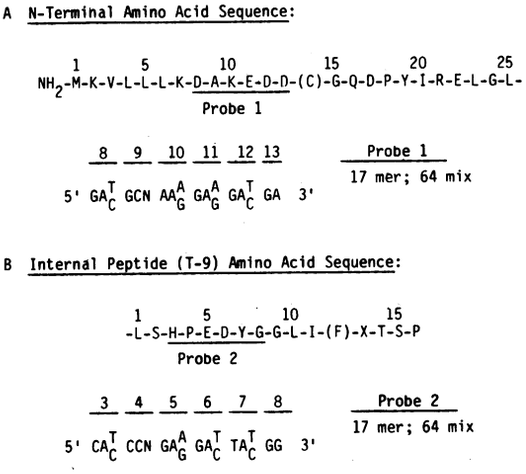

FIG. 2. Amino acid sequences (single-letter code) of the human URO-synthase amino terminus and internal peptide T-9 and the corresponding oligonucleotide mixtures that were used as probes for library screening. Uncertain residues are in parentheses.

from human URO-synthase identified a total of 81 nonoverlapping residues representing 30% of the total amino acid sequence. Synthetic oligonucleotide mixtures were constructed to contain all possible codons for an N-terminal sequence (probe 1) and a sequence in peptide T-9 (probe 2; Fig. 2). Both probes were 17 nucleotides (nt) long and were mixtures of 64 species.

Library Screening and Characterization of Positive cDNA Clones. Of the 28 clones initially selected, eight remained positive after purification. Three clones (designated pUROS-1, -2, and -3) each had a 1.3-kilobase (kb) insert and *Pst* I fragments of approximately 800, 400, and 100 bp; two (pUROS-4 and -5) had 1.25-kb inserts and 750- and 500-bp *Pst* I fragments; and three (pUROS-6, -7, and -8) had a single 800-bp *Pst* I fragment. Probes 1 and 2 both hydridized to the 800- and 750-bp fragments. The inserts from the three groups of positive clones were shown to be related by cross-hybridization, restriction mapping with *Hin*dIII and *Sau*3AI, and by partial sequencing.

Nucleotide and Protein Sequence Analyses. The 1.3-kb pUROS-2 insert was completely sequenced on both strands (Fig. 3). An open reading frame of 798 nt encoded a protein of 265 amino acids. The predicted molecular mass of 28,607 Da was consistent with that (29,500 Da) estimated by $NaDodSO_4$/PAGE of the purified enzyme (7). Colinearity was observed between the predicted sequence and the 81 nonoverlapping amino acids in the microsequenced N terminus and four tryptic peptides (Fig. 3). In addition, the amino acid composition predicted by the URO-synthase hepatic cDNA and that of the purified erythrocyte enzyme (7) were remarkably similar (data not shown). There were no ATG triplets in the 196-nt 5′ untranslated sequence. In the 284-nt 3′ untranslated sequence, a consensus polyadenylylation signal AATAAA (21) was identified 21 nt preceding the poly(A) tract. The consensus recognition sequence (CACTG) for the U4 small nuclear ribonucleoprotein (22) was located 13 nt before the poly(A) tract.

Searches of nucleic acid and protein data bases revealed no significant direct or inverted repeats in the cDNA and no extensive amino acid sequence similarities between the open reading frame of URO-synthase and any other coding sequence. Notably, there were no significant similarities with sequences for the available cDNAs encoding the other human heme biosynthetic enzymes (23–26). Amino acid sequence similarities were detected for short regions in URO-synthase and several other plant and animal proteins. As shown in Fig. 4, residues 24–71 in URO-synthase contained regions that were similar to sequences in the enzymes for potassium-transporting ATPase B chain (KdpB) from *E. coli* (27), *N*-(5′-phosphoribosyl)anthranilate isomerase (trpF) from *E. coli* (28), human cytochrome *c* oxidase, polypeptide II (29), and human globin α-chain (30). Other shorter regions of similarity were observed for URO-synthase residues 227–240 with residues 16–28 in *Agrobacterium tumefaciens* Ti plasmid Ach5 tmr protein (31) and URO-synthase residues 99–155 and 217–228 with residues 216–269 and 101–112, respectively, in the *Rhizobium meliloti* nif-specific regulatory protein (*nifA*) (32).

Prokaryotic Expression of Human URO-Synthase. Blunt-end ligation of the URO-synthase *Ava* II fragment into the *Eco*RI or *Sma* I sites of pKK223-3 positioned the URO-synthase coding sequence 45 and 48 bp, respectively, from the Shine–Dalgarno sequence in the vector (33) and 13 bp from the Shine–Dalgarno consensus sequence in the URO-synthase 5′ untranslated region (Fig. 5*A*). These vectors were designed to express the native protein from the URO-synthase initiation codon. Fig. 6*A* shows an immunoblot demonstrating the comigration of purified human URO-synthase with the 29.5-kDa enzyme protein expressed in *E. coli* JM105 by pKK223-3E-UROS. IPTG induction was required to express sufficient enzyme protein for immuno-

```
                                                                                                          Pst I
-196                  TCCTGG GGCCCAGCGC GGGTGGCTGC CGCGGCCCCT CGGGCTGCGT GGGGAGGGGG CTTCCGCCCC TGTTGTCATT GCTCCTGCAG C  -110
                                                                                       Ava II
-109  CTTTTCGCT GGGACTGCGC GACACCGCCC CCCGACCGGG TGCCCGCTGT GTGCCAGGCC GGGTGCTGGG CACGGTCCCG CGAGTGCCCT ATAAGGACTG CCAGGCAATA   -1

   1  ATG AAG GTT CTT TTA CTG AAG GAT GCG AAG GAA GAT GAC TGT GGC CAG GAT CCG TAT ATC AGG GAA TTA GGA TTA TAT GGA CTT GAA GCC   90
   1  Met Lys Val Leu Leu Leu Lys Asp Ala Lys Glu Asp Asp Cys Gly Gln Asp Pro Tyr Ile Arg Glu Leu Gly Leu Tyr Gly Leu Glu Ala   30
N-Ter ----------------------------------------------------- X ----------------------------------------------------------------
                                                                                      T-20 ---------------------------------
  91  ACT TTG ATC CCT GTT TTA TCG TTT GAG TTT TTG TCT CTT CCC AGT TTC TCT GAG AAG CTT TCT CAT CCT GAA GAT TAC GGG GGA CTC ATT  180
  31  Thr Leu Ile Pro Val Leu Ser Phe Glu Phe Leu Ser Leu Pro Ser Phe Ser Glu Lys Leu Ser His Pro Glu Asp Tyr Gly Gly Leu Ile   60
      -------------------------------- Ala Thr ---                                     T-9 ----------------------------------------
      -X - X  --
 181  TTT ACC AGC CCC AGA GCA GTG GAA GCA GCA GAG TTA TGT TTG GAG CAA AAC AAT AAA ACT GAA GTC TGG GAA AGG TCT CTG AAA GAA AAA  270
  61  Phe Thr Ser Pro Arg Ala Val Glu Ala Ala Glu Leu Cys Leu Glu Gln Asn Asn Lys Thr Glu Val Trp Glu Arg Ser Leu Lys Glu Lys   90
      ---------------
 271  TGG AAT GCC AAG TCA GTG TAT GTG GTT GGA AAT GCT ACT GCT TCT CTA GTG AGT AAA ATT GGC CTG GAT ACA GAA GGA GAA ACC TGT GGA  360
  91  Trp Asn Ala Lys Ser Val Tyr Val Val Gly Asn Ala Thr Ala Ser Leu Val Ser Lys Ile Gly Leu Asp Thr Glu Gly Glu Thr Cys Gly  120

 361  AAT GCA GAA AAG CTT GCA GAA TAT ATT TGT TCC AGG GAG TCC TCA GCA CTG CCT CTT CTA TTT CCC TGT GGA AAC CTC AAA AGA GAA ATC  450
 121  Asn Ala Glu Lys Leu Ala Glu Tyr Ile Cys Ser Arg Glu Ser Ser Ala Leu Pro Leu Leu Phe Pro Cys Gly Asn Leu Lys Arg Glu Ile  150

 451  CTG CCA AAA GCG CTC AAG GAC AAA GGG ATT GCC ATG GAA AGC ATA ACT GTG TAT CAG ACA GTT GCA CAC CCA GGA ATC CAA GGG AAC CTG  540
 151  Leu Pro Lys Ala Leu Lys Asp Lys Gly Ile Ala Met Glu Ser Ile Thr Val Tyr Gln Thr Val Ala His Pro Gly Ile Gln Gly Asn Leu  180
                                T-15 ------------------------------------------------------- X ----------- X --
 541  AAC AGC TAC TAT TCC CAG CAG GGG GTT CCA GCC AGC ATC ACA TTT TTT AGT CCC TCT GGC CTC ACA TAC AGT CTC AAG CAC ATT CAG GAG  630
 181  Asn Ser Tyr Tyr Ser Gln Gln Gly Val Pro Ala Ser Ile Thr Phe Phe Ser Pro Ser Gly Leu Thr Tyr Ser Leu Lys His Ile Gln Glu  210

 631  TTA TCT GGT GAC AAT ATC GAT CAA ATT AAG TTT GCA GCC ATC GGC CCC ACT ACG GCT CGC GCG CTG GCC GCC CAG GGC CTT CCT GTA AGC  720
 211  Leu Ser Gly Asp Asn Ile Asp Gln Ile Lys Phe Ala Ala Ile Gly Pro Thr Thr Ala Arg Ala Leu Ala Ala Gln Gly Leu Pro Val Ser  240
           Pst I                      T-5 ----------------------------------------
 721  TGC ACT GCA GAG AGC CCC ACG CCA CAA GCC CTG GCC ACT GGC ATC AGG AAG GCT CTC CAG CCC CAT GGC TGC TGC TGA GTCAGCCACC TAGC  812
 241  Cys Thr Ala Glu Ser Pro Thr Pro Gln Ala Leu Ala Thr Gly Ile Arg Lys Ala Leu Gln Pro His Gly Cys Cys Ter                  265

 813  GCTGGCCCCA TGCAGCCTCC CTGGGCTGGG CTGGCTCTGG ATGGAGCCAG GCATCGGCAA GGGCTCTCGG GAGCTGCTGC CGTCAGACTC CTGCCTCAAG CCTGAGTGG  921
                Ava II
 922  A AGCACCTGAG GACCGGGGAT CGGGACCTGA CCTGGGGCTG GCCTCAGGCC CACGTGCACG TGACTGCCCT CTGTGGAAGC CAGCTTAAAC CCTAGCCCTG TGAGAGC 1029

1030  TTC CTGTGCCCAG CAGGAAGGAA GTCAAATAAA CCACACTGAC TACCTGTGCT TAAAAAAAAA AAAAAAAA                                       1100
```

FIG. 3. Nucleotide and predicted amino acid sequences of the pUROS-2 cDNA insert containing the complete coding region for human URO-synthase. Amino acid 1 is the N-terminal residue. Bold underlines indicate colinear amino acid sequence obtained by microsequencing of amino-terminal (N-Ter) and tryptic peptides (T) of the purified enzyme. Differences between microsequenced and predicted amino acids are indicated; X, unassigned amino acids. Overlines indicate the polyadenylation signal (AATAAA) and the pentanucleotide sequence (CACTG) recognized by the U4 small nuclear ribonucleoprotein. The positions of the *Ava* II and *Pst* I sites are indicated.

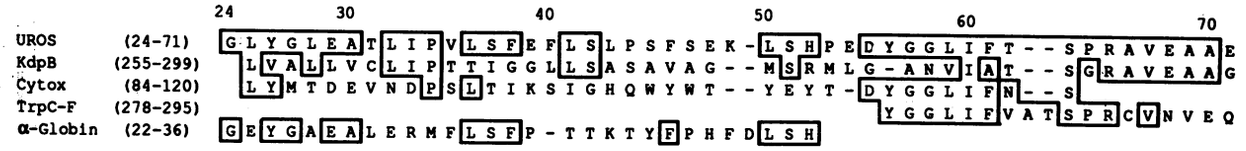

FIG. 4. URO-synthase amino acid sequence (single-letter code) similarities with other proteins. Human URO-synthase residues 24–71 were aligned with residues 255–299 of the potassium-transporting ATPase B chain (KdpB) from *E. coli* (27), residues 278–295 of *N*-(5′-phosphoribosyl)anthranilate isomerase from *E. coli* (TrpC-F) (28), residues 22–36 of human α-globin α-chain (α-Globin) (30), and residues 84–120 of the human mitochondrial cytochrome *c* oxidase polypeptide II (Cytox) (29). These regions were identified as significant when compared to random sequences by using the IDEAS program (20).

detection. Identical results were obtained after transformation with pKK223-3S-UROS (data not shown). Blunt-end ligation of the *Ava* II fragment into the *Kpn* I site of pBSKS (Fig. 5*B*) resulted in a construct, pBSKS-UROS, designed to express a fusion protein (32,085 Da) containing 19 residues of *E. coli* β-galactosidase and 12 residues encoded by the URO-synthase 5′ untranslated sequence. An immunoblot of the lysate from *E. coli* JM109 transformed with pBSKS-UROS revealed an expressed protein (with or without IPTG induction) with an apparent mass of ≈33 kDa, consistent with the presence of the additional 31 residues. As summarized in Table 1, all three vector constructs with the *Ava* II fragment in the sense orientation expressed high levels of active recombinant enzyme. Although the human enzyme was expressed without induction (presumably due to insufficient endogenous repressor synthesis by the different *E. coli* hosts), induction with IPTG increased the expression, most markedly of pKK223-3S-UROS.

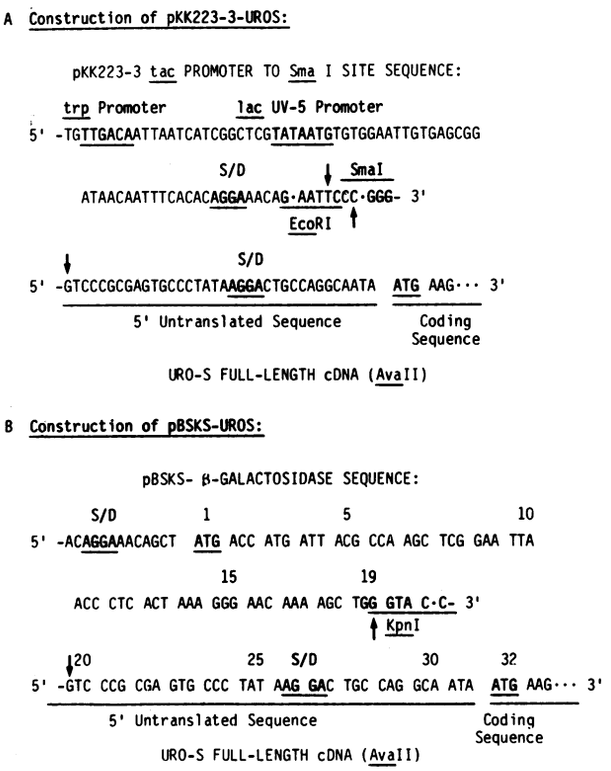

FIG. 5. Construction of plasmids pKK223-3-UROS and pBSKS-UROS. The 971-bp blunt-ended pUROS-*Ava* II fragment, which contains 36 bp (12 codons) of 5′ untranslated sequence including a Shine–Dalgarno (S/D) consensus sequence 13 bp from the coding sequence ATG, was blunt-end ligated to the indicated nucleotides (arrows) in pKK223-3 (*A*) and pBSKS (*B*). pKK223-3 has a *tac* promoter containing the *trp* and *lac* UV-5 promoter sequences and a S/D consensus sequence 5 and 13 bp from the *Eco*RI and *Sma* I ligation sites, respectively. The plasmid pBSKS has a S/D consensus sequence 7 bp from the initiation codon of the β-galactosidase gene.

DISCUSSION

A full-length human cDNA encoding URO-synthase was isolated by screening an adult liver cDNA library with synthetic oligonucleotide mixtures corresponding to microsequenced N-terminal and internal peptides from the purified enzyme. Authenticity of the full-length cDNA was established by colinearity (Fig. 3) and by microbial expression of active recombinant enzyme. Notably, the predicted N-terminal sequence of the liver cDNA was colinear with the N-terminal residues of the erythrocyte enzyme, consistent with the first ATG encoding the N-terminal residue and the absence of a precursor or signal peptide sequence in this cytosolic enzyme. In addition, the predicted amino acid composition and that of the purified erythrocyte enzyme were essentially identical (data not shown). These findings suggested that the hepatic and erythroid forms of the enzyme were the same. In contrast, erythroid and nonerythroid enzyme forms have been reported for human HMB-synthase (34, 35) and chicken δ-aminolevulinate synthase (36).

High levels of human URO-synthase activity were expressed in transformed bacteria either as the native protein (40.9–64.2 units per mg of protein) or as a fusion protein (1210 units per mg of protein; Fig. 6). For comparison, the mean activity of URO-synthase in human erythrocyte lysates was 7.5 units per mg of protein (9). In the pKK223-3S-UROS and pKK223-3E-UROS constructs, which have the strong *tac* promoter, the Shine–Dalgarno tetranucleotide consensus sequence (AGGA) was 48 and 45 bp, respectively, upstream from the URO-synthase initiation codon. Since the optimal location of the ribosome binding site is 5–9 nt from the initiation codon (37), it is likely that both plasmids used the endogenous Shine–Dalgarno sequence in the URO-synthase

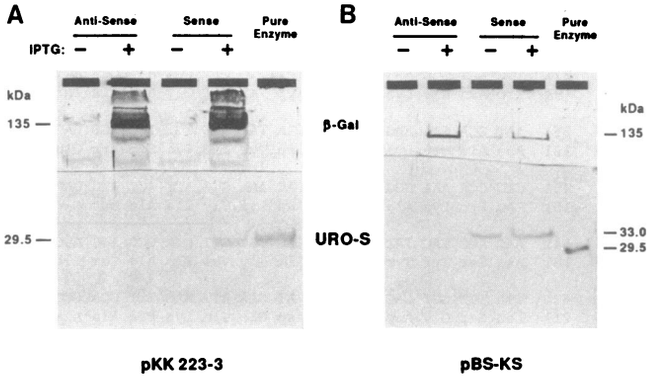

FIG. 6. Immunoblot demonstrating expression of human URO-synthase by *E. coli* transformed with pKK223-3E-UROS (*A*) and pBSKS-UROS (*B*). *E. coli* transformed with the pKK223-3E-UROS and pBSKS-UROS in the sense and antisense orientation (see Fig. 5) were grown in the presence or absence of IPTG; lysates were electrophoresed on 12.5% $NaDodSO_4$/polyacrylamide gel and transferred to nitrocellulose, and the filter was cut to stain separately for *E. coli* β-galactosidase (β-Gal) with rabbit anti-β-galactosidase and for URO-synthase (UROS) with goat anti-human URO-synthase antibodies.

Table 1. Expression of human URO-synthase by transformed *E. coli*

Plasmid	Specific activity,* units per mg of protein	
	Uninduced	Induced†
pKK223-3	2.23	1.55
pKK223-3E-UROS		
Antisense	1.97	1.64
Sense	13.70	64.20
pKK223-3S-UROS		
Antisense	1.64	1.83
Sense	3.91	40.90
pBSKS	3.52	3.70
pBSKS-UROS		
Antisense	4.30	3.77
Sense	913.0	1210.0

*Mean activity of four independent clones determined in duplicate by the coupled-enzyme assay (9).
†Induced with 5 mM IPTG.

5′ untranslated sequence, which was 13 nt from the initiation codon. The 20-fold higher expression of the fusion protein may have resulted from enhanced translation due to the more efficient β-galactosidase ribosome binding site as well as the higher replication rate of the plasmid and the greater stability of the fusion protein (38). Although these plasmid constructs produced high levels of enzyme, elimination of the 5′ untranslated sequence should facilitate the expression and isolation of even larger amounts of native enzyme for investigation of the enzyme's precise reaction mechanism and the nature of the interaction between URO-synthase and HMB-synthase.

Computer searches of nucleic acid and protein data bases did not reveal sequence similarities with the four other recently cloned human heme biosynthetic enzymes (23–26), suggesting the absence of common functional domains. However, short regions with significant amino acid sequence similarity (20) were identified in other proteins. These included a region of 48 residues in URO-synthase, which had similarity with sequences in *E. coli* *N*-(5′-phosphoribosyl)anthranilate isomerase, *E. coli* potassium-transporting ATPase B chain, human cytochrome *c* oxidase peptide II, and human globin α-chain (Fig. 4). Within this region there was a highly conserved sequence (Asp-Tyr-Gly-Gly-Leu-Ile-Phe-Thr-Ser-Pro-Arg-Ala-Val-Glu-Ala-Ala). The *E. coli* enzymes are not hemoproteins nor are they known to be functionally related; however, the *N*-(5′-phosphoribosyl)anthranilate isomerase is responsible for rearrangement of the ribose moiety and indole ring closure in tryptophan biosynthesis (28). In addition, the region of similarity in the human globin α-chain is involved in binding the D ring of heme (39), the pyrrole group in HMB which undergoes rearrangement and ring closure by URO-synthase.

In conclusion, the availability of the full-length cDNA for human URO-synthase should facilitate the localization and isolation of the chromosomal gene for investigation of its structure and tissue-specific regulation and for the identification and characterization of the molecular lesions that cause congenital erythropoietic porphyria. In addition, the ability to express large amounts of URO-synthase will permit structure–function analyses by site-specific mutagenesis.

Note Added in Proof. The recently reported sequence for the *E. coli* *hemD* (i.e., URO-synthase) gene (40, 41) had 44% nucleotide (with 42 gaps of 1–4 nt) and 22% amino acid (with 14 gaps of 1–3 residues) sequence homology with human URO-synthase.

We thank Dr. Kenneth Williams (Yale University School of Medicine) for amino acid microsequencing; Mr. Raman Reddy, Ms. Lily Kang, and Mr. Kuni Nishino for technical assistance; and Ms. Jackie Bilski for clerical assistance. This work was supported in part by Grants 5 R01 DK26824 and 1 U41 RR01685 from the National Institutes of Health.

1. Bogorad, L. & Granick, S. (1953) *Proc. Natl. Acad. Sci. USA* **39**, 1176–1188.
2. Bogorad, L. (1958) *J. Biol. Chem.* **233**, 501–515.
3. Batlle, A. M. d. C. & Rossetti, M. V. (1977) *Int. J. Biochem.* **8**, 251–267.
4. Battersby, A. R., Fookes, C. J. R., Matcham, G. W. J. & McDonald, E. (1980) *Nature (London)* **285**, 17–21.
5. Battersby, A. R., Fookes, C. J. R., Gustafson-Potter, K. E., McDonald, E. & Matcham, G. W. J. (1982) *J. Chem. Soc. Perkin Trans. 1*, 2427–2444.
6. Hart, G. J. & Battersby, A. R. (1985) *Biochem. J.* **232**, 151–160.
7. Tsai, S. F., Bishop, D. F. & Desnick, R. J. (1987) *J. Biol. Chem.* **262**, 1268–1273.
8. Kappas, A., Sassa, S. & Anderson, K. 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. 1325–1332.
9. Tsai, S. F., Bishop, D. F. & Desnick, R. J. (1987) *Anal. Biochem.* **166**, 120–133.
10. Allen, G. (1981) *Sequencing of Proteins and Peptides* (Elsevier, New York), pp. 53–54.
11. Hunkapiller, M. W. & Hood, L. E. (1983) *Science* **219**, 650–659.
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
13. Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G. & Colton, H. R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5661–5665.
14. Hanahan, D. & Meselson, M. (1980) *Gene* **10**, 63–67.
15. Birnboim, H. D. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
16. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Bilofsky, H. S., Burks, C., Fickett, J. W., Goad, W. B., Lewitter, F. I., Rindone, W. P., Swindell, C. D. & Tung, C. S. (1986) *Nucleic Acids Res.* **14**, 1–4.
19. Lipman, D. J. & Pearson, W. P. (1985) *Science* **227**, 1435–1441.
20. Kanehisa, M., Klein, P., Greif, P. & Delisi, C. (1984) *Nucleic Acids Res.* **12**, 417–428.
21. Wickens, M. & Stephenson, P. (1984) *Science* **226**, 1045–1051.
22. Berget, S. M. (1984) *Nature (London)* **309**, 179–182.
23. Bawden, M. J., Borthwick, I. A., Healy, H. M., Morris, C. P., May, B. K. & Elliott, W. H. (1987) *Nucleic Acids Res.* **15**, 8863.
24. Wetmur, J. G., Bishop, D. F., Cantelmo, C. & Desnick, R. J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7703–7707.
25. Raich, N., Romeo, P.-H., Dubart, A., Beaupain, D., Cohen-Solal, M. & Goossens, M. (1986) *Nucleic Acids. Res.* **14**, 5955–5968.
26. Romeo, P.-H., Raich, N., Dubart, A., Beaupain, D., Pryor, M., Kushner, J., Cohen-Solal, M. & Goossens, M. (1986) *J. Biol. Chem.* **261**, 9825–9831.
27. Hesse, J. E., Wieczorek, L., Altendorf, K., Reicin, A. S., Dorus, E. & Epstein, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4746–4750.
28. Priestle, J. P., Grütter, M. G., White, J. L., Vincent, M. G., Kania, M., Wilson, E., Jardetzky, T. S., Kirschner, K. & Jansonius, J. N. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5690–5694.
29. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1981) *Nature (London)* **290**, 457–465.
30. Michelson, A. M. & Orkin, S. H. (1980) *Cell* **22**, 371–377.
31. Heidekamp, F., Dirkse, W. G., Hille, J. & van Ormondt, H. (1983) *Nucleic Acids Res.* **11**, 6211–6223.
32. Buikema, W. J., Szeto, W. W., Lemley, P. V., Orme-Johnson, W. H. & Ausubel, F. M. (1985) *Nucleic Acids Res.* **13**, 4539–4555.
33. de Boer, H. A., Comstock, L. J. & Vasser, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 21–25.
34. Grandchamp, B., de Verneuil, H., Beaumont, C., Chretien, S., Walter, O. & Nordmann, Y. (1987) *Eur. J. Biochem.* **162**, 105–110.
35. Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M., & Romeo, P.-H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6–10.
36. Yamamoto, M., Yew, N. S., Federspiel, M., Dodgson, J. B., Hayashi, N. & Engel, J. D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3702–3706.
37. Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–45.
38. Stanley, K. (1983) *Nucleic Acids Res.* **11**, 4077–4092.
39. Weissbluth, M. (1974) *Hemoglobin. Cooperativity and Electronic Properties*, (Springer, New York), p. 17.
40. Sasarman, A., Nepveu, A., Echelard, Y., Dymetryszyn, J., Drolet, M. & Goyer, C. (1987) *J. Bacteriol.* **169**, 4257–4262.
41. Jordan, P. M., Mgbeje, B. I. A., Thomas, S. D. & Alwan, A. F. (1988) *Biochem. J.* **249**, 613–616.