Nucleotide Sequence of a Full-Length Complementary DNA Clone and Amino Acid Sequence of Human Phenylalanine Hydroxylase†

Simon C. M. Kwok, Fred D. Ledley, Anthony G. DiLella, Kathryn J. H. Robson,‡ and Savio L. C. Woo*

Howard Hughes Medical Institute, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030
Received November 13, 1984

ABSTRACT: A full-length human phenylalanine hydroxylase complementary DNA (cDNA) clone was isolated from a human liver cDNA library, and the nucleotide sequence encoding the entire enzyme was determined. The cDNA clone contains an inserted DNA fragment of 2448 base pairs, including 19 base pairs of poly(A) at the 3′ end. The first methionine codon occurs at nucleotide position 223, followed by an open reading frame of 1353 base pairs, encoding 451 amino acids. Translation of the nucleotide sequence in the open reading frame predicts the amino acid sequence of human phenylalanine hydroxylase. The human protein shows a 96% amino acid sequence homology with the corresponding rat enzyme. The determination of the complete primary structure for phenylalanine hydroxylase represents the first among mixed-function oxidases.

Phenylalanine hydroxylase (phenylalanine 4-monooxygenase, EC 1.14.16.1) is a hepatic enzyme that catalyzes the irreversible conversion of phenylalanine to tyrosine. This enzyme is deficient in phenylketonuria, which is the most common inborn error of amino acid metabolism (Kaufman, 1976). Phenylalanine hydroxylase occupies a central place in aromatic amino acid metabolism. The hydroxylation of phenylalanine represents an important means of producing tyrosine and is the rate-limiting step in the catabolism of phenylalanine. Because of the rate-limiting nature of the hydroxylation reaction, phenylalanine hydroxylase is principally responsible for regulating the clearance of phenylalanine from the body and maintaining homeostasis of the body's phenylalanine pools. When this enzyme is deficient as it is in phenylketonuria, phenylalanine accumulates in excess, leading to hyperphenylalaninemia and widespread abnormalities in the metabolism of many compounds, primarily those derived from aromatic or indole amino acids (Blau, 1979). If untreated, these metabolic abnormalities cause severe postnatal brain damage and mental retardation (Blau, 1979). Phenylketonuria is transmitted as an autosomal recessive trait, with an average prevalence of about 1/10000 among Caucasians (Scriver & Clow, 1980).

Phenylalanine hydroxylase has been purified from livers of rat, monkey, and human (Kaufman & Fisher, 1970; Cotton & Grattan, 1975; Woo et al., 1974; Friedman & Kaufman, 1973; Abita et al., 1983), and much has been learned about its structure and activity. The active enzyme has a molecular weight of about 100000 and is a dimer comprised of subunits whose individual molecular weights are 49000–52000. Though there has been considerable uncertainty about the identity of the subunits, there is evidence that in the rat the enzyme is composed of identical monomers. The rat enzyme is known to be a substrate for cAMP-dependent protein kinase, and monomers can be purified, revealing both phosphorylated (1 mol/mol) and dephosphorylated forms (Donlon & Kaufman, 1980). The holoenzyme is a metalloprotein containing 1 mol of iron/mol of subunit and requires the cofactor tetrahydrobiopterin for activity. However, neither the primary structure of phenylalanine hydroxylase nor the molecular basis for phenylketonuria has been established.

The development of recombinant DNA techniques enables us to study the primary structure of the enzyme and the molecular basis for phenylketonuria at the gene level. We have previously reported the purification of phenylalanine hydroxylase mRNA[1] from rat liver by polysome immunoprecipitation and the cloning of its cDNA (Robson et al., 1982), as well as the identification of several corresponding cDNA clones from a human liver cDNA library (Woo et al., 1983). Analysis of these human cDNA clones revealed that they did not comprise the entire mRNA molecule, which is about 2.5 kb in length. We report here the molecular cloning and nucleotide sequence determination of a full-length human phenylalanine hydroxylase cDNA. We also present the complete amino acid sequence of the human enzyme deduced from the nucleotide sequence, which has not yet been determined by conventional amino acid sequencing of the purified enzyme.

Experimental Procedures

Enzymes and Reagents. Restriction endonucleases and T_4 DNA ligase were purchased from New England Biolabs or Bethesda Research Laboratories. *Escherichia coli* DNA polymerase I (Klenow fragment) and calf alkaline phosphatase were products of Boehringer Mannheim. T_4 polynucleotide kinase and the M13 cloning/sequence kit were obtained from P-L Biochemicals. [α-^{32}P]dNTP and [α-^{32}P]ATP (specific radioactivity >3000 Ci/mmol) were purchased from Amersham Corp.

Screening of a λgt11 Human Liver cDNA Library. Our laboratory has recently constructed a human liver cDNA library of 14 million primary recombinants using the λgt11 vector system of Young & Davis (1983). About 100000 recombinant phage plaques were screened with a partial-length

† This work was partially supported by an NIH grant (HD-17711) and a March of Dimes grant (1-566) to S.L.C.W., who is also an Investigator of the Howard Hughes Medical Institute, and by an NIH research fellowship award (HD-06495) to A.G.D.

‡ Present address: MRC Unit of Molecular Haematology, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 7DU, U.K.

[1] Abbreviations: mRNA, messenger ribonucleic acid; cDNA, complementary deoxyribonucleic acid; bp, base pair(s); kb, kilobase(s); EDTA, ethylenediaminetetraacetic acid; SSC, 150 mM NaCl and 15 mM sodium citrate, pH 7.0; SDS, sodium dodecyl sulfate; SM buffer, 10 mM $MgSO_4$, 50 mM Tris-HCl, pH 7.4, and 100 mM NaCl; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.

0006-2960/85/0424-0556$01.50/0

human phenylalanine hydroxylase cDNA clone phPAH72 (Woo et al., 1983) as the hybridization probe, according to the phage amplification method of Woo (1979). The positive clones were picked in plugs containing three to ten independent plaques, and phage DNA was isolated for characterization by using a modified method of Cameron et al. (1977). Phage stocks were prepared by plating the mixed population of phage at high density on 100-mm plates and eluting phage into 5 mL of SM buffer. Phage DNA was prepared by sequential treatments with 1 μL of diethyl oxydiformate (room temperature), 10 μL of 10% SDS, and then 50 μL of 2 M Tris-HCl and 0.2 M EDTA (pH 8.5) at 70 °C for 5 min. Sediment was removed by centrifugation after the addition of 50 μL of 5 M potassium acetate and incubation on ice for 30 min. The supernatant was decanted and DNA precipitated in ethanol. The resulting DNA pellet was redissolved in TE (10 mM Tris-HCl, pH 8, and 1 mM EDTA), extracted twice with phenol saturated with TE and twice with ether, and precipitated with 2 volumes of ethanol. DNA from the miniphage lysate was digested with *Eco*RI, electrophoresed on a 0.8% agarose gel, transferred to a nitrocellulose filter, and hybridized to a ^{32}P-labeled probe consisting of the insert from phPAH72. Filters were prehybridized in 45% formamide, 4× SSC, 0.1 M sodium phosphate (pH 6.5), 0.1% sodium pyrophosphate, 0.1% SDS, 5× Denhardt's solution (Denhardt, 1966), and 250 μg/mL sheared salmon sperm DNA at 42 °C for 6 h. Hybridization was performed in 45% formamide, 4× SSC, 0.075% sodium pyrophosphate, 0.1% SDS, 1× Denhardt's solution, 100 μg/mL sheared salmon sperm DNA, and 10% dextran sulfate at 42 °C for 12–16 h. Filters were washed in 2× SSC and 0.1% SDS for 1 h at room temperature and then for 3–4 h at 68 °C. Phages containing the longest inserts (2–2.4 kb) were subsequently plaque purified.

DNA Sequence Analysis. The chemical modification method of Maxam & Gilbert (1980) was used for most of the DNA sequence determination. The "dideoxy" DNA sequencing method of Sanger et al. (1977) was also used for two small *Rsa*I fragments. For the chemical modification method, the longest human phenylalanine hydroxylase cDNA insert was subcloned into the *Eco*RI site of pBR322. The plasmid was digested with the appropriate restriction endonuclease and either 5′ end labeled with polynucleotide kinase and [γ-^{32}P]ATP or 3′ end labeled with DNA polymerase I (Klenow fragment) and [α-^{32}P]dNTP. After digestion with a second restriction enzyme, single end-labeled fragments then were isolated from 5% polyacrylamide gels for DNA sequence analysis. For the dideoxy sequencing method, the *Rsa*I fragments of the cDNA insert were subcloned into the *Sma*I site of the M13 vector. Subsequent procedures were those described by Messing (1983).

Computer Analysis. Dot-matrix sequence analysis was performed on an IBM PC/XT with a program provided by Zweig (1984). Data bank homology searches were performed with the Atlas of Protein Sequence and Structure Data Bank.

Results

Isolation of a Full-Length Human Phenylalanine Hydroxylase cDNA Clone. A library of human liver cDNA was constructed in the cloning vector λgt11 kindly provided to us by Drs. Richard Young and Ron Davis of Stanford by using polyadenylated RNA from a fresh human liver of a renal donor. Phenylalanine hydroxylase mRNA is thought to constitute approximately 0.1% of the total cellular mRNA of the normal human liver. Approximately 100 000 recombinants of the library were screened with a partial-length human phenylalanine hydroxylase cDNA clone as the hybridization probe. About 100 positive signals were obtained on the primary screen, corresponding to the expected abundance of phenylalanine hydroxylase mRNA in the human liver. The positive clones were picked from the primary plates in plugs that contained approximately three to ten individuals' plaques. Miniphage lysates were prepared from the mixed population of phages, followed by *Eco*RI digestion and Southern hybridization, in order to determine the length of the cDNA insert in each of the positive clones (Figure 1). Of the 50 independent miniphage preparations examined, 4 contained cDNA inserts greater than 2 kb in length that hybridized to the human phenylalanine hydroxylase cDNA probe. The longest cDNA insert was estimated to be approximately 2500 bp in length (Figure 1, lane h). This size corresponded roughly to the length of the phenylalanine hydroxylase mRNA estimated previously from Northern blot analysis of human liver poly(A)-containing RNA.

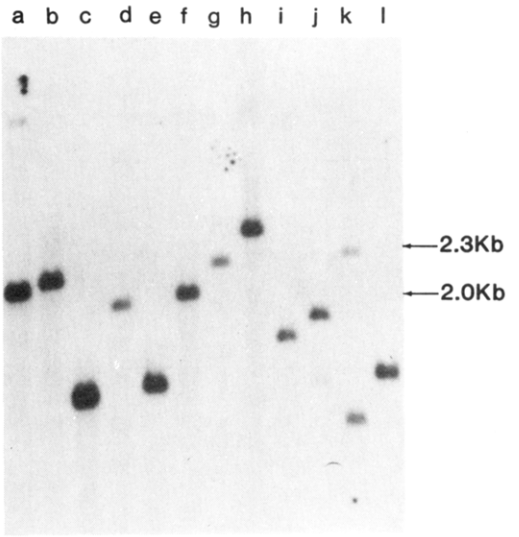

FIGURE 1: Southern blot analysis of λgt11 clones containing human phenylalanine hydroxylase cDNA. Phage DNA was digested with *Eco*RI, which separates the phage arms from the cDNA insert. Most of the clones contain incomplete cDNAs. The longest clone (lane h) corresponded to the length of phenylalanine hydroxylase mRNA and represents a full-length human phenylalanine hydroxylase cDNA clone.

The four clones containing inserts greater than 2 kb were plaque purified, and the inserted *Eco*RI fragments were subcloned from λgt11 into the *Eco*RI site of plasmid pBR322. They range in size from 2000 to 2500 bp and contain several identical internal restriction sites but differ in their 5′ and 3′ extents (data not shown). The possibility that the phage clones might contain additional internal *Eco*RI fragments was ruled out by end labeling *Eco*RI-digested phage DNA and visualizing the radiolabeled products on polyacrylamide gels capable of resolving fragments as small as 5 bases in length. The only fragments visualized in this experiment corresponded to the 2–2.5-kb insert and the phage arms (data not shown).

Nucleotide Sequence Analysis of a Full-Length Human Phenylalanine Hydroxylase cDNA Clone. The plasmid containing the longest phenylalanine hydroxylase cDNA insert, designated phPAH247, was further characterized by restriction endonuclease mapping and found to contain 700 bp of additional 5′ sequence that was not present in clones obtained previously (Figure 2). The presence of several unique restriction endonuclease sites, such as *Xba*I, *Bgl*II, *Xho*I, and *Bam*HI, as well as two internal *Hin*dIII and *Nco*I sites, facilitated nucleotide sequence determination by the method of Maxam and Gilbert, and the sequencing strategy is shown in

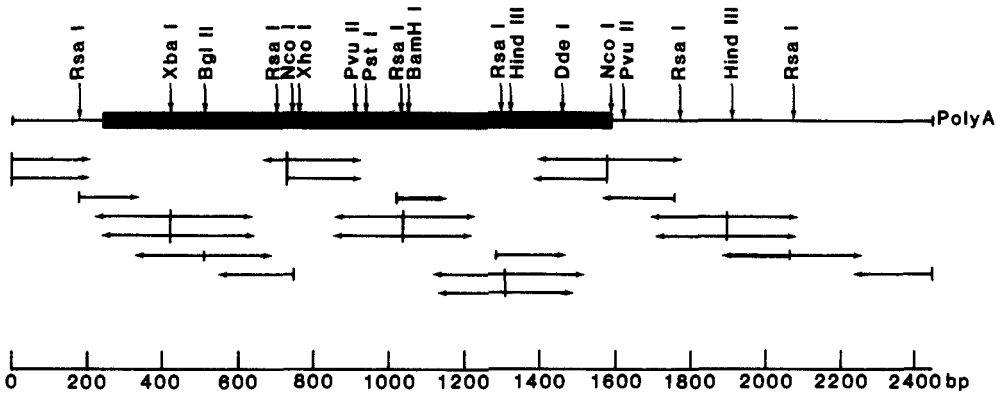

FIGURE 2: Restriction endonuclease map and strategy for nucleotide sequence analysis of human phenylalanine hydroxylase cDNA. The insert DNA in phPAH247 is presented in the 5′ to 3′ orientation. The long solid box represents the coding region, and the solid lines represent the untranslated regions. The horizontal arrows indicate the direction and extent of sequence analyses by the procedures of Maxam & Gilbert (1980) and Messing (1983).

Figure 2. Both strands of the entire cDNA insert were completely sequenced in order to ensure accuracy of the sequence.

The completed sequence of inserted cDNA in phPAH247 is 2448 bp in length, excluding 19 bp of poly(A) tail at the 3′ end (Figure 3). A continuous open reading frame is present, starting from the first ATG at position 223 and ending at the TAA codon at position 1579. Although the N-terminal amino acid sequence of human or rat phenylalanine hydroxylase has never been established, it can be assumed that the ATG at position 223 must be the initiation codon for methionine because there are termination codons in all three reading frames within the 5′ untranslated region. Thus the coding region is 1353 bp in length and codes for 451 amino acid residues. The 5′ untranslated region is in excess of 200 bp and at position 105–114 contains an oligonucleotide sequence (UCGUUACCGC) with partial homology to the 3′ end of the 18S ribosomal RNA (Hagenbuchle et al., 1978). The overall homology of the alignment is 80%. This sequence may function as a recognition site for binding to the 18S ribosomal RNA during translation. This region also includes two inverted repeat sequences of 13 bp with 77% homology at positions 44–56 and 108–120 (Figure 3), which could be involved in a stable stem–loop structure. The 3′ untranslated region is 851 bp long, with many termination codons in all reading frames. It also contains three AATAAA sequences at positions 2198, 2347, and 2413. The distance between the last AATAAA sequence and the poly(A) tail is 13 bp and is apparently used for RNA processing and polyadenylation normally (Proudfoot & Brownlee, 1976).

Amino Acid Sequence of Human Phenylalanine Hydroxylase. Translation of the nucleotide sequence in the coding region of the cDNA predicts the primary structure of human phenylalanine hydroxylase (Figure 3). The deduced amino acid sequence constitutes a protein of 51 672 daltons, which is in accordance with the reported values for the subunit molecular weight of the enzyme (Woo et al., 1974; Friedman & Kaufman, 1973; Abita et al., 1983). The predicted amino acid composition from the sequence data agrees very well with that reported for the corresponding rat enzyme (Shiman, 1980) but differs somewhat from the composition of the human enzyme recently reported by Abita et al. (1983), particularly in the contents of methionine residues (Table I).

Rat liver phenylalanine hydroxylase is a phosphoprotein and its phosphorylation site has an amino acid sequence of Ser-Arg-Lys-Leu-[^{32}P]SerP-Asx-Phe-Gly-Glx-Glx (Wretborn et al., 1980). An identical sequence is found in the deduced amino acid sequence of the human enzyme at amino acid residues 11–20 from the amino terminus (Figure 3), with the exception that the first serine residue in the rat enzyme fragment is replaced by a glycine residue in the human enzyme.

Table I: Comparison of Amino Acid Composition of Human and Rat Liver Phenylalanine Hydroxylase

amino acid residues	residues per mole of subunit		
	human[a]	human[b]	rat[c]
Lys	30	33.1	27.8
His	13	10.7	10.8
Arg	24	26.2	24.4
Asp + Asn	40	49.9	37.6
Thr	24	27.8	21.9
Ser	28	29.5	29.1
Glu + Gln	56	61.1	51.9
Pro	23	27.2	19.6
Gly	25	43.3	24.1
Ala	28	46.2	27.2
cysteic acid	9	8.7	7.7
Val	19	36.9	21.0
Met	2	10.8	0.9
Ile	28	20.7	23.5
Leu	50	41.2	44.3
Tyr	22	14.9	20.8
Phe	27	20.9	24.8
Trp	3	ND	4

[a] Deduced from the nucleotide sequence of phPAH247. [b] From Abita et al. (1983). [c] From Shiman (1980).

Sequence Homology with Rat Phenylalanine Hydroxylase. Comparison of the human phenylalanine hydroxylase cDNA sequence with those of partial rat cDNA clones containing the 3′ half of the mRNA sequnce (Robson et al., 1984) using a dot-matrix computer program (Zweig, 1984) showed that the two sequences share considerable homology in the coding regions up to where the termination codons appear (data not shown). This finding is in general agreement with other eukaryotic genes in that the 3′ and 5′ untranslated regions have higher rates of sequence divergence among different species than the coding regions. Within the coding regions compared, sequence homology is 89% at the nucleotide level and 96% at the amino acid level. In fact, 82% of the nucleotide substitutions are silent mutations that do not affect the amino acid sequence. Only 10 out of the 245 amino acid residues compared in these two sequences are different, and most of the amino acid replacements are conservative substitutions (Table II). One unusual substitution is the isoleucine residue at position 373 of human phenylalanine hydroxylase that was replaced by a cysteine residue in the rat enzyme.

Discussion

The present work describes the isolation and sequence

```
                15                  30                  45                  60                  75
CTT CAT CGT CGT CCA ACT GAC CTT GAG TGT TAG TTT CGC GGT AAG TTT GGG TAT AAG TGC CAC CAC CAG TGC CGG
                90                 105                 120                 135                 150
CAG TGT AGT CAG TAG TTT GTT GCT GGA AGT CGT TAC CGC CAA ACT GCG TGT TAC CGC CTA GAT TAG ACT GTG GCT
               165                 180                 195                 210                 225
GCT GGC GTT GAG GGA AAC CTG CCT GTA CGT GAG GCC CTA AAA AGC CAG AGA CCT CAC TCC CGG GGA GCC AGC ATG
                                                                                               Met
               240                 255                 270                 285                 300
TCC ACT GCG GTC CTG GAA AAC CCA GGC TTG GGC AGG AAA CTC TCT GAC TTT GGA CAG GAA ACA AGC TAT ATT GAA
Ser Thr Ala Val Leu Glu Asn Pro Gly Leu Gly Arg Lys Leu Ser Asp Phe Gly Gln Glu Thr Ser Tyr Ile Glu
               315                 330                 345                 360                 375
GAC AAC TGC AAT CAA AAT GGT GCC ATA TCA CTG ATC TTC TCA CTC AAA GAA GAA GTT GGT GCA TTG GCC AAA GTA
Asp Asn Cys Asn Gln Asn Gly Ala Ile Ser Leu Ile Phe Ser Leu Lys Glu Glu Val Gly Ala Leu Ala Lys Val
               390                 405                 420                 435                 450
TTG CGC TTA TTT GAG GAG AAT GAT GTA AAC CTG ACC CAC ATT GAA TCT AGA CCT TCT CGT TTA AAG AAA GAT GAG
Leu Arg Leu Phe Glu Glu Asn Asp Val Asn Leu Thr His Ile Glu Ser Arg Pro Ser Arg Leu Lys Lys Asp Glu
               465                 480                 495                 510                 525
TAT GAA TTT TTC ACC CAT TTG GAT AAA CGT AGC CTG CCT GCT CTG ACA AAC ATC ATC AAG ATC TTG AGG CAT GAC
Tyr Glu Phe Phe Thr His Leu Asp Lys Arg Ser Leu Pro Ala Leu Thr Asn Ile Ile Lys Ile Leu Arg His Asp
               540                 555                 570                 585                 600
ATT GGT GCC ACT GTC CAT GAG CTT TCA CGA GAT AAG AAG AAA GAC ACA GTG CCC TGG TTC CCA AGA ACC ATT CAA
Ile Gly Ala Thr Val His Glu Leu Ser Arg Asp Lys Lys Lys Asp Thr Val Pro Trp Phe Pro Arg Thr Ile Gln
               615                 630                 645                 660                 675
GAG CTG GAC AGA TTT GCC AAT CAG ATT CTC AGC TAT GGA GCG GAA CTG GAT GCT GAC CAC CCT GGT TTT AAA GAT
Glu Leu Asp Arg Phe Ala Asn Gln Ile Leu Ser Tyr Gly Ala Glu Leu Asp Ala Asp His Pro Gly Phe Lys Asp
               690                 705                 720                 735                 750
CCT GTG TAC CGT GCA AGA CGG AAG CAG TTT GCT GAC ATT GCC TAC AAC TAC CGC CAT GGG CAG CCC ATC CCT CGA
Pro Val Tyr Arg Ala Arg Arg Lys Gln Phe Ala Asp Ile Ala Tyr Asn Tyr Arg His Gly Gln Pro Ile Pro Arg
               765                 780                 795                 810                 825
GTG GAA TAC ATG GAG GAA GAA AAG AAA ACA TGG GGC ACA GTG TTC AAG ACT CTG AAG TCC TTG TAT AAA ACC CAT
Val Glu Tyr Met Glu Glu Glu Lys Lys Thr Trp Gly Thr Val Phe Lys Thr Leu Lys Ser Leu Tyr Lys Thr His
               840                 855                 870                 885                 900
GCT TGC TAT GAG TAC AAT CAC ATT TTT CCA CTT CTT GAA AAG TAC TGT GGC TTC CAT GAA GAT AAC ATT CCC CAG
Ala Cys Tyr Glu Tyr Asn His Ile Phe Pro Leu Leu Glu Lys Tyr Cys Gly Phe His Glu Asp Asn Ile Pro Gln
               915                 930                 945                 960                 975
CTG GAA GAC GTT TCT CAA TTC CTG CAG ACT TGC ACT GGT TTC CGC CTC CGA CCT GTG GCT GGC CTG CTT TCC TCT
Leu Glu Asp Val Ser Gln Phe Leu Gln Thr Cys Thr Gly Phe Arg Leu Arg Pro Val Ala Gly Leu Leu Ser Ser
               990                1005                1020                1035                1050
CGG GAT TTC TTG GGT GGC CTG GCC TTC CGA GTC TTC CAC TGC ACA CAG TAC ATC AGA CAT GGA TCC AAG CCC ATG
Arg Asp Phe Leu Gly Gly Leu Ala Phe Arg Val Phe His Cys Thr Gln Tyr Ile Arg His Gly Ser Lys Pro Met
              1065                1085                1095                1110                1125
TAT ACC CCC GAA CCT GAC ATC TGC CAT GAG CTG TTG GGA CAT GTG CCC TTG TTT TCA GAT CGC AGC TTT GCC CAG
Tyr Thr Pro Glu Pro Asp Ile Cys His Glu Leu Leu Gly His Val Pro Leu Phe Ser Asp Arg Ser Phe Ala Gln
              1140                1155                1170                1185                1200
TTT TCC CAG GAA ATT GGC CTT GCC TCT CTG GGT GCA CCT GAT GAA TAC ATT GAA AAG CTC GCC ACA ATT TAC TGG
Phe Ser Gln Glu Ile Gly Leu Ala Ser Leu Gly Ala Pro Asp Glu Tyr Ile Glu Lys Leu Ala Thr Ile Tyr Trp
              1215                1230                1245                1260                1275
TTT ACT GTG GAG TTT GGG CTC TGC AAA CAA GGA GAC TCC ATA AAG GCA TAT GGT GCT GGG CTC CTG TCA TCC TTT
Phe Thr Val Glu Phe Gly Leu Cys Lys Gln Gly Asp Ser Ile Lys Ala Tyr Gly Ala Gly Leu Leu Ser Ser Phe
              1290                1305                1320                1335                1350
GGT GAA TTA CAG TAC TGC TTA TCA GAG AAG CCA AAG CTT CTC CCC CTG GAG CTG GAG AAG ACA GCC ATC CAA AAT
Gly Glu Leu Gln Tyr Cys Leu Ser Glu Lys Pro Lys Leu Leu Pro Leu Glu Leu Glu Lys Thr Ala Ile Gln Asn
              1365                1380                1395                1410                1425
TAC ACT GTC ACG GAG TTC CAG CCC CTG TAT TAC GTG GCA GAG AGT TTT AAT GAT GCC AAG GAG AAA GTA AGG AAC
Tyr Thr Val Thr Glu Phe Gln Pro Leu Tyr Tyr Val Ala Glu Ser Phe Asn Asp Ala Lys Glu Lys Val Arg Asn
              1440                1455                1470                1485                1500
TTT GCT GCC ACA ATA CCT CGG CCC TTC TCA GTT CGC TAC GAC CCA TAC ACC CAA AGG ATT GAG GTC TTG GAC AAT
Phe Ala Ala Thr Ile Pro Arg Pro Phe Ser Val Arg Tyr Asp Pro Tyr Thr Gln Arg Ile Glu Val Leu Asp Asn
              1515                1530                1545                1560                1575
ACC CAG CAG CTT AAG ATT TTG GCT GAT TCC ATT AAC AGT GAA ATT GGA ATC CTT TGC AGT GCC CTC CAG AAA ATA
Thr Gln Gln Leu Lys Ile Leu Ala Asp Ser Ile Asn Ser Glu Ile Gly Ile Leu Cys Ser Ala Leu Gln Lys Ile
              1590                1605                1620                1635                1650
AAG TAA AGC CAT GGA CAG AAT GTG GTC TGT CAG CTG TGA ATC TGT TGA TGG AGA TCC AAC TAT TTC TTT CAT CAG
Lys ***
              1665                1680                1695                1710                1725
AAA AAG TCC GAA AAG CAA ACC TTA ATT TGA AAT AAC AGC CTT AAA TCC TTT ACA AGA TGG AGA AAC AAC AAA TAA
              1740                1755                1770                1785                1800
GTC AAA ATA ATC TGA AAT GAC AGG ATA TGA GTA CAT ACT CAA GAG CAT AAT GGT AAA TCT TTT GGG GTC ATC TTT
              1815                1830                1845                1860                1875
GAT TTA GAG ATG ATA ATC CCA TAC TCT CAA TTG AGT TAA ATC AGT AAT CTG TCG CAT TTC ATC AAG ATT AAT TAA
              1890                1905                1920                1935                1950
AAT TTG GGA CCT GCT TCA TTC AAG CTT CAT ATA TGC TTT GCA GAG AAC TCA TAA AGG AGC ATA TAA GGC TAA ATG
              1965                1980                1995                2010                2025
TAA AAC ACA AGA CTG TCA TTA GAA TTG AAT TAT TGG GCT TAA TAT AAA TCG TAA CCT ATG AAG TTT ATT TTC TAT
              2040                2055                2070                2085                2100
TTT AGT TAA CTA TGA TTC CAA TTA CTA CTT TGT TAT TGT ACC TAA GTA AAT TTT CTT TAG GTC AGA AGC CCA TTA
              2115                2138                2145                2160                2175
AAA TAG TTA CAA GCA TTG AAC TTC TTT AGT ATT ATA TTA ATA TAA AAA CAT TTT TGT ATG TTT TAT TGT AAT CAT
              2190                2205                2220                2235                2250
AAA TAC TGC TGT ATA AGG TAA TAA AAC TCT GCA CCT AAT CCC CAT AAC TTC CAG TAT CAT TTT CCA ATT AAT TAT
              2265                2280                2295                2310                2325
CAA GTC TGT TTT GGG AAA CAC TTT GAG GAC ATT TAT GAT GCA GCA GAT GTT GAC TAA AGG CTT GGT TGG TAG ATA
              2340                2355                2370                2385                2400
TTC AGG AAA TGT TCA CTG AAT AAA TAA GTA AAT ACA TTA TTG AAA AGC AAA TCT GTA TAA ATG TGA AAT TTT TAT
              2415                2430                2445
TTG TAT TAG TAA TAA AAC ATT AGT AGT TTA AAA AAA AAA AAA AAA AAA
```

FIGURE 3: Nucleotide sequence of human phenylalanine hydroxylase cDNA. The complete sequence of clone phPAH247 and its translation into human phenylalanine hydroxylase are presented in the 5′ to 3′ direction. The underlined sequences in the 5′ untranslated region are the inverted repeat sequences that could be involved in a stable stem–loop structure. Also underlined are three AATAAA sequences at the end of the 3′ untranslated region, the last of which could be involved in transcription termination and/or polyadenylation.

characterization of a full-length human phenylalanine hydroxylase cDNA clone. The authenticity of the clone is established by two observations. First, the predicted amino acid sequence of human phenylalanine hydroxylase matches a

Table II: Amino Acid Differences between Human and Rat Phenylalanine Hydroxylase (PAH) in the Region of Residues 207–451

residue position	human PAH		rat PAH	
	amino acid	codon	amino acid	codon
119	His	CAU	Arg	CGU
335	Gln	CAA	Glu	GAA
373	Ile	AUC	Cys	UGC
375	Asn	AAU	Glu	GAG
377	Thr	ACU	Ser	UCU
400	Asn	AAC	Thr	ACU
420	Ile	AUU	Val	GUU
437	Asn	AAC	Ile	AUU
440	Ile	AUU	Val	GUC
445	Ser	AGU	Asn	AAU

fragment of the rat enzyme that contained the phosphorylation site. The homology to the phosphorylation site occurs at residues 11–20 in the amino terminus of the protein. Second, the nucleotide sequence of the human cDNA clone is 89% homologous to the rat cDNA sequence that specifically hybrid selected the mRNA coding for phenylalanine hydroxylase (Robson et al., 1982) and, more recently, was shown to contain an identical amino acid sequence with a cyanogen bromide fragment of the purified rat enzyme (Robson et al., 1984). The human enzyme also contains an identical amino acid sequence following the internal methionine codon at residue 275. This cDNA clone contains the entire coding information for phenylalanine hydroxylase, since preliminary gene transfer studies demonstrate that it is capable of expressing authentic phenylalanine hydroxylase activity in transformed mammalian cells (F. Ledley, unpublished observations).

The amino acid composition predicted by our sequence is congruous with that experimentally determined for the rat enzyme reported by several laboratories (Shiman, 1980; Fisher et al., 1972; Nakata & Fujisawa, 1980). It differs somewhat, however, from a single report for the human enzyme (Abita et al., 1983). It is difficult to reconcile the discrepancies between the amino acid composition of the human enzyme reported by Abita and co-workers with the deduced values from the nucleotide sequence of the cDNA clone, particularly in the number of methionine residues (Table I). Recently, Smith et al. (1984) have independently reported a much lower number of methionine residues in their purified human phenylalanine hydroxylase preparation, which is in closer agreement with our findings. Because of rapid degradation of the human enzyme after death, one possible reason for the discrepancy could be the contamination of the enzyme preparation used for amino acid composition analysis by Abita et al. (1983).

The degree and pattern of homology between the rat and human phenylalanine hydroxylase nucleotide sequences are consistent with the evolutionary distance between rat and humans. Homology between the nucleotide sequences is much more pronounced within the open reading frame (89%) than in the 3′ untranslated region where the degree of homology is only 35%. Moreover, within the open reading frames of the cDNA clones in the two species, the divergence in sequences occurs predominantly at the third position of codons leading to silent mutations, which do not alter the primary structure of the enzymes. The overall amino acid sequence homology between rat and human phenylalanine hydroxylase is 96%. Of particular interest is the preservation of the consensus phosphorylation site in the human sequence that retains the determinants common to substrates of the cAMP-dependent protein kinase (Williams, 1976; Matsuo et al., 1978). Although Abita et al. (1983) reported that their purified human phenylalanine hydroxylase could not be phosphorylated by cAMP-dependent protein kinase, recently Smith et al. (1984) showed that two apparent molecular weight forms of human and monkey phenylalanine hydroxylase can be interconverted by phosphorylation and dephosphorylation.

The primary structure of human phenylalanine hydroxylase represents the first reported among mixed-function oxidases. No homology was found between the deduced amino acid sequence of human phenylalanine hydroxylase and sequences compiled in the Atlas of Protein Sequence and Structure data base. The sequences of proteins with which phenylalanine hydroxylase shares considerable structural and functional attributes, such as tyrosine hydroxylase, tryptophan hydroxylase, and dopamine β-hydroxylase, have not yet been reported. These proteins may be expected to share considerable sequence homology at the amino acid and nucleotide levels on the basis of the results of peptide mapping, cross-immunoreactivity, and cross-hybridization of cDNA of one enzyme with mRNAs of the others (Robson et al., 1982; Joh et al., 1983; Chikaraishi et al., 1983). The molecular cloning of rat tyrosine hydroxylase (Lamouroux et al., 1982; Chikaraishi et al., 1983), dopamine β-hydroxylase (Joh et al., 1983; O'Malley et al., 1983), and phenylethanolamine *N*-methyltransferase (Baetge et al., 1983) has been reported. It would be of interest to study the structural similarities of these enzymes with phenylalanine hydroxylase when their nucleotide sequences become available.

The cloning of a full-length human phenylalanine hydroxylase cDNA presented in this paper has enabled a more detailed characterization of its chromosomal gene. Using this full-length cDNA clone as a hybridization probe, we were able to detect five restriction endonucleases that yielded restriction fragment length polymorphisms in the human phenylalanine hydroxylase locus (A. S. Lidsky, unpublished observation) in addition to the three reported previously (Woo et al., 1983). The additional restriction fragment length polymorphisms elevate the percent heterozygosity of the phenylalanine hydroxylase locus in the population such that about 90% of the phenylketonuric families are informative for prenatal diagnosis and carrier detection of the genetic disorder by gene analysis.

Acknowledgments

We thank Drs. Richard Young and Ron Davis for λgt11 and host bacterial strains used for preparing the human liver cDNA library, which was constructed by Drs. T. Chandra and V. Kidd of our laboratory. We also thank Judy Stanfield for her assistance in preparing the manuscript.

References

Abita, J. P., Blandin-Savoja, F., & Rey, F. (1983) *Biochem. Int. 7*, 727–737.

Baetge, E. E., Moon, H. M., Kaplan, B. B., Park, D. H., Reis, D. J., & Joh, T. H. (1983) *Neurochem. Int. 5*, 611–618.

Blau, K. (1979) in *Aromatic Amino Acid Hydroxylases and Mental Disease* (Youdim, M. B. H., Ed.) pp 79–139, Wiley, New York.

Cameron, J. R., Philippsen, P., & Davis, R. W. (1977) *Nucleic Acids Res. 4*, 1429–1448.

Chikaraishi, D. M., Brilliant, M. H., & Lewis, E. J. (1983) *Cold Spring Harbor Symp. Quant. Biol. 48*, 309–318.

Cotton, R. G. H., & Grattan, P. J. (1975) *Eur. J. Biochem. 60*, 427–430.

Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun. 23*, 641–652.

Donlon, J., & Kaufman, S. (1980) *J. Biol. Chem. 255*, 2146–2152.

Fisher, D. B., Kirkwood, R., & Kaufman, S. (1972) *J. Biol. Chem. 247*, 5161–5167.
Friedman, P. A., & Kaufman, S. (1973) *Biochim. Biophys. Acta 293*, 56–61.
Hagenbuchle, O., Santer, M., Steitz, J. A., & Mans, R. J. (1978) *Cell (Cambridge, Mass.) 13*, 551–563.
Joh, T. H., Baetge, E. E., Ross, M. E., & Reis, D. J. (1983) *Cold Spring Harbor Symp. Quant. Biol. 48*, 327–335.
Kaufman, S. (1976) *Adv. Neurochem. 2*, 1–132.
Kaufman, S., & Fisher, D. B. (1970) *J. Biol. Chem. 235*, 4745–4750.
Lamouroux, A., Biguet, N. F., Samolyk, D., Privat, A., Salomon, J. C., Pujol, J. F., & Mallet, J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3881–3885.
Matsuo, M., Huang, C., & Huang, L. C. (1978) *Biochem. J. 173*, 441–447.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
Messing, J. (1983) *Methods Enzymol. 101*, 20–78.
Nakata, H., & Fujisawa, H. (1980) *Biochim. Biophys. Acta 614*, 313–327.
O'Malley, K. L., Mauron, A., Raese, J., Barchas, J. D., & Kedes, L. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2161–2165.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.
Robson, K. J. H., Chandra, T., MacGillivray, R. T. A., & Woo, S. L. C. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4701–4705.
Robson, K. J. H., Beattie, W., James, R. J., Cotton, R. C. H., Morgan, F. J., & Woo, S. L. C. (1984) *Biochemistry 23*, 5671–5675.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Scriver, C. R., & Clow, C. R. (1980) *Annu. Rev. Genet. 14*, 179–202.
Shiman, R. (1980) *J. Biol. Chem. 255*, 10029–10032.
Smith, S. C., Kemp, B. E., McAdam, W. J., Mercer, J. F. B., & Cotton, R. G. H. (1984) *J. Biol. Chem. 259*, 11284–11289.
Williams, R. E. (1976) *Science (Washington, D.C). 192*, 473–474.
Woo, S. L. C. (1979) *Methods Enzymol. 68*, 389–395.
Woo, S. L. C., Gillam, S. S., & Woolf, L. I. (1974) *Biochem. J. 139*, 741–749.
Woo, S. L. C., Lidsky, A. S., Guttler, F., Chandra, T., & Robson, K. J. H. (1983) *Nature (London) 306*, 151–155.
Wretborn, M., Humble, E., Ragnarsson, U., & Engstrom, L. (1980) *Biochem. Biophys. Res. Commun. 93*, 403–408.
Young, R. A., & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1194–1198.
Zweig, S. E. (1984) *Nucleic Acids Res. 12*, 767–776.

Articles

Membrane Polypeptide in Rabbit Erythrocytes Associated with the Inhibition of L-Lactate Transport by a Synthetic Anhydride of Lactic Acid†

Jerald A. Donovan* and Michael L. Jennings

Department of Physiology and Biophysics, The University of Iowa, Iowa City, Iowa 52242
Received July 30, 1984

ABSTRACT: The synthetic lactyl anhydride isobutylcarbonyl lactyl anhydride (iBCLA), a selective and potent inhibitor of L-(+)-lactate transport in rabbit erythrocytes, reduces the chemical labeling of a 40–50-kdalton polypeptide by tritiated 4,4′-diisothiocyanato-2,2′-dihydrostilbenedisulfonate ([^{3}H]H_2DIDS). iBCLA does so in a dose-dependent manner at concentrations that strongly inhibit lactate exchange but not chloride–phosphate exchange. These labeling experiments and inhibition reversal studies using iBCLA, *p*-(chloromercuri)benzenesulfonic acid (pCMBS), and dithiothreitol (DTT) suggest that iBCLA does not act at sulfhydryl groups but at or near an amino group that is near a disulfide linkage in the polypeptide which catalyzes lactate transport. These experiments support the association between specific monocarboxylate transport and a 40–50-kdalton membrane-bound polypeptide of the rabbit erythrocyte.

Specific monocarboxylate transport appears to be a common feature in a variety of mammalian cells. Similarities in kinetics and susceptibility to such inhibitors as organomercurials (Deuticke et al., 1978; Spencer & Lehninger, 1976), cyanocinnamates (Spencer & Lehninger, 1976; Halestrap, 1976), and anhydrides (Johnson et al., 1980) occur in the lactate transport systems of human, rabbit, and other mammalian erythrocytes (Deuticke et al., 1978; Deuticke, 1982; Halestrap, 1976; Dubinsky & Racker, 1978), Ehrlich ascites tumor cells (Spencer & Lehninger, 1976; Johnson et al., 1980), and thymocytes (Anderson et al., 1978; Regen & Tarpley, 1978). Rabbit erythrocytes have been shown (Deuticke et al., 1978) to have a high capacity for lactate transport, suggesting that these cells have an abundance of this transport protein in the plasma membrane. Therefore, we have chosen these cells for the chemical labeling of the lactate transport protein. Jennings & Adams-Lackey (1982) presented evidence suggesting that a 40–50-kdalton integral membrane polypeptide is associated with lactate transport in rabbit erythrocytes. The evidence

†This work was supported by Research Grant GM-26861 from the National Institutes of Health. J.A.D. received support from Medical Scientist Training Program Grant GM-07337 from the NIH. M.L.J. received support from NIH RCDA AM-01137.

0006-2960/85/0424-0561$01.50/0