Structural Characterization of the 5′ Regions of the Human Phenylalanine Hydroxylase Gene†

David S. Konecki,‡ Yibin Wang,§ Friedrich K. Trefz,‡ Uta Lichter-Konecki,‡ and Savio L. C. Woo*,§

Universitäts-Kinderklinik, Im Neuenheimer Feld 150, W-6900 Heidelberg, FRG, and Howard Hughes Medical Institute, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030

Received February 3, 1992; Revised Manuscript Received May 12, 1992

ABSTRACT: Human phenylalanine hydroxylase (PAH) is expressed in a liver-specific manner and catalyzes the enzymatic conversion of phenylalanine to tyrosine. Genetic deficiency of PAH results in the autosomal-recessive disorder phenylketonuria (PKU). Through the application of genomic and cDNA cloning, primer extension studies, SI mapping experiments, and PCR methodologies, the transcription initiation (CAP) site has been identified and the 5′-flanking region determined. The most upstream CAP site for the human hepatic PAH gene transcript is located 154 nucleotides upstream of the first translation codon. The genomic and cDNA sequences analyzed demonstrated that the previously reported cDNA sequence, phPAH247 [Kwok et al. (1985) *Biochemistry 24*, 556–561], contained a 164-nucleotide cloning artifact at its 5′-end. The 319 base pair region immediately upstream of the CAP site is characterized by the lack of a proximal TATA box and the presence of sequences similar to GC boxes, CACCC boxes, CCAAT boxes, activator protein 2 (Ap–2) sites, partial glucocorticoid response elements (GREs), and partial cyclic AMP response elements (CREs). This suggests that the human PAH gene has a TATA-less promoter regulated by multiple transcription factors.

Phenylalanine hydroxylase (PAH), a mixed-function oxidase synthesized in the liver, catalyzes the hydroxylation of phenylalanine to tyrosine by molecular oxygen in the presence of the cofactor tetrahydrobiopterin (Kaufman, 1976). While the expression of this enzyme in rat kidney has been documented (Tourian et al., 1969; Rao & Kaufman, 1986), reports of nonhepatic human phenylalanine hydroxylase expression (Ayling et al., 1974; Hoffbauer & Schrempf, 1976) have since been discounted (Crawfurd et al., 1981). Clinically, PAH is important in that its deficiency results in the autosomal-recessive human disorder phenylketonuria (PKU), which is characterized by the excretion of large quantities of phenylpyruvate in the urine and accumulation of phenylalanine in the blood, resulting in hyperphenylalaninemia, mental retardation, and abnormal formation of the myelin sheath around neuronal axons in the central nervous system in untreated patients (Alvord et al., 1950; Shaw et al., 1972).

In 1983, Woo et al. (1983) reported the isolation of a recombinant clone possessing the complementary (c) DNA for the human phenylalanine hydroxylase enzyme. Subsequently, a full-length human PAH cDNA was isolated (Kwok et al., 1985), which allowed the determination of the primary structure of this enzyme, the localization of the gene to the long arm of chromosome 12 (Lidsky et al., 1985a), the establishment of a restriction fragment length polymorphism (RFLP) haplotype analysis system (Lidsky et al., 1985b; Woo, 1988), and the delineation of the organization of the human PAH gene (DiLella et al., 1986). Eukaryotic as well as prokaryotic expression studies using the coding region derived from the phPAH247 recombinant have conclusively demonstrated that a single mRNA species, about 2500 nucleotides in length, contains all the genetic information necessary to code for the functional PAH enzyme (Ledley et al., 1985). While considerable information concerning the molecular biology of the human PAH gene has been collected, the characterization of the transcription initiation site and the sequences 5′ of the transcription unit of the human PAH gene have been elusive.

We now report the structural characterization of the immediate 5′-flanking and 5′-transcribed regions of the human PAH gene, its major transcription initiation site in hepatic cell lines and liver, and potential regulatory sequences responsible for the hepatic-specific expression. This study also determined that the first 164 nucleotides of the previously reported full-length human PAH cDNA, phPAH247 (Kwok et al., 1985), were an artifact of cloning, which contributed significantly to the difficulties encountered in our characterization of the 5′-proximal region of the human PAH gene.

EXPERIMENTAL PROCEDURES

Materials

Restriction and DNA-modifying enzymes were obtained from New England Biolabs or Promega Corp. Reagents for tissue culture were from GIBCO and Flow Laboratory. Nylon transfer membranes were from New England Nuclear—DuPont. Taq DNA polymerase was obtained from Perkin-Elmer/Cetus. [α-^{32}P]dCTP, [α-^{35}S]dATP, and [γ-^{32}P]ATP were purchased from Amersham Corp. and New England Nuclear—DuPont. Reagents for dideoxy sequencing were from Pharmacia LKB Biotechnology and U.S. Biochemical Corp. Synthetic oligonucleotides were synthesized by American Synthesis, Inc., or by W. Fleischer of the German Cancer

† This work was supported in part by NIH Grant HD-17711 to S.L.C.W., who is also an investigator with the Howard Hughes Medical Institute. This work was also supported by Deutsche Forschungsgemeinschaft Grant Li.375/2-1/2-2 (to U.L.-K.) and Fritz Thyssen Foundation Grants 1990/51 (to Professor H. J. Bremer, U.L.-K., and D.S.K.), and 8.11/91 (to D.S.K.).

‡ Universitäts-Kinderklinik.

§ Baylor College of Medicine.

[1] Abbreviations: PAH, phenylalanine hydroxylase; PKU, phenylketonuria; c, complementary; AP-2, activator protein 2; GRE, glucorticoid response element; CRE, cyclic AMP response element; PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism; bp, base pair(s); kb, kilobase pair(s); nt, nucleotide(s); EDTA, ethylenediaminetetraacetic acid; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.

0006-2960/92/0431-8363$03.00/0

Research Center. Human liver tissue was obtained through the Department of Pediatric Surgery, University of Heidelberg Hospital of Surgery.

Cell Lines. The human hepatoblastoma-derived cell line Hep G2 (Knowles et al., 1984) was obtained from the American Type Culture Collection (ATCC HB 8065). Hep G2 cells were grown as a monolayer in culture media consisting of 25% Waymouth medium and 75% modified Eagle's medium, supplemented with 10% fetal bovine serum. Folse cells, a primary human fibroblast cell line generously provided by Dr. F. D. Ledley (Baylor College of Medicine, Houston, TX), were grown in monolayer in modified Eagle's medium supplemented with 20% fetal bovine serum, 2 × modified Eagle's medium amino acid solution (Gibco 320-1130 AG), and 5 mL of modified Eagle's medium vitamin solution (Gibco 320-1120 AG).

Methods

RNA Isolation. Total cellular RNA was isolated from cultured cells and human liver tissue according to the guanidine hydrochloride method of Przybyla et al. (1979). Poly(A)$^+$ and poly(A)$^{2+}$ RNA were obtained by one and two cycles of oligo(dT)–cellulose affinity chromatography, respectively. Concentrations of RNA samples were determined according to the ultraviolet absorption at 260 nm.

Genomic Clones. The cosmid recombinant cPAH-15, containing exons 1 and 2 of the human PAH gene, was isolated as previously described (DiLella et al., 1986). Subsequent analysis revealed the first exon of the human PAH gene to be contained in a *Pvu*II/*Eco*RI fragment, approximately 650 bp in length, derived from cPAH-15. This fragment was subcloned into the plasmid vector pBlueScript II-SK(–) (Stratagene). The insert of the resulting plasmid recombinant, designated pBS-PAH-1, was subjected to DNA sequence analysis, utilized in the screening of cDNA recombinants, and employed in the generation of radiolabeled probes for SI mapping studies.

Complementary DNA Library Construction and Screening. A human liver cDNA library was constructed by the method of Gubler and Hoffman (1987), employing the adaptor strategy of Sartoris et al. (1987) and the cloning vector λgt11 (Young & Davis, 1983). The resulting library, from 2.4 μg of human liver poly(A)$^{2+}$ RNA, consisted of about 2.0 × 10^8 independent recombinants. Approximately 1.5 × 10^7 recombinant bacteriophage from the unamplified library were screened by filter hybridization (Herrmann et al., 1986) using a ^{32}P-labeled (Bucan et al., 1990) 426 bp fragment generated by *Rsa*I digestion of the *Pvu*II/*Eco*RI insert from pBS-PAH-1. Hybridization-positive recombinants were plaque-purified by three rounds of purification. Phage DNA inserts were isolated, purified (Pohl et al., 1988), and subcloned into pBlueScript II-SK(–) vector for DNA sequence analysis.

DNA Sequence Analysis. Sequencing was performed by the double-stranded dideoxy chain termination technique using the Pharmacia kit or Sequenase (Sambrook et al., 1989). Sequencing of the exon 1-containing genomic fragment from pBS-PAH-1 employed the universal and reverse primers for pBlueScript, as well as internal primers derived from the coding sequence of phPAH247. This 650 bp insert was sequenced in its entirety on both strands. The 5′- and 3′-ends of cDNA fragments were determined through the use of the universal and reverse pBlueScript primers, although inserts from selected recombinants were sequenced completely, through the use of additional primers derived from the sequence of phPAH247. Computer analysis of DNA sequence data was performed at the European Molecular Biology Laboratory (Heidelberg, FRG) using the University of Wisconsin Genetics Computer Group (UWGCG) DNA sequence analysis package, version 6.0, or the on-line facilities of the MBIR services provided by Baylor College of Medicine.

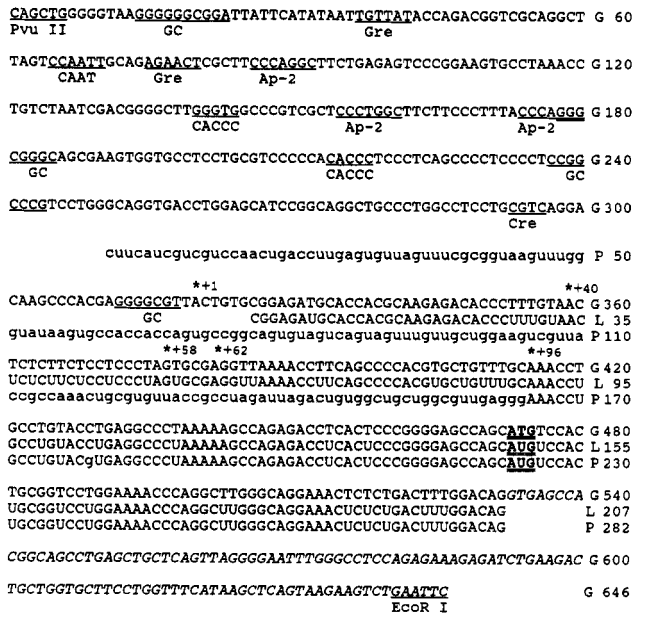

FIGURE 1: Nucleotide sequence of the genomic *Pvu*II/*Eco*RI fragment containing the 5′-regulatory region, exon 1, and part of intron 1 of the human PAH gene. Aligned with the genomic sequence (G) is the phPAH247 cDNA sequence (P) and the 5′ sequence (L) determined from the independent PAH cDNA recombinants K1024 and K1133. The 5′-regulatory region and PAH exon 1 are shown in upper case letters and intron sequences in upper case italic letters. The nonhomologous region of phPAH247 sequence is shown in lower case letters. Restriction endonuclease sites defining the ends of the genomic fragment, as well as putative cis elements, are underlined and indicated. The codon for the initiator methionine is boldfaced and double-underlined. The nucleotides marked by an asterisk are the putative CAP sites with their relative positions labeled.

Primer Extension Analysis. For primer extension reactions, a 21-base oligonucleotide, KW-D, complementary to the sense strand for the first 7 codons of the human PAH enzyme (G 493–473, Figure 1) was designed to hybridize with mRNA and prime reverse transcription. The oligonucleotide was 5′-end-labeled with [τ-^{32}P]ATP by T4 polynucleotide kinase (Sambrook et al., 1989) and purified by size-exclusion chromatography (Biogel P-2, Pharmacia). The primer extension reactions were performed according to Sambrook et al. (1989) and contained 10 ng of radiolabeled KW-D, preannealed with 25 μg of total RNA or 10 μg of poly(A)$^+$ RNA (Hep G2 or Folse cells), or 2.5 μg of human liver poly(A)$^{2+}$ RNA, or 40 μg of wheat germ tRNA (Sigma R-7876), at 58 °C for at least 16 h. Primer extension of annealed components (20-μL reaction volume) was accomplished by incubation at 42 °C for 90 min in the presence of 700 units/mL avian myeloblastosis virus reverse transcriptase (Promega). The primer-extended products were then resolved on 6% polyacrylamide [gradient (Biggin et al., 1983) or nongradient]/7 M urea sequencing gels in formamide-loading buffer containing 50 mM NaOH and detected by autoradiography. Size markers were either 5′-end-labeled fragments of pBR322 digested with *Msp*I or double-stranded sequencing reactions of pBS-PAH-1, using KW-D as the sequencing primer.

Amplification of Primer Extension Products. Primer extension products selected for amplification by the polymerase chain reaction (PCR) method (Saiki et al., 1988) were extracted from the polyacrylamide sequencing gel. In brief,

the area of the gel containing the primer extension band was removed to a microcentrifuge tube containing 50 μL of TE buffer (10 mM Tris-HCl/1 mM EDTA, pH 8.0) and subjected to a regimen of freezing, thawing, and vortexing. Ten microliters of the DNA extracted in this manner was used as template for subsequent PCR amplifications. One primer common to all PCR amplifications of extended products was KW-D (5′-TTCCAGGACCGCAGTGGACAT-3′), which was employed in their initial generation. Three additional primers, synthesized complementary to the antisense strand of sequence derived from the genomic insert of pBS-PAH-1, were KW-A (5′-TCCCTGGCTTCTTCCCTTTA-3′), G 154–173, Figure 1), KW-B (5′-CGTCAGGACAAGCCCACGAG-3′, G 293–312, Figure 1), and KW-C (5′-TGCGGAGAT-GCACCACGCAA-3′, G 324–343, Figure 1). Reaction mixtures contained 10 μL of gel-extracted primer extension product, 0.6 μg of each primer (KW-D + KW-A, -B, or -C), 2 units of Taq DNA polymerase (Perkin-Elmer/Cetus), and 250 μM deoxynucleotide triphosphates (dNTP) in a total volume of 100 μL of buffer containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgCl_2$, 0.01% gelatin, 5 mM β-mercaptoethanol, and 7 μM EDTA. Thirty-six cycles of amplification were performed, with the initial round consisting of denaturation for 7 min at 97 °C, annealing at 55 °C for 45 s, and extension for 45 s at 72 °C. The remaining 35 cycles contained denaturation steps of 30 s at 93 °C, annealing at 55 °C for 30 s, and extension at 72 °C for 45 s. Purification and sequencing of the resulting double-stranded amplified DNA were performed according to Wang et al. (1990).

S1 Nuclease Protection Experiments. The probe to be used in S1 mapping studies was generated by PCR amplification as described above, using the priming oligonucleotides KW-A and KW-D, with DNA from the genomic subclone pBS-PAH-1 as template. Following 36 cycles of amplification as described above, the PCR product was resolved on 4% NuSeive agarose (FMC Corp.), purified by electroelution from the excised band, and concentrated by ethanol precipitation. S1 nuclease analysis was performed using standard procedures (Sambrook et al., 1989). The same quantities of total, poly-$(A)^+$, or poly$(A)^{2+}$ RNA or tRNA as described for the primer extension studies were hybridized with 0.1 pmol of 5′-end-labeled probe at 53 °C for 16 h. Hybrids were digested with S1 nuclease (Pharmacia) at 25 °C for 45 min at a concentration of 1000 units of S1/mL of reaction. Following precipitation in the presence of wheat germ tRNA (50 μg/mL), samples were dissolved in formamide-loading buffer containing 100 mM NaOH, heated at 85 °C for 5 min, and separated on 6% polyacrylamide (gradient or nongradient)/7 M urea sequencing gels. The size markers were the same as described for analysis of the primer extension products.

RESULTS AND DISCUSSION

Genomic and cDNA Cloning. The previously characterized cosmid recombinant, cPAH-15 (DiLella et al., 1986), was subjected to various restriction endonuclease digestions and analyzed by Southern blot analysis, using a radiolabeled 211 bp 5′ *Eco*RI/*Sma*I fragment of phPAH247 as probe. The smallest hybridizing fragment detected (data not shown), about 650 bp in length, resulted from the digestion of the cPAH-15 DNA with the restriction endonucleases *Pvu*II and *Eco*RI. Subsequent restriction mapping studies and DNA sequencing (Figure 1) of this genomic *Pvu*II/*Eco*RI fragment verified that it possessed the previously determined PAH exon 1 sequence. Figure 1 shows the alignment of the phPAH247 cDNA insert sequence (P) with that of the genomic fragment (G). The first base of phPAH247 sequence homology (depicted by upper case letters in this sequence) occurs at genomic position 415, corresponding to phPAH247 nucleotide position 165. The observation that the homology between phPAH247 and the *Pvu*II/*Eco*RI genomic fragment did not begin at the 5′-end of the phPAH247 cDNA insert suggested the existence either of new exon sequences upstream in the gene or of a cloning artifact in the proximal region of this cDNA. To aid in resolving these two possibilities, a new human liver cDNA library was constructed.

Following the construction and screening of the human liver cDNA library, as described under Experimental Procedures, 90 of the more than 400 hybridization-positive recombinants were plaque-purified. On the basis of the size of the cDNA inserts determined for these clones by restriction analysis, inserts from 68 independent cDNA recombinants were subcloned into the *Eco*RI site of the plasmid vector pBlueScript II-SK(–). Subsequent double-stranded sequence analysis of the proximal and terminal regions of these cDNA inserts revealed that 57 of the cDNA inserts initiated upstream of the genomic position 415, which corresponded to the beginning of phPAH247 sequence homology with the genomic *Pvu*II/*Eco*RI fragment. Two independent recombinants (K1024 and K1133) from the newly constructed library initiated at genomic position 326, as shown in Figure 1 (L). In all cases, the insert sequences of the cDNA recombinants were contiguous and identical with the genomic sequence through genomic position 532, which represents the last nucleotide of PAH exon 1 (with the PAH intron 1 sequence shown in italics). These results excluded the possibility that the first 164 nucleotides of the phPAH247 cDNA sequence came from an upstream exon. Therefore, the transcription initiation site of the human PAH gene may be located at or upstream of genomic position 326, which corresponds to the first nucleotide of the 2 longest cDNA recombinants isolated and analyzed in this study.

Primer Extension and S1 Nuclease Protection Analysis. To map the transcription initiation start (CAP) site, both primer extension and S1 nuclease protection experiments were performed on RNA samples from surgically removed human liver tissue, human hepatoma (Hep G2) cells, and human fibroblast (Folse) cells as described under Experimental Procedures. Primer extension analyses utilized the 21mer oligonucleotide KW-D which was synthesized complementary to the sense strand coding for the first seven amino acids of the human PAH enzyme (see Figure 4). As shown in Figure 2A, specific primer extension products with sizes of 174, 135 cluster, 117, 113, and 79 nucleotides were detected from Hep G2 RNA, and the bands at 174, 113 and 79 nucleotides were also detected from human liver RNA (lanes 1, 2, and 4), but not from primary human fibroblasts (lane 3). To determine which of these primer-extended products truly represented the correct 5′-ends of the PAH gene transcript, S1 protection studies were performed as shown in Figure 2B, using a 5′-end-labeled probe generated by PCR amplification of the genomic subclone pBS-PAH-1 with oligonucleotides KW-A and KW-D. The protected fragments with sizes of 174, 135 cluster, 117, 113, and 79 nucleotides were revealed from Hep G2 RNA samples, while 135 cluster, 117, 113, and 79 bands were also detected from human liver RNA samples (lanes 3–5). As with the primer extension analyses, these protected fragments were not detected in reactions performed with RNA from human fibroblast cells (lanes 1 and 2). Since the same labeled oligonucleotide (KW-D) was used to perform the primer extension analyses and to generate the probe for S1 protection

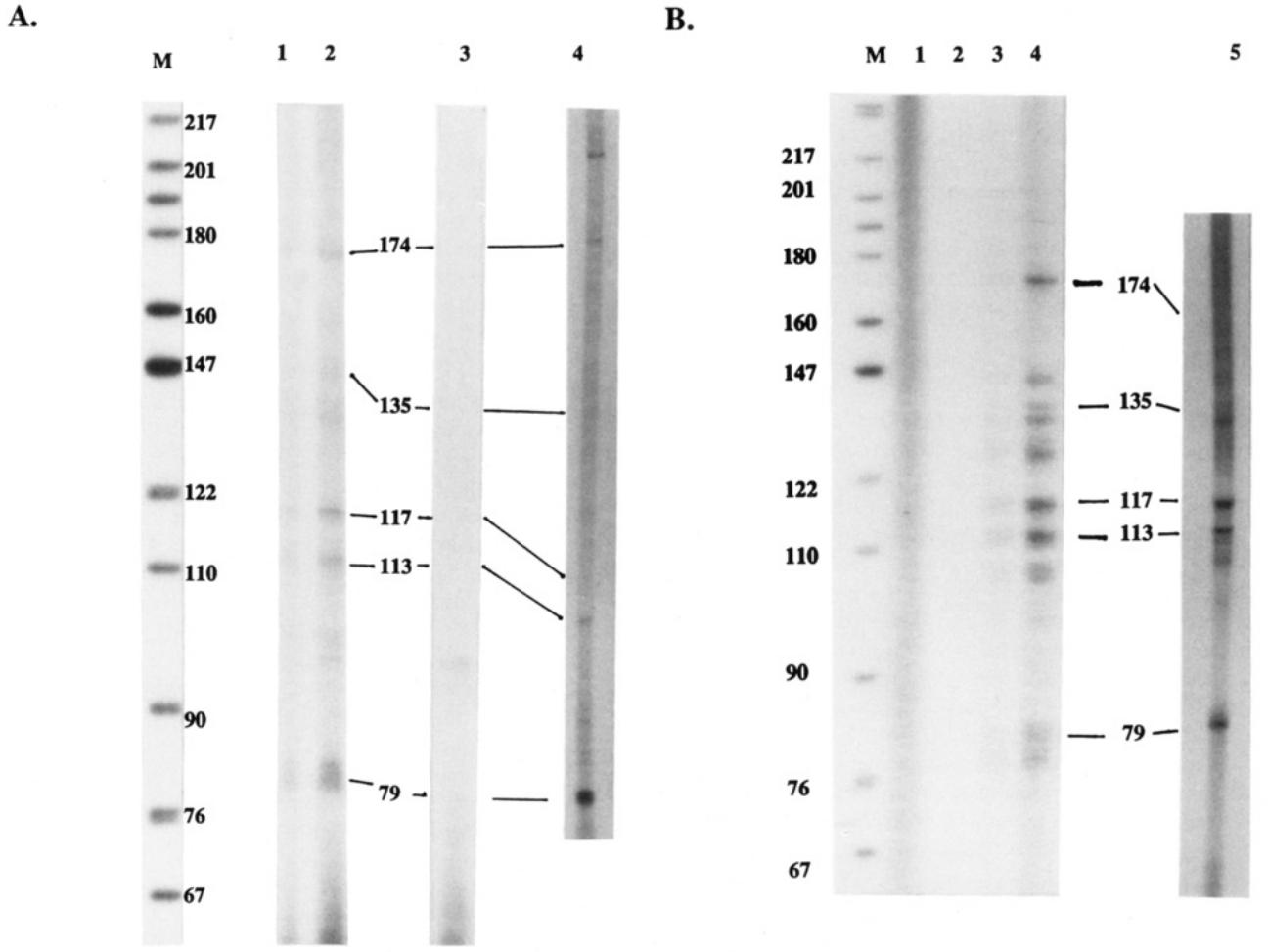

FIGURE 2: Determination of the hepatic transcription initiation site of the human PAH gene by (A) primer extension analysis and (B) S1 nuclease mapping. (A) Primer extension reactions performed with 25 μg of Hep G2 total RNA (lane 1), 10 μg of Hep G2 poly(A)$^+$ RNA (lane 2), 10 μg of human fibroblast poly(A)$^+$ RNA (lane 3), or 2.5 μg of human liver poly(A)$^{2+}$ RNA (lane 4). The priming oligonucleotide was [τ-^{32}P]ATP-labeled KW-D, synthesized complementary to the sense strand coding for the first seven amino acids of the PAH enzyme (G 493–473, Figure 1). Preannealed RNA/KW-D samples were extended at 42 °C for 90 min by the action of AMV reverse transcriptase. Size markers (lane M) were 5′-end-labeled fragments of *Msp*I-digested pBR322 plasmid DNA, whose sizes are indicated to the left of the lane. Indicated between lanes 2 and 3 are the sizes of the hepatic primer extension products. (B) S1 nuclease mapping experiments, using the probe generated by PCR amplification of the genomic subclone pBSPAH-1 with oligonucleotides KW-A and KW-D (5′-end-labeled, were performed with 25 μg of human fibroblast total RNA (lane 1), 10 μg of human fibroblast poly(A)$^+$ RNA (lane 2), 25 μg of Hep G2 total RNA (lane 3), 10 μg of Hep G2 poly(A)$^+$ RNA (lane 4), or 2.5 μg of human liver poly(A)$^{2+}$ RNA (lane 5). The size markers (lane M) were as described in (A). The sizes of the S1 nuclease-protected fragments are indicated between lanes 4 and 5.

studies, each coinciding S1-protected fragment and primer-extended product must have the same 5′-end.

The fact that the primer-extended products matched the corresponding S1 nuclease protected fragments suggested that the exonic region from oligonucleotide KW-D to the CAP site was contiguous. To ascertain the contiguity of this exonic region and the specificity of the primer-extended product, the longest (174 nt) primer extension product from Hep G2 cells was amplified by PCR using KW-D as the 3′-primer paired with KW-A , or -B, or -C as the 5′-primer (as schematically shown in Figure 4) and subjected to DNA sequence analysis of the amplified product (data not shown). As expected, the 174-nt primer-extended product could be amplified only with primer set KW-C and KW-D, but not with KW-A and KW-D or KW-B and KW-D (Figure 3). The sequence of the PCR product was identical to the genomic sequence (data not shown), which further excluded the possibility of RNA processing within this region. In conclusion, the combined results from primer extension analysis and S1 mapping pointed toward the most upstream transcription initiation site at genomic position 320, with multiple potential downstream CAP sites at positions 359, 377, 381, and 415 (Figure 4). Multiple CAP sites for one gene have been observed in many cases, especially among housekeeping family members (Perlino et al., 1987; Chelly et al., 1990; Chen et al., 1990; Kunze et al., 1990; Semenza et al., 1990). In these instances, most of the genes possess a TATA-less promoter with multiple GC-rich domains, which are conceivably responsible for determining the transcription initiation sites (Blake et al., 1990; Chen et al., 1990; Kunze et al., 1990). Though the human PAH 5′-flanking sequence has the characteristics of a "typical" TATA-less promoter (as discussed in the following section), we cannot rule out the possibility that the appearance of multiple downstream CAP sites is merely the result of ribonuclease degradation of a single transcription product initiated from genomic position 320.

Sequence and Structural Analysis of the 5′-Flanking Region. Having determined the CAP sites, the organization of the 5′-flanking region and the first exon of the human PAH gene could be demonstrated as shown in Figure 5. The +1 position corresponds to the most 5′ transcription initiation site at genomic position 320, with the putative downstream initiation sites at +40, +58, +62, and +96. Sequence analysis of the 5′-flanking region revealed several putative cis-acting

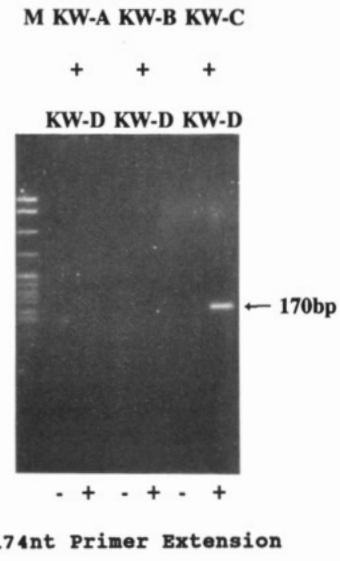

FIGURE 3: Determination of sequence contiguity between the longest primer extension product and the *Pvu*II/*Eco*RI genomic fragment by PCR amplification. Amplification primers KW-A, KW-B, or KW-C were used in combination with the primer extension oligonucleotides KW-D. PCR amplification reactions contained amplification reagents, the indicated oligonucleotides, and no DNA (–) or the gel-purified 174-nt primer extension product (+). The 5′-end of this extension product coincides with base 320 of the genomic fragment shown in Figure 1. The position of the resulting 170 bp amplified fragment is indicated. The size markers (M) were generated by *Msp*I digestion of pBR322 DNA.

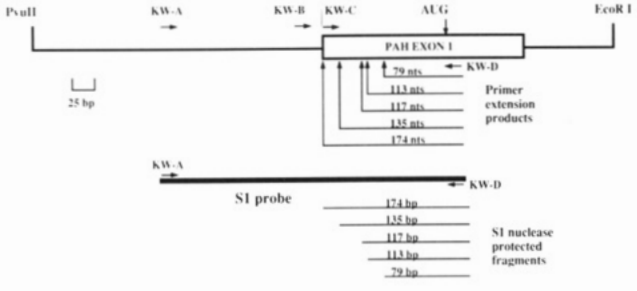

FIGURE 4: Schematic representation of the results of the primer extension analysis and S1 nuclease mapping studies which allowed determination of the human PAH hepatic CAP site. The forward amplification primers (KW-A, -B, and -C) are indicated above the line representing the *Pvu*II/*Eco*RI genomic fragment, with the reverse primer (KW-D) shown below this line. The KW-D oligonucleotide was used to prime extension reactions, and in combination with KW-A used to generate the probe for S1 nuclease protection experiments. The sizes and positions of the primer extension products and S1 nuclease protected fragments with regard to the PAH exon 1-containing genomic fragment are indicated.

elements that may be involved in the transcriptional regulation of human PAH gene expression. No TATA-like sequence was found in the proximal region, as in many classic TATA-(+) promoters. Four GC-rich domains were detected, at genomic positions 14–23, 178–185, 237–244, and 312–318, as putative Sp1-binding sites (Dynan, 1986). The occurrence of multiple GC boxes has been shown to promote transcription of several housekeeping genes [i.e., adenosine deaminase (Valerio et al., 1985), hypoxanthine phosphoribosyltransferase (Melton et al., 1984), dihydrofolate reductase (Azizkhan et al., 1986), and hydroxymethylglutaryl coenzyme A reductase (Reynolds et al., 1984)] as well as nonhousekeeping genes [i.e., the epidermal growth factor receptor (Ishii et al., 1985a), nerve growth factor receptor (Sehgal et al., 1988), human c–Ha–ras (Ishii et al., 1985b), and transforming growth factor α (Blasband et al., 1990) genes]. Positions 65–69 perfectly

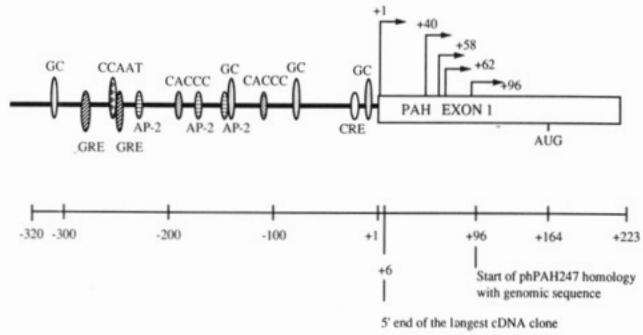

FIGURE 5: Schematic diagram showing the organization of the 5′-regulatory region and exon 1 of the human PAH gene. The relative positions and identification of sequences similar to the known cis-acting elements in the 319 bp region upstream of PAH exon 1 (open rectangle) are shown. The PAH liver-specific CAP site is marked +1 and corresponds to the 5′-end of the longest (174 nt) primer extension product (and S1 nuclease protected fragment of the same size). The 5′-ends of the other extension products and protected fragments are denoted by +40, +58, +62, and +96. The initiation of the most 5′ cDNA clones and the start of homology between phPAH247 and the genomic fragment are also indicated.

match the canonical CCAAT sequence, which is the target of factors regulating the efficiency of transcription (Stewart et al., 1991; Chodosh et al., 1988). It has been shown that glucocorticoid affects PAH enzyme activity (Kaufman, 1986; Dahl & Mercer, 1986); thus, the detection of two potential glucocorticoid response element (GRE) half–sites at genomic positions 37–42 and 75–80 was of great interest, since the glucocorticoid receptor monomers can bind to GRE half-sites (Eriksson & Wrange, 1990). Also of importance concerning the regulation of the PAH gene by glucocorticoid was the identification of the two potential CACCC-binding protein sites at positions 140–144 and 214–218. The CACCC-binding protein has been shown to be capable of interacting cooperatively with the glucocorticoid receptor (Schule et al., 1988a,b). Additionally, two consensus binding sites for activator protein 2 (AP-2; Mitchell et al., 1987) and one partial cyclic AMP response element (CRE; Lin & Green, 1989) site were detected at genomic positions 86–92, 174–181, or 293–296, respectively. Notably, no sequences were identified within this region showing significant homology with known cis-acting elements responsible for liver-specific regulation. The exact locations and functions of the cis-acting elements responsible for the PAH gene regulation are the subject of current investigation.

From our in vivo studies, a 9-kb fragment 5′-flanking to the human PAH gene has been shown to be able to confer tissue- and development-specific regulation to a reporter gene in transgenic animals (Wang et al., 1992). Importantly, the utilization of the same cluster of CAP sites was also observed from the transcription products of the transgene. All this evidence further suggested that the PAH 5′-flanking region of the human PAH gene contains functional promoter elements which can direct specific transcription initiated from genomic site 320, as well as multiple downstream sites.

Conclusion. This study has determined that the most 5′ transcription initiation site occurs 154 nucleotides upstream of the start of translation. Both cDNA cloning and CAP site mapping results suggested that the transcripts of the human PAH gene are contiguous at their 5′-ends, with no alternative splicing. This work also provided conclusive evidence that the first 164 nucleotides of the cDNA insert from phPAH247 (Kwok et al., 1985) result from a cloning artifact. The organization of the 5′-flanking region of the human PAH gene suggested that it has a TATA-less promoter, presumably interacting with multiple transcription factors through the

putative cis elements. This information opens new areas of investigation for the understanding of PAH expression and regulation. Additionally, this knowledge makes possible the detection of potential promoter mutations resulting in PAH deficiency and PKU.

ACKNOWLEDGMENT

We thank Lori Reed-Fourquet and Magdelena Schlotter for excellent technical assistance and Professor H. J. Bremer (Universitäts-Kinderklinik, Heidelberg) and Professor G. Schütz (German Cancer Research Center, Heidelberg) for support and helpful discussions.

REFERENCES

Alvord, E. C., Stevenson, L. D., Vogel, F. S., & Engle, R. L. (1950) *J. Neuropathol. Exp. Neurol. 9*, 298–310.

Ayling, J. E., Pirson, W. D., Al-Janabi, J. M., & Helfland, G. D. (1984) *Biochemistry 13*, 78–85.

Azizkhan, J. C., Vaughn, J., Christy, R. J., & Hamlin, J. H. (1986) *Biochemistry 25*, 6228–6236.

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Blake, M. C., Jambou, R. C., Swick, A. G., Kahn, J. W., & Azizkhan, J. C. (1990) *Mol. Cell. Biol. 10*, 6632–6641.

Bucan, M., Zimmer, M., Whaley, W. L., Poustka, A. Youngman, S., Allitto, A., Ormondroyd, E., Smith, B., Pohl, T. M., MacDonald, M., Bates, G. P., Richards, J., Volinia, S., Gilliam, T. C., Sedlacek, Z., Collins, F. S., Wasmuth, J. J., Shaw, D. J., Gusella, J. F., Frischauf, A.-M., & Lehrach, H. (1990) *Genomics 6*, 1–15.

Chelly, J., Hamard, G., Koulakoff, A., Kaplan, J.-C., Kahn, A., & Berwald-Netter, Y. (1990) *Nature 344*, 64–65.

Chen, A., Reyes, A., & Akeson, R. (1990) *Mol. Cell. Biol. 10*, 3314–3324.

Chodosh, L. A., Baldwin, A. S., Carthew, R. W., & Sharp, P. A. (1988) *Cell 53*, 11–24.

Crawfurd, M. A., Gibbs, D. A., & Sheppard, D. M. (1981) *J. Inherited Metab. Dis. 4*, 191–195.

Dahl, H. H.-M., & Mercer, J. F. B. (1986) *J. Biol. Chem. 261*, 4148–4153.

DiLella, A. G., Kwok, S. C. M., Ledley, F. D., Marvit, J., & Woo, S. L. C., (1986) *Biochemistry 25*, 743–749.

Dynan, W. S. (1986) *Trends Genet. 2*, 196–197.

Eriksson, P., & Wrange, O. (1990) *J. Biol. Chem. 265*, 3535–3542.

Gubler, U., & Hoffman, B. J. (1983) *Gene 25*, 227–237.

Herrmann, B., Bucan, M., Mains, P. E., Frischauf, A.-M., Silover, L. M., & Lehrach, H. (1986) *Cell 44*, 469–476.

Hoffbauer, R. W., & Schrempf, G. (1976) *Lancet II*, 194.

Ishii, S., Xu, Y.-H., Stratton, R. T., Roe, B. A., Merlino, G. T., & Pastan, I. (1985a) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4920–4924.

Ishii, S., Merlino, G. T., & Pastan, I. (1985b) *Science 230*, 1378–1381.

Kaufman, S. (1986) *Adv. Neurochem. 2*, 1–132.

Kaufman, S. (1986) *Adv. Enzyme Regul. 25*, 37–64.

Knowles, B. B., Searles, D. B., & Aden, D. P. (1984) *Advances in Hepatitis Research*, Masson Publishing Co., Chicago, IL.

Kunze, N., Klein, M., Richter, A., & Knippers, R. (1990) *Eur. J. Biochem. 194*, 323–330.

Kwok, S. C. M., Ledley, F. D., DiLella, A. G., Robson, K. J. H., & Woo, S. L. C. (1985) *Biochemistry 24*, 556–561.

Ledley, F. D., Grenett, H. E., DiLella, A. G., Kwok, S. C. M., & Woo, S. L. C. (1985) *Science 228*, 77–79.

Lidsky, A. S., Law, M. L., Morse, H. G., Kao, F. T., Rabin, M., Ruddle, F. H., & Woo, S. L. C. (1985a) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6221–6225.

Lidsky, A. S., Ledley, F. D., DiLella, A. G., Kwok, S. C. M., Daiger, S. P., Robson, K. J. H., & Woo, S. L. C. (1985b) *Am. J. Hum. Genet. 37*, 619–634.

Lin, Y.-S., & Green, M. R. (1989) *Nature 340*, 656–659.

Melton, D. W., Konecki, D. S., Brennand, J., & Caskey, C. T. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2147–2151.

Mitchell, P. J., Wang, C., & Tijan, R. (1987) *Cell 50*, 847–861.

Perlino, E., Cortese, R., & Ciliberto, G. (1987) *EMBO J. 6*, 2767–2771.

Pohl, T. M., Zimmer, M., MacDonald, M. E., Smith, B., Bucan, M., Poustka, A., Volinia, S., Searle, S., Zehetner, G., Wasmuth, J. J., Gusella, J., Lehrach, H., & Frischauf, A.-M. (1988) *Nucleic Acids Res. 16*, 9185–9198.

Przybyla, A. E., MacDonald, R. J., Warding, J. D., Pictet, R. L., & Rutter, W. J. (1979) *J. Biol. Chem. 254*, 2154–2159.

Rao, D. N., & Kaufman, S. (1986) *J. Biol. Chem. 261*, 8866–8876.

Reynolds, G. A., Basu, S. K., Osborne, T. F., Chin, D. J., Gil, G., Brown, M. S., Goldstein, J. L., & Luskey, K. L. (1984) *Cell 38*, 275–285.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487–491.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sartoris, S., Cohen, E. B., & Lee, J. S. (1987) *Gene 56*, 301–307.

Schule, R., Muller, M., Otsuka-Murakami, H., & Renkawitz, R. (1988a) *Nature 332*, 87–90.

Schule, R., Muller, M., Kaltschmidt, C., & Renkawitz, R. (1988b) *Science 242*, 1418–1420.

Sehgal, A., Patil, N., & Chao, M. (1988) *Mol. Cell. Biol. 8*, 3160–3167.

Sememza, G. L., Dureza, R. C., Traystman, M. D., Gearhart, J. D., & Antonarakis, S. E. (1990) *Mol. Cell. Biol. 10*, 930–938.

Shah, S. N., Peterson, N. A., & McKean, C. M. (1972) *J. Neurochem. 19*, 479–485.

Stewart, M. J., Shean, M. L., Paeper, B. W., & Duester, G. (1991) *J. Biol. Chem. 266*, 11594–11603.

Tourian, A., Goddard, J., & Puck, T. T. (1969) *J. Cell. Physiol. 73*, 159–170. Valerio, D., Duyvesteyn, M. G. C., Dekker, B. M. M. Weeda, G., Berkvens, T. M., Vander Voorn, L., van Ormondt, H., & van der Eb, A. J. (1985) *EMBO J. 4*, 437–443.

Wang, T., Okano, Y., Eisensmith, R. C., Fekete, G., Schuler, D., Berensci, G., Nasz, I., & Woo, S. L. C. (1990) *Somat. Cell Mol. Genet. 16*, 85–89.

Wang, Y., DeMayo, J. L., Hahn, T. M., Finegold, M. J., Konecki, D. S., Lichter-Konecki, U., & Woo, S. L. C. (1992) *J. Biol. Chem.* (in press).

Woo, S. L. C. (1988) *Am. J. Hum. Genet. 43*, 781–783.

Woo, S. L. C., Lidsky, A. S., Güttler, F., Chandra, T., & Robson, K. J. H. (1983) *Nature 306*, 151–155.

Young, R. A., & Davis, R. W. (1983) *Science 222*, 778–782.