Biochemistry

 Volume 25, Number 4 February 25, 1986

Accelerated Publications

Molecular Structure and Polymorphic Map of the Human Phenylalanine Hydroxylase Gene†

Anthony G. DiLella, Simon C. M. Kwok,‡ Fred D. Ledley, Joshua Marvit, and Savio L. C. Woo*

Howard Hughes Medical Institute, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030

Received November 6, 1985; Revised Manuscript Received December 20, 1985

ABSTRACT: Human phenylalanine hydroxylase is a liver-specific enzyme that catalyzes the conversion of phenylalanine to tyrosine. Absence of enzymatic activity results in phenylketonuria, a genetic disorder that causes development of severe mental retardation in untreated children. In this paper we report the cloning and structure of the normal human phenylalanine hydroxylase gene, which was isolated in four overlapping cosmid clones that span more than 125 kilobases (kb) of the genetic locus. The peptide coding region of the gene is about 90 kb in length and contains 13 exons, with intron sizes ranging from 1 to 23 kb. Exons at the 3′ half of the gene are compact, whereas those at the 5′ half are separated by large introns. The human phenylalanine hydroxylase gene codes for a mature messenger RNA of approximately 2.4 kb, and its noncoding to coding DNA ratio is one of the highest among eukaryotic genes characterized to date. The map positions of nine polymorphic restriction sites identified within the locus were established by restriction enzyme mapping of the cloned gene fragments. Two clusters of polymorphic sites were demonstrated: (1) *Bgl*II, *Pvu*II(a), and *Pvu*II(b) at the 5′ end of the gene and (2) *Eco*RI, *Xmn*I, *Msp*I(a), *Msp*I(b), *Eco*RV, and *Hin*dIII at the 3′ end. The polymorphic site distribution within this gene is a useful tool for prenatal diagnosis and carrier detection of the genetic disorder, while knowledge of normal gene structure is a prerequisite for future characterization of mutant alleles.

Classical phenylketonuria (PKU) is an autosomal recessive human genetic disorder caused by a deficiency in hepatic phenylalanine hydroxylase (PAH, phenylalanine 4-monooxygenase, EC 1.14.16.1). The enzyme is a mixed-function oxidase that catalyzes the rate-limiting step in the hydroxylation of phenylalanine to tyrosine (Scriver & Clow, 1980; Kaufman, 1976). A deficiency of this enzyme causes an accumulation of serum phenylalanine, resulting in hyperphenylalaninemia and abnormalities in the metabolism of many compounds derived from the aromatic amino acids (Blau, 1979). If left untreated, these metabolic abnormalities cause postnatal brain damage and severe mental retardation. PKU is the most common inborn error in amino acid metabolism, with an average incidence of about 1 in 10 000 Caucasians. The mutant gene frequency is such that 1 in 50 individuals are carriers of the disease trait (Scriver & Clow, 1980).

Purified PAH from rat (Kaufman & Fisher 1970), monkey (Cotton & Grattan, 1975), and human liver (Woo et al., 1974; Friedman & Kaufman, 1973; Abita et al., 1983) has a molecular weight of about 100 000 and is a dimer comprised of 49 000–50 000-dalton subunits. PAH is a substrate for cAMP-dependent protein kinase, and monomers composed of phosphorylated and dephosphorylated forms were recently identified for rat (Donlan & Kaufman, 1980), human, and monkey (Smith et al., 1984). The rat enzyme is a metalloprotein containing 1 mol of iron/mol of subunit and requires the cofactor tetrahydrobiopterin for activity (Scriver & Clow, 1980).

We have recently reported the isolation of a full-length human PAH cDNA clone, phPAH247 (Kwok et al., 1985). The amino acid sequence deduced from this cDNA clone predicts a protein of 451 amino acids with a molecular weight of 51 900. DNA-mediated gene transfer experiments dem-

† This work was partially supported by National Institutes of Health Grant HD-17711 to S.L.C.W., who is also an investigator of the Howard Hughes Medical Institute. A.G.D. is the recipient of NIH Postdoctoral Fellowship HD-06495.

‡ Present address: Department of Biochemistry, College of Health Sciences and Hospital, The University of Kansas, School of Medicine, Kansas City, KS 66103.

0006-2960/86/0425-0743$01.50/0

onstrated that phPAH247 contains all the genetic information necessary to code for a functional enzyme (Ledley et al., 1985a), providing direct evidence that the dimeric structure of PAH is comprised of identical subunits. We have further used this cDNA clone as a probe to demonstrate by restriction fragment length polymorphism (RFLP) analysis that PKU mutations are linked to the PAH locus (Woo et al., 1983), supporting the observation that the active enzyme is encoded by a single gene (Ledley et al., 1985a).

The genetic lesions causing the loss of enzyme activity in PKU are not known. Southern hybridization analysis of DNA isolated from a large number of PKU individuals demonstrated that the disorder is not caused by any obvious deletions in the PAH gene (Woo et al., 1983). We recently detected PAH mRNA in some PKU liver biopsy specimens (DiLella et al., 1985a), supporting initial observations that PKU is manifested by a structurally altered enzyme having less than 1% of normal enzyme activity (Bartholome & Dresel, 1982; Friedman et al., 1973; Bartholome & Ertel, 1976). However, the cause of the disease must be heterogeneous, in that PAH mRNA was not detected in all PKU liver biopsies (DiLella et al., 1985a). The liver-specific expression of the PAH gene and the scarcity of human liver biopsy tissue have precluded the use of patient material for studying the molecular basis of PKU at the mRNA and protein levels. Alternatively, PKU mutations can be identified at the genomic DNA level by sequencing the mutant genes followed by a study of mutational effects on protein structure and activity by gene transfer experiments. These studies require prior knowledge of the normal PAH gene structure, which is the subject of the present report.

Materials and Methods

Materials

Restriction endonucleases and T4 DNA ligase were purchased from Bethesda Research Laboratories. *Escherichia coli* DNA polymerase I (Klenow fragment) and calf intestine alkaline phosphatase were products of Boehringer Mannheim. The M13 cloning/sequencing kit was obtained from P-L Biochemicals. [α-^{32}P]dNTP (specific activity >3000 Ci/mmol) for nick translation and [α-^{35}S]dATP (650 Ci/mmol) for sequencing were purchased from Amersham Corp.

Methods

Construction and Screening of a Human Cosmid Genomic DNA Library. Lymphocyte DNA was isolated from a homozygous normal individual as evidenced by RFLP analysis and used to construct a cosmid DNA library according to established procedures (Lau & Kan, 1983). Genomic DNA was partially digested with MboI to a length of 30–50 kilobases (kb) and size-fractionated on a 1.25–5 M NaCl gradient. Cosmid vector pCV107 (Lau & Kan, 1983) was linearized with *Bam*HI and treated with calf intestine alkaline phosphatase (2.5 units/50 pmol of 5′ termini). This vector contains two *Eco*RI sites flanking the *Bam*HI cloning site, which facilitates mapping of recombinant clones. Size-fractionated genomic DNA (1.5 μg) was ligated to pCV107 (3 μg) in a 20-μL reaction at 14 °C overnight. The resulting concatenated DNA molecules were packaged with extracts prepared by the procedure of Grosveld et al. (1981). The infectious bacteriophage particles were used to transduce *E. coli* ED8767, with an efficiency of about 500 000 transformants per microgram of size-fractionated genomic DNA. Approximately 800 000 recombinant colonies were grown on nitrocellulose filters (40 000 colonies/137-mm filter) and hybridized to ^{32}P-labeled human PAH cDNA probes according to the screening procedure of Hanahan and Meselson (1980). DNA was isolated from positive colonies by the alkaline lysis (Birnboim & Doly, 1979) procedure and characterized by restriction enzyme mapping.

Restriction Enzyme Mapping, Gel Electrophoresis, and Southern Blotting. Enzyme digests of cosmid and genomic DNA were electrophoresed on agarose gels, transferred to nitrocellulose filters, and hybridized to ^{32}P-nick-translated PAH cDNA probes (10^8 cpm/μg) as previously described (Woo et al., 1983). Restriction enzyme mapping of cosmid clones and selected DNA fragments was carried out as described by Milbrandt et al. (1983) and Smith and Birnstiel (1976).

DNA Sequence Analysis. Exon-containing *Eco*RI fragments were subcloned from the cosmid clones. Exon regions were then isolated from the subclones, inserted into M13 mp18 (Messing, 1983), and sequenced by the dideoxynucleotide chain-termination method of Sanger et al. (1977). For exons containing unique restriction endonuclease sites present in the cDNA, the corresponding enzymes were used to generate DNA fragments so that, after blunt-end cloning into M13 mp18, the exons would be adjacent to the priming site in the M13 vector for sequence analysis. Alternatively, large DNA fragments were digested with restriction enzymes that cut the DNA frequently. These digests were then shotgun-cloned into M13 mp18 (Messing, 1983), and recombinant clones were screened for the presence of exon segments by using PAH cDNA probes (Benton & Davis, 1977). Overlapping sequence data for these exons were then generated by using clones containing inserts of both orientations. In addition, some exons were sequenced by using exon synthetic oligonucleotide primers (deduced from the PAH cDNA sequence) specific for both ends of each exon (Kwok et al., 1985).

Results

Isolation and Characterization of the Human Phenylalanine Hydroxylase Gene. Figure 1A shows a Southern blot hybridization analysis of genomic DNA digested with *Eco*RI and probed with human PAH cDNA. Seven hybridizing *Eco*RI fragments representing 66 kb of genomic DNA were detected with the full-length PAH cDNA probe (lane 1). The structural complexity of the PAH gene in genomic DNA was then determined by using 5′- and 3′-specific PAH cDNA probes. The 5′-cDNA probe detected six *Eco*RI fragments containing 54 kb of genomic DNA (lane 2). In addition to a common 10-kb *Eco*RI fragment, the 3′-cDNA probe detected a 12-kb band not detected by the 5′-cDNA probe (lane 3). These data indicated that the 3′ half of the gene was compact and contained in approximately 12 kb of genomic DNA, whereas the 5′ exons were separated by very large introns, spanning over 54 kb of DNA.

Due to the large size of the human PAH gene, we constructed a genomic DNA library using a cosmid vector (pCV107) which can accommodate large DNA inserts. We screened the library with the PAH cDNA probe and isolated the corresponding genomic sequences from a number of positive clones. Figure 1B shows a Southern blot of four overlapping PAH cosmid clones digested with *Eco*RI and probed with 5′- and 3′-cDNA probes. All four clones hybridized to the 5′ probe (lanes 4–7), while only cPAH14 and cPAH25 hybridized to the 3′ probe (lanes 8 and 9, respectively). This demonstrated that cPAH15 and cPAH10 contained genomic sequences corresponding to the 5′ half of the gene. The 10-kb *Eco*RI fragment detected in genomic DNA by using the 5′- and 3′-cDNA probes (lanes 2 and 3, respectively) was also observed in cPAH 14 using these probes (lanes 6 and 8, respectively). Clone cPAH25 contained a 12-kb *Eco*RI frag-

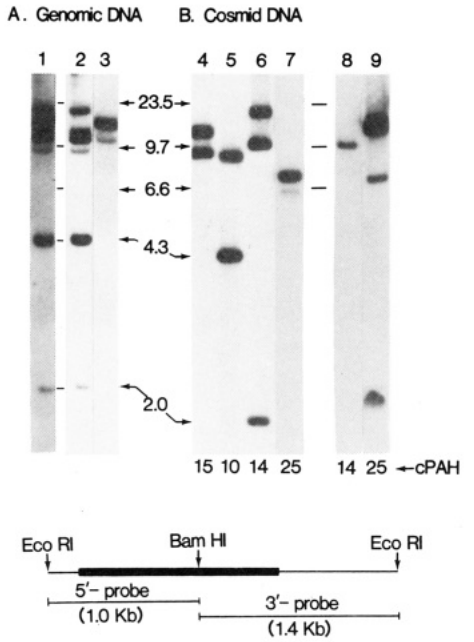

FIGURE 1: *Eco*RI and Southern hybridization analysis of genomic and cosmid DNA. Genomic DNA (panel A) was digested with *Eco*RI and hybridized to full-length PAH cDNA (lane 1) and 5′- and 3′-specific cDNA probes (lanes 2 and 3, respectively) generated by *Bam*HI restriction of phPAH247 cDNA (bottom). PAH cosmid clones (panel B) were digested with *Eco*RI and hybridized to 5′-cDNA (lanes 4–7) and 3′-cDNA (lanes 8 and 9) probes. The solid box and lines in pPAH247 (bottom) represent the coding and untranslated cDNA regions, respectively.

ment which hybridized specifically to the 3′-cDNA probe (lane 9), which was in agreement with the genomic DNA hybridization data (lane 3) and suggested that cPAH25 contained genomic DNA sequences located at the 3′ end of the gene.

Figure 2 (top) shows the restriction map of the PAH gene which is contained in four overlapping cosmid clones covering more than 125 kb of contiguous genomic DNA. Exon-containing *Eco*RI fragments detected in Southern blot analysis of genomic DNA (Figure 1, lane 1) were contained within the overlapping cosmid clones and were labeled A–G (Figure 2). Clones cPAH10 and cPAH14 overlapped within an intron. To determine the extent of overlap or any possible gap between these clones, we characterized several cosmid clones spanning this region and containing exons 3 and 4 (data not shown). Restriction enzyme mapping of these clones demonstrated that exons 3 and 4 were interrupted by a 23-kb intron, verifying the overlap between cPAH10 and cPAH14 and excluding the possibility of any gaps between these clones.

Sequence and Detailed Structure of the Human PAH Gene. The detailed structure of the PAH gene is shown at the bottom of Figure 2. The precise locations of the exon–intron boundaries were defined by sequencing the appropriate regions of the cloned genomic DNA (Figure 3) and aligning them with the PAH cDNA sequence reported previously (Kwok et al., 1985). The gene contains 13 exons spanning approximately 90 kb of DNA. Introns divide the protein-coding portion of the gene into exons of 57–892 base pairs (bp), with a mean exon length of 114 bp (Figure 4).

Figure 4 lists the sequences at the 5′ and 3′ boundaries of each intron. Intron–exon boundaries were consistent with established consensus splice sites (Mount, 1982). All introns in the PAH gene begin and end with the canonical GT and AG dinucleotide, respectively. The final 17 nucleotides at the 3′ end of each intron contain only one AG dinucleotide and are pyrimidine rich. However, for 8 out of 12 donor splice sites in the PAH gene, a second GT dinucleotide was present within the consensus sequence. Similar findings were reported for other genes (Mount, 1982). The role of the second GT dinucleotide in the splicing process is not known. Three types of intron–exon junctions were observed in the PAH gene. Type O junctions (those between codons) occur 6 times, whereas type I and II junctions (interrupting codons after the first and second dinucleotide, respectively) each occur 3 times. These observations are in accord with intron–exon boundaries of other eukaryotic genes (Mount, 1982; Sharp, 1981; Reynolds et al., 1984).

The most 3′ exon (Figure 4, exon 13) is 892 bp in length. It contains 41 bp of protein-coding sequence, the termination codon, and the entire 848 bp of 3′-untranslated sequence previously reported for PAH cDNA (Kwok et al., 1985). The dinucleotide site of polyadenylation (TA) was inferred from PAH cDNA clones (Kwok et al., 1985) and previously reported consensus sequences (Birnstiel et al., 1985) and is preceded 13 bases by the common polyadenylation signal AAUAAA (Figure 4). The CACTG sequence motif identified in other eukaryotic genes (Birnstiel et al., 1985) occurs 9 bp downstream of the polyadenylation site in the PAH gene (Figure 4) and is believed to function in RNA processing and

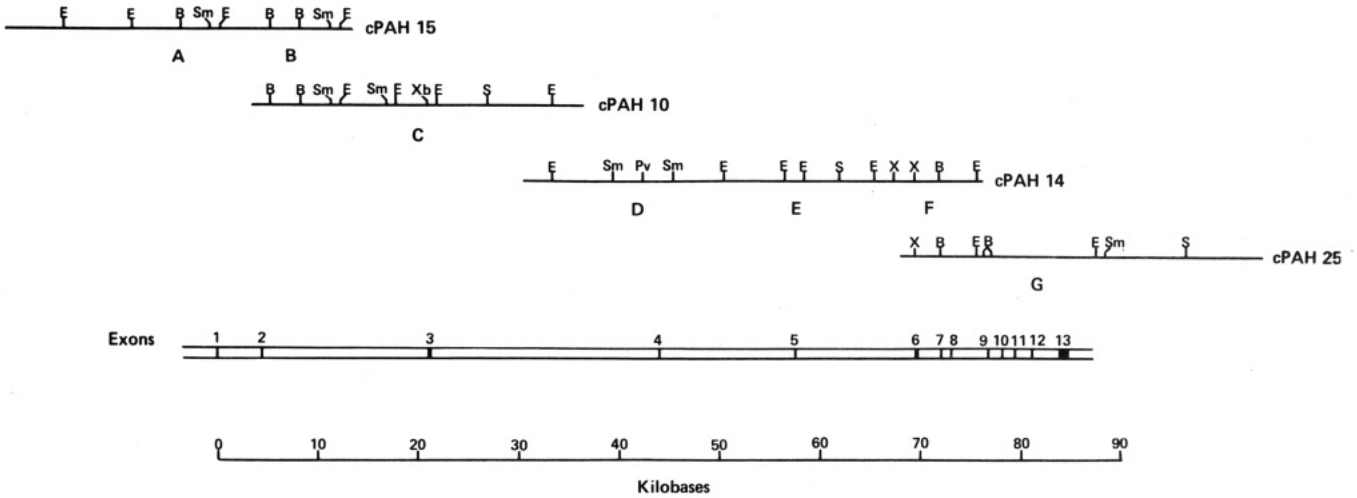

FIGURE 2: Physical map of the human PAH gene. The map of the four overlapping cosmid clones (cPAH15, cPAH10, cPAH14, and cPAH25) that were used to characterize the gene is indicated at the top. The subcloned *Eco*RI fragments derived from the cosmid clones and used for DNA sequence analysis are labeled A–G. The detailed structure of the PAH gene in the 5′ to 3′ orientation is shown (bottom). Exons and introns are denoted by solid and open boxes, respectively. Restriction enzyme abbreviations: E, *Eco*RI; B, *Bam*HI; Sm, *Sma*I; Xb, *Xba*I; S, *Sal*I; Pv, *Pvu*I; X, *Xho*I.

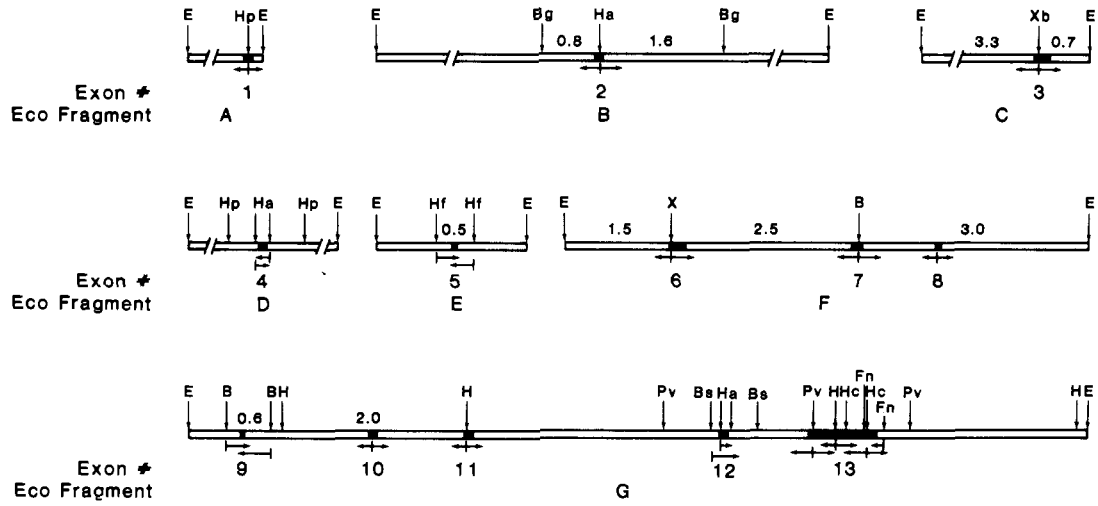

FIGURE 3: Sequencing strategy used for determining the exon–intron junctions of the human PAH gene. *Eco*RI fragments A–G were digested with the indicated restriction enzymes, and the exons were cloned into M13 mp18 for sequence analysis. The arrows indicate the direction of sequencing for each fragment. Exons 8 and 10 were sequenced by using exon-specific synthetic oligonucleotide primers as described under Materials and Methods.

Exon #	Size(bp)	Intron #	Size(kb)	Exon	Intron	Exon
1	118	1	4.5	GGA CAG Glu 20	gtgagccacg.....tgcatcttatcctgtag	GAA AGA Glu 21
2	108	2	16.5	TTT GAG Glu 56	gtcagtacta.....cccattctctcttctag	GAG AAT Glu 57
3	184	3	23.5	ACA G Val 118	gtaagaatta.....ccttctctgtgtttcag	TG CCC Val 118
4	89	4	13.5	CAC CCT Pro 147	gtgagtccat.....tgtctcttttctcctag	GGT TTT Gly 148
5	68	5	12.0	CGC CA His 170	gtaagtctgc.....ttgtgcctgtattctag	T GGG His 170
6	197	6	2.5	CAG A Thr 235	gtaagtccac.....tcttcttttcatcccag	CT TGC Thr 235
7	136	7	1.0	GAA CC Pro 280	gtgagtactg.....tgctttctgtctttcag	T GAC Pro 280
8	70	8	3.5	TCC CAG Gln 303	gtaaggaatg.....ttttcccccaattacag	GAA ATT Glu 304
9	57	9	1.5	GCC ACA Thr 322	gtaagtccct.....attgactttccatccag	ATT TAC Ile 323
10	96	10	1.2	TTA CAG Gln 354	gtatgacctt.....tcacttggggcctacag	TAC TGC Tyr 355
11	134	11	3.2	GTA AG Arg 399	gtgaggtggt.....tgtggttttggtcttag	G AAC Arg 399
12	116	12	1.2	AAC A Ser 438	gtaagtaatt.....gttgtttttctttgtag	GT GAA Ser 438
13	892			AAG TAA Lys Stop	..... AATAAACATTAGTAGTTTAAACAATTTCACTGTTT	
					CAGTTATTTCCTGGAATGAAATATTCCTCTCACTGCCTTAGAAATACAT -3'	

FIGURE 4: Exon–intron junctions of the human PAH gene. The nucleotide sequences of the exon–intron junctions were determined from genomic subclones according to the strategy shown in Figure 3. Exon sequences are in capital letters; intron sequences are in lower case letters. Intron junctions were positioned by applying the gt-ag splice rule (Mount, 1982). The numbers shown at the intron–exon splice junctions denote the positions of the corresponding amino acids in PAH, as deduced from the previously published PAH cDNA sequence (Kwok et al., 1985). The polyadenylation site is indicated by an arrow. Polyadenylation signals are underlined.

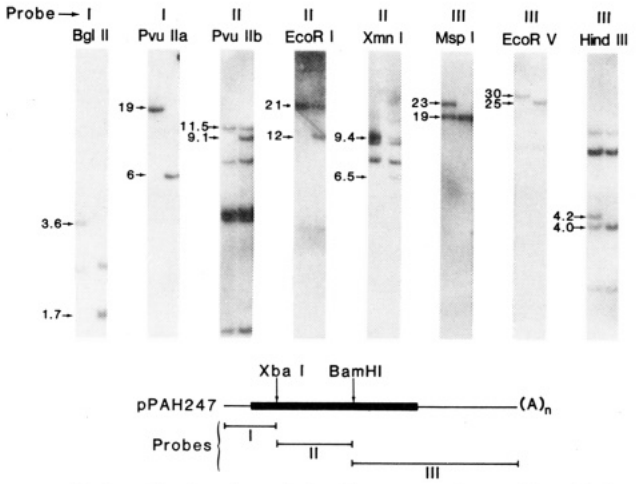

FIGURE 5: Localization of restriction fragment polymorphisms in the human PAH gene. Genomic DNA was isolated from lymphocytes of normal individuals with previously characterized RFLP haplotypes (Lidsky et al., 1985a), which illustrate the restriction site variation when DNA is cleaved with the indicated restriction enzymes and hybridized to probes derived from different regions of phPAH247 cDNA. Individuals that were either homozygous or heterozygous for a particular RFLP were used in this analysis. In each panel, the polymorphic fragment sizes (in kilobases) are indicated. Bottom: cDNA probes I, II, and III were generated by *Xba*I and *Bam*HI digests of phPAH247 cDNA. The solid box and lines in the cDNA represent the coding and untranslated regions, respectively.

polyadenylation. Exon 1 (149 bp) contains 89 bp of the 5′-untranslated region upstream of the initiation methionine codon and codes for 20 amino acids (Figure 4). Characterization of the transcription start site is in progress.

Localization of RFLP's in the Human Phenylalanine Hydroxylase Locus. We recently identified nine restriction site polymorphisms in the PAH locus using the full-length human PAH cDNA probe (Lidsky et al., 1985a). Subfragments of the PAH cDNA clone (Figure 5, bottom) were used to tentatively localize these polymorphisms in the PAH gene by Southern blot analysis of lymphocyte DNA isolated from individuals with previously characterized RFLP haplotypes (Lidsky et al., 1985a) (Figure 5). *Bgl*II and *Pvu*II(a) polymorphisms were detected with cDNA probe I which contains exons 1 and 2. *Pvu*II(b), *Eco*RI, and *Xmn*I polymorphisms were detected with cDNA probe II which contains exons 3–7. Probe III, comprised of exons 8–13, identified *Msp*I, *Eco*RV, and *Hin*dIII polymorphisms. The exact locations of these RFLP's were then determined by restriction endonuclease mapping of the corresponding exon regions of the PAH cosmid clones and verified by exploiting restriction fragment length differences that exist between cosmid clones derived from alleles of different RFLP haplotypes (data not shown).

Figure 6 shows the RFLP map of the PAH locus. The RFLP's are clustered at the 5′ and 3′ regions of the gene. The 5′-polymorphic cluster [*Bgl*II, *Pvu*II(a), and *Pvu*II(b)] spans 30 kb of DNA and includes exons 1, 2, and 3 and only 2 kb of 5′-flanking DNA. The 3′ cluster [*Eco*RI, *Xmn*I, *Msp*I, *Hin*dIII, and *Eco*RV] spans about 25 kb of the gene (exons 6–13) and 15 kb of 3′-flanking DNA (Figure 6). Two types of RFLP's were detected in this gene. One type, generated by simple base substitution, is due to the presence or absence of a particular recognition sequence for the restriction enzymes *Bgl*II, *Pvu*II, *Xmn*I, *Eco*RI, *Eco*RV, and *Msp*I (Figure 6, heavy arrows). The other type, detected as a variation in fragment length produced by digestion with *Hin*dIII and several other enzymes, has three allelic forms (i.e., 4.4, 4.2, and 4.0 kb) as previously reported (Woo et al., 1983) and differed by 200 bp of inserted DNA (A. G. DiLella, unpublished results).

DISCUSSION

We determined the structural organization of the human chromosomal PAH gene by constructing and characterizing cosmid clones containing over 120 kb of contiguous genomic DNA from the corresponding genetic locus. The protein-coding region of the PAH gene is unusually large. The 13 exons contained in this gene range in size from 57 to 892 bp (mean size 114 ± 43 bp), consistent with published distributions of exon sizes for 20 proteins (Blake, 1983). Introns varied considerably in size, ranging from 1 to 24 kb. The complete gene sequence consists of 2.3 kb of exons and 85 kb of introns, and the ratio of coding to noncoding DNA is among the lowest in eukaryotic genes. Although genes containing large introns are common, there does not appear to be a direct correlation of mRNA size with gene structure. For example, the human factor VIII (Gitschier et al., 1984) and thyroglobulin (van Ommen et al., 1983) genes exceed 100 kb, although these genes code for mRNA transcripts of 7.8 kb. In contrast, the

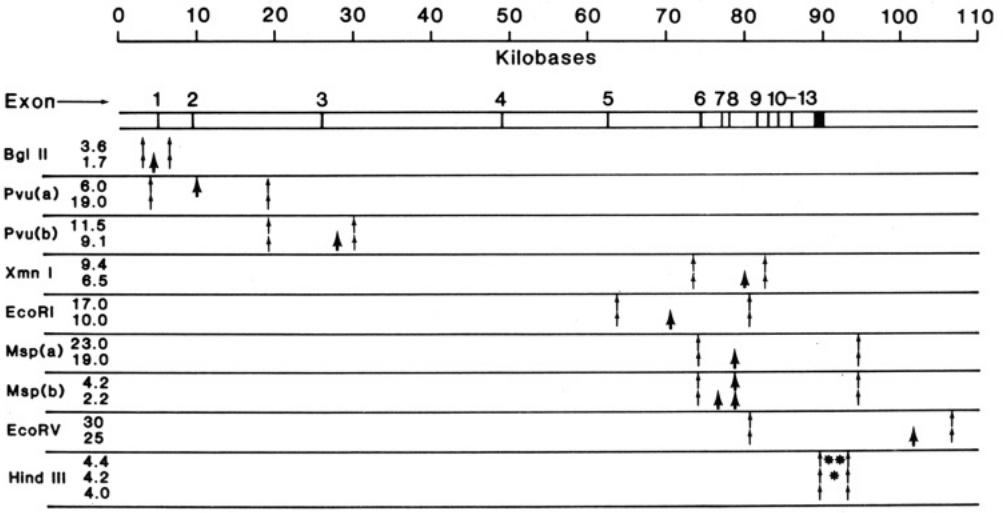

FIGURE 6: Physical map of restriction fragment length polymorphisms in the human PAH gene. The exact locations of the restriction enzyme site polymorphisms were determined by restriction endonuclease mapping of the PAH cosmid clones shown in Figure 1. *Bgl*II, *Pvu*II, *Xmn*I, *Eco*RI, *Msp*I, and *Eco*RV restriction site polymorphisms were generated by the presence or absence of a particular recognition sequence (heavy arrow). The *Hin*dIII fragment length variation has three allelic forms (4.4/4.2/4.0) differing by 200 bp (one asterisk) or 400 bp (two asterisks) of DNA inserted into the prototype 4.0-kb fragment.

vitellogenin gene codes for a 6.6-kb mRNA product and contains 25 introns (mean size 940 bp) spanning only 23 kb (Wilks et al., 1981).

Although it appears that introns are not essential for gene expression (Perler et al., 1980; Wallace et al., 1980), there is evidence that they function in gene assembly. Correlations between the location of introns and features of protein structure have been reported for several genes and argue against the insertional model for the origin of introns (Blake, 1983). We recently reported extensive homology between tyrosine hydroxylase and PAH at the mRNA and amino acid levels (Ledley et al., 1985b), suggesting that these two proteins evolved from a common ancestral gene. Virtually all of the homology between these two enzymes resides within the C-terminal two-thirds of the protein, and it has been proposed that this region contains the common catalytic sequences shared between these proteins (Ledley et al., 1985b). It is interesting that this homologous region in the PAH gene is encoded by exons 6–13 comprising 1698 bp of mRNA, which is clustered in only 16 kb of DNA at the 3′ end of the gene. In contrast, no significant homology exists within the amino-terminal 150 amino acids of tyrosine hydroxylase and PAH. For PAH, this nonhomologous region is encoded by the first five exons comprising only 567 bp of mRNA and which is distributed over 72 kb of DNA at the 5′ end of the gene. It is possible that the amino-terminal regions of tyrosine hydroxylase and PAH may have arisen by recruitment of unrelated exons carrying functional domains. The large introns detected in this region of the PAH gene could be a vestige of this genetic reorganization, which would also involve the recruitment of signals that are involved in the tissue-specific regulation of these genes. The observation that a 23.5-kb intron interrupts the coding sequence between amino acids 117 and 118, i.e., adjacent to the region homologous to tyrosine hydroxylase, lends support to this hypothesis. Future characterization of tyrosine hydroxylase gene structure will provide insight into the evolution and structure/function relationships of these proteins.

All restriction fragments detected by Southern blotting of genomic DNA using the PAH cDNA probe are represented within the cosmid clones and are confined to a single genetic locus, providing strong structural evidence that there are no PAH pseudogenes in the genome. Taken together with previously observed biochemical evidence that the human PAH cDNA clone phPAH247 contains all necessary genetic information to code for normal enzymatic activity (Ledley et al., 1985a), cytogenetic evidence that there is only a single genetic locus containing the PAH gene on the q22–q24.1 region of chromosome 12 (Lidsky et al., 1984), and genetic evidence that the RFLP's detected by the cDNA probe segregated concordantly with the PKU trait (Woo et al., 1983), it can be concluded that there is only a single PAH gene in the human genome and mutations in this genetic locus are the direct cause of PKU. Consequently, all RFLP's detected by the PAH cDNA probe can be confidently used for prenatal diagnosis of the genetic disorder (Lidsky et al., 1985b; DiLella et al., 1985b). In this report, we further demonstrate that there are two clusters of RFLP's in the PAH locus: (1) *Bgl*II–*Pvu*II(a)–*Pvu*II(b) at the 5′ half and (2) *Eco*RI–*Msp*I–*Xmn*I–*Eco*RV–*Hin*dIII at the 3′ half of the gene. For future prenatal diagnosis of PKU, we determined that *Pvu*II, *Xmn*I, and *Eco*RV (detecting RFLP's in the 5′- and 3′-polymorphic clusters) will establish disease status in about 85% of PKU families at risk (unpublished experiments).

We have recently used the RFLP's to define haplotypes of normal and PKU chromosomes in the Danish population. It is most interesting that the majority of PKU genes in the Danish population are confined to four common RFLP haplotypes, two of which are different from the predominant haplotypes of the normal gene (DiLella et al., 1985b). Close associations between RFLP haplotypes of the β-globin locus and specific β-thallasemia mutations in different ethnic populations were previously observed (Kazazian et al., 1984). Whether such tight association between RFLP haplotypes and specific mutations exists in PKU due to linkage disequilibrium is not yet known. Having established the molecular structure of the normal PAH gene, this important issue can now be addressed by cloning and characterization of the mutant alleles of the various RFLP haplotypes.

ACKNOWLEDGMENTS

We thank Drs. Y. F. Lau and Y. W. Kan for cosmid vector pCV107 and *E. coli* ED8767. We thank Kelly Brayton and Susan Sullivan for excellent technical assistance and Kelly Porter for typing the manuscript.

Registry No. Phenylalanine hydroxylase, 9029-73-6.

REFERENCES

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

Bartholome, K., & Ertel, E. (1976) *Lancet 2*, 862.

Bartholome, K., & Dresel, A. (1982) *J. Inherited Metab. Dis. 5*, 7.

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 180.

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513.

Birnstiel, M. L., Busslinger, M., & Strub, K. (1985) *Cell (Cambridge, Mass.) 41*, 349.

Blake, C. (1983) *Nature (London) 308*, 535.

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

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

DiLella, A. G., Ledley, F. D., Rey, F., Munnich, A., & Woo, S. L. C. (1985a) *Lancet 1*, 160.

DiLella, A. G., Ledley, F. D., & Woo, S. L. C. (1985b) in *Biotechnology in Diagnostics: Proceedings of the International Symposium on the Impact of Biotechnology on Diagnostics* (Koprowski, H., Ferrone, S., & Albertini, A., Eds.) pp 295–307, Elsevier Science Publishers, Amsterdam.

Donlan, J., & Kaufman, S. (1980) *J. Biol. Chem. 255*, 2146.

Friedman, P. A., & Kaufman, S. (1973) *Biochim. Biophys. Acta 293*, 56.

Friedman, P. A., Fisher, D. B., Kang, E. S., & Kaufman, S. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 552.

Gitschier, J., Wood, W. I., Goralka, T. M., Wion, K. L., Chen, E. Y., Eaton, D. H., Vehar, G. A., Capon, D. J., & Lawn, R. M. (1984) *Nature (London) 312*, 326.

Grosveld, F. G., Dahl, H. H. M., deBoer, E., & Flavell, R. A. (1981) *Gene 13*, 227.

Kaufman, S. (1976) *Adv. Neurochem. 2*, 1.

Kaufman, S., & Fisher, D. B. (1970) *J. Biol. Chem. 245*, 4745.

Kazazian, H. H., Orkin, S. H., Markham, A. F., Chapman, C. R., Youssoufian, H., & Waber, P. G. (1984) *Nature (London) 310*, 152.

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

Hanahan, D., & Meselson, M. (1980) *Gene 10*, 63.

Lau, Y. F., & Kan, Y. W. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5225.
Ledley, F. D., Grenett, H. E., DiLella, A. G., Kwok, S. C. M., & Woo, S. L. C. (1985a) *Science (Washington, D.C.) 228*, 77.
Ledley, F. D., DiLella, A. G., Kwok, S. C. M., & Woo, S. L. C. (1985b) *Biochemistry 24*, 3389.
Lidsky, A. S., Robson, K. J. H., Thirumalachary, C., Barker, P. E., Ruddle, F. H., & Woo, S. L. C. (1984) *Am. J. Hum. Genet. 36*, 527.
Lidsky, A. S., Ledley, F. D., DiLella, A. G., Kwok, S. C. M., Daiger, S. P., Robson, K. J. H., & Woo, S. L. C. (1985a) *Am. J. Hum. Genet. 37*, 619.
Lidsky, A. S., Güttler, F., & Woo, S. L. C. (1985b) *Lancet 1*, 549.
Messing, J. (1983) *Methods Enzymol. 101*, 20.
Milbrandt, J. D., Azizkahn, J. C., Greisen, K. S., & Hamlin, J. L. (1983) *Mol. Cell. Biol. 3*, 1266.
Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459.
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 (Cambridge, Mass.) 38*, 275.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463.
Scriver, C. R., & Clow, C. L. (1980) *Annu. Rev. Genet. 14*, 179.
Sharp, P. A. (1981) *Cell (Cambridge, Mass.) 23*, 643.
Smith, H. O., & Birnstiel, M. L. (1976) *Nucleic Acids Res. 3*, 2387.
Smith, S. C., Kemp, B. E., McAdam, W. J., Mercer, J. F. B., & Cotton, R. G. H. (1984) *J. Biol. Chem. 259*, 11284.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503.
Wallace, B. R., Johnson, P. F., Tonaka, S., Schold, M., Itakura, K., & Abelson, J. (1980) *Science (Washington, D.C.) 209*, 1396.
Wilks, A., Cató, A. C. B., Cozens, P. J., Mattaj, I. W., & Jost, J. P. (1981) *Gene 16*, 249.
Woo, S. L. C., Gillam, S. S., & Woolf, S. I. (1974) *Biochem. J. 139*, 741.
Woo, S. L. C., Lidsky, A. S., Güttler, F., Chandra, T., & Robson, K. J. H. (1983) *Nature (London) 306*, 151.
van Ommen, G. J. B., Arnberg, A. C., Baas, F., Brocus, H., Sterk, A., Tegelaers, W. H. H., Vassart, G., & de Vijlder, J. J. M. (1983) *Nucleic Acids Res. 11*, 2273.

Affinity Chromatography of the Anterior Pituitary D_2-Dopamine Receptor†

Susan E. Senogles, Nourdine Amlaiky, Anne L. Johnson, and Marc G. Caron*
Departments of Physiology and Medicine, Howard Hughes Medical Institute, Duke University Medical Center, Durham, North Carolina 27710

Received November 25, 1985; Revised Manuscript Received December 23, 1985

ABSTRACT: The D_2-dopamine receptor from bovine anterior pituitary has been solubilized with digitonin and purified ~1000-fold by affinity chromatography on a new affinity support. This support consists of a (carboxymethylene)oximino derivative of the D_2-selective antagonist spiperone (CMOS) covalently attached to Sepharose 4B through a long side chain. The interaction of the solubilized receptor activity with the affinity gel was biospecific. Dopaminergic drugs blocked adsorption of solubilized receptor activity to the CMOS–Sepharose with the appropriate D_2-dopaminergic potency and stereoselectivity. For agonists, (−)-*N*-*n*-propylnorapomorphine > 2-amino-6,7-dihydroxytetrahydronaphthalene ≃ apomorphine > dopamine, whereas for antagonists (+)-butaclamol >> (−)-butaclamol. The same D_2-dopaminergic specificity was observed for elution of receptor activity from the gel. To observe eluted receptor binding activity, reconstitution of the eluted material into phospholipid vesicles was necessary. Typically, 70–80% of the solubilized receptor was adsorbed by CMOS–Sepharose, and 40–50% of the adsorbed activity could be recovered after reconstitution of the eluted material. The overall recovery of D_2-receptor activity from bovine anterior pituitary membranes was 12–15% with specific binding activity of ~150 pmol/mg. The reconstituted affinity-purified receptor bound ligands with the expected D_2-dopaminergic specificity, stereoselectivity, and rank order of potency.

The D_2-dopamine receptor mediates many of the physiological actions of dopamine. It is found most prominently in the pituitary gland and the corpus striatum region of the brain [reviewed in Stoof & Kebabian (1984) and Seeman (1980)]. In the pituitary gland, the D_2 receptor mediates both the dopaminergic inhibition of prolactin release from the mammotrophs (Caron et al., 1978) and the inhibition of release of α-melanocyte-stimulating hormone from cells of the intermediate lobe (Cote et al., 1982), an effect which may be mediated through inhibition of adenylate cyclase activity (Giannattasio et al., 1981; Enjalbert & Bockaert, 1983). Recent evidence suggests that the D_2 receptor may also function via a different signal transfer pathway in the anterior pituitary (Canonico et al., 1982; Schofield, 1983). The ultimate understanding of the molecular events involved in the dopaminergic signal transduction which control hormone secretion will require the eventual purification of the receptor as well as the various other components of the signal transfer system (Sternweis et al., 1981; Bokoch et al., 1984; Codina et al., 1984; Pfeuffer et al., 1985).

Purification of the D_2-dopamine receptor has not yet been accomplished mainly due to the lack of suitable tools to this

† This work was supported in part by a grant from the National Institutes of Health (NS-19576).
* Correspondence should be addressed to this author at the Departments of Physiology and Medicine, Box 3287, Duke University Medical Center.

0006-2960/86/0425-0749$01.50/0