BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **225,** 238–242 (1996)
ARTICLE NO. 1160

# Structure of the Phenylalanine Hydroxylase Gene in *Drosophila melanogaster* and Evidence of Alternative Promoter Usage[1]

Pilar Ruiz-Vázquez, Mireille Moulard, and Francisco J. Silva[2]

*Departament de Genètica, Universitat de València, c/ Dr. Moliner 50, 46100 Burjassot, València, Spain*

Received June 20, 1996

The complete *Drosophila melanogaster* phenylalanine hydroxylase gene isolated from a genomic library was sequenced. Gene structure consisted of five exons covering a region of around 3 kb. Position of introns in the C-terminal domain was conserved with mammalian aromatic amino acid hydroxylase genes. Putative promoter sequences in the 5′UTR and intron 1 were identified. A novel transcript was detected differing from that previously reported by the inclusion of a part of the intron 1 sequence. It could be produced using an alternative promoter. The deduced open reading frame would code a protein with a small difference at the N-terminus. Expression of the alternative transcripts was examined throughout development. 

Phenylalanine hydroxylase (EC 1.14.16.1, PAH) is a member of the pterin-dependent aromatic L-amino acid hydroxylase family to which tryptophan (EC 1.14.16.4) and tyrosine hydroxylases (EC 1.14.16.2) belong. These enzymes share many physical, structural and catalytic properties (1). Sequences of the genes encoding these proteins are available, not only for mammals but for more evolutionary distant species such as *Drosophila.* They exhibit a high degree of similarity both at nucleotide and amino acid levels, reflecting a common origin from an ancestral gene (2,3). This similarity spans the genomic organization (*i.e.* the position of intron/exon boundaries) (2,4). These enzymes are reported to have a domain structure on the basis of their homology. The C-terminal two-thirds of the protein possibly represents the catalytic domain, including the pterin cofactor binding site, while the less conserved N-terminus would consist of a regulatory domain reflecting the unique properties of each enzyme (1).

In *Drosophila melanogaster*, phenylalanine hydroxylase (*Pah*) cDNA has been isolated and sequenced (5,6), although its genomic structure has not been determined. It has been proposed that the same gene encodes both phenylalanine and tryptophan hydroxylase activities in this organism (6,7) providing additional support for a model of gene duplication. *Pah* locus involvement in the pteridine metabolism has also been shown (7,8,9). This last function has been suggested since the *Henna* locus, involved in the synthesis of pteridines, encodes PAH (10). Moreover, in mammals, purified PAH has been shown to function as a tetrahydropterin oxidase (11).

We report here the structure of the *Drosophila Pah* gene and demonstrate the existence of alternative transcripts, which would generate two isoform proteins.

## MATERIAL AND METHODS

*Isolation of genomic clones.* A full-length *Drosophila Pah* cDNA clone (a gift of Dr. C. Alonso) was used as a probe to screen a *Canton-S* strain genomic library prepared in EMBL4 (a gift of Dr. M. Pérez-Alonso).

---

[1] Sequence data from this article have been deposited with the EMBL/GeneBank Data Libraries under Accession number X98116.

[2] To whom correspondence should be addressed. Fax: 34-6-3983029.



0006-291X/96 $18.00

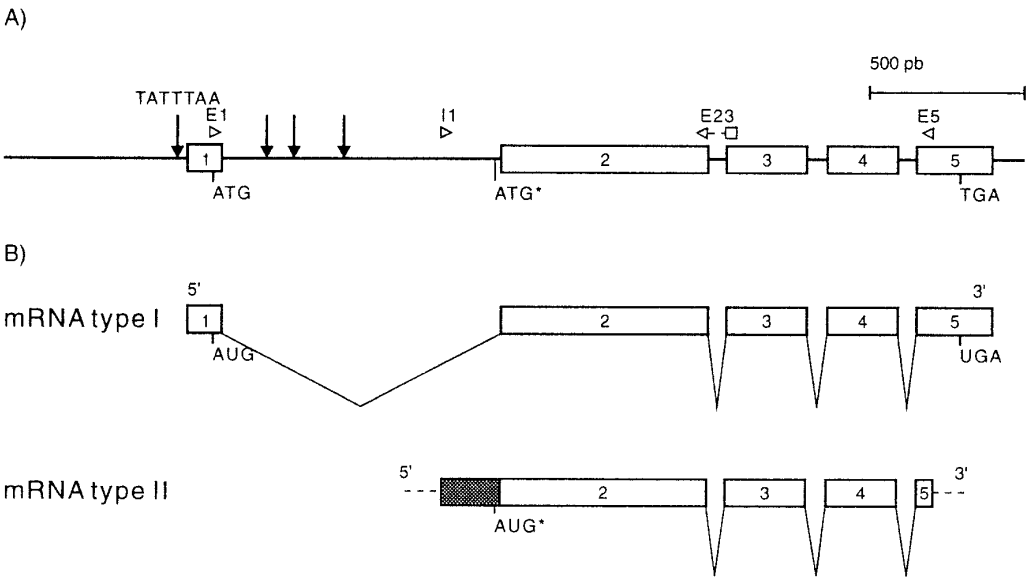


**FIG. 1.** (A) Genomic structure of the *Pah* gene determined by comparing the sequence of the genomic clone herein described and the cDNA clones previously reported (5,6). (B) Structure of the *Pah* transcripts detected in *Drosophila* adult heads. Exons are represented by boxes. Initiation (ATG) and stop (TGA) codons are indicated. The TATA-box-like (TATTTAA) sequence observed in the 5′-untranscribed region is indicated. Putative initiator codon (ATG*) and promoter sequences (vertical arrows) found in intron 1 are also shown. Grid box represents the novel exonic sequence in mRNA type II. Locations of primers used for RT-PCR are indicated (horizontal arrowheads).

*Sequencing.* Sequencing was performed over both strands using the Applied Biosystems 373A Automated DNA Sequencer (SCSI, Universitat de València, Spain).

*RNA extraction.* Around 40 mg of insect sample were frozen in liquid nitrogen and reduced mechanically in a fine slurry. RNA extraction was performed using the Microscale-Total RNA Separator Kit from Clontech. RNA purity and yield was determined by absorbance.

*Generation and analysis of RT-PCR products.* The oligonucleotides used as primers were: **E1** (*Pah* sense exon 1), 5′GAATCTTGTGGAATTCTACCAGCGG; **I1** (*Pah* sense intron 1), 5′GATTTGCTGATAGACATTGC; **E23** (*Pah* antisense spanning junction exon 2 and 3) 5′GTGAAACCTGTGCAATCCC; **E5** (*Pah* antisense exon 5), 5′TCTAGA-ACCTCAACACTCTG. cDNA synthesis was performed using the First-strand cDNA Synthesis kit from Pharmacia. DNAs were amplified 30 cycles in 50 $\mu$l reactions containing 10 pmol of each primer using a GeneAmp PCR System 2400 thermocycler (Perkin Elmer).

## RESULTS AND DISCUSSION

Genomic DNA encoding PAH was identified by screening a *Canton-S* genomic library using the entire *Pah* cDNA of *Drosophila* as a probe. Three positive clones which contained the entire *Pah* gene were isolated. After subcloning in pUC18, the genomic region homologous to the *Pah* cDNA was sequenced. Gene structure was deduced by comparison with its cDNA (5,6). The complete *Drosophila Pah* gene was approximately 3 kb in length and consisted of a total of five exons interrupted by four introns (Fig. 1A). The sequences of the exon/intron boundaries are shown in Fig. 2. The 5′ and 3′ splice sites were in line with the consensus sequences in *Drosophila* (12) showing the compulsory GT...AG dinucleotides. Putative branch-point sequences were found in all four introns. No polyadenylation signal other than the previously described AATAAA (5,6) was identified within the 409 bp downstream from the stop codon TGA. Although the number of exons was higher in the human gene (13 exons compared with 5 in *Drosophila*), location of *Drosophila* introns 2, 3 and 4 were accurately conserved with respect to that of introns 6, 9 and 11 from the human *Pah* sequence. This

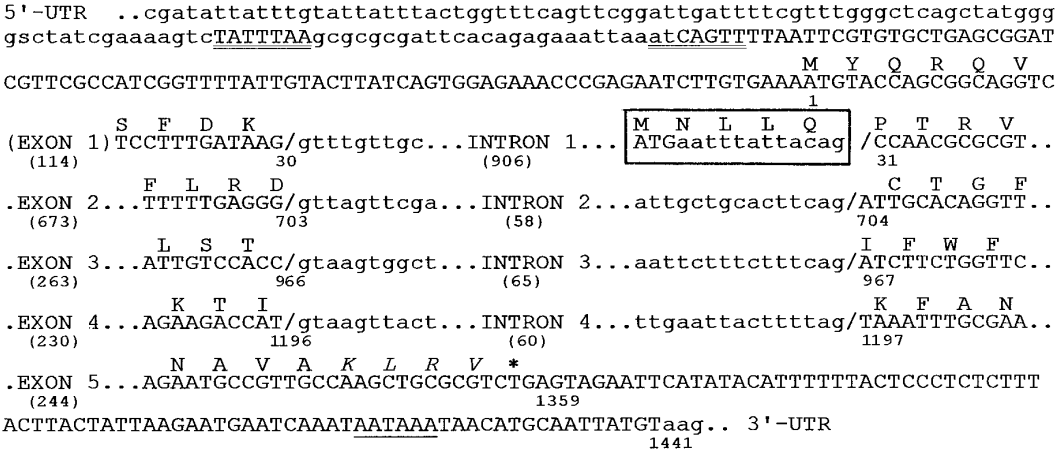


**FIG. 2.** Partial nucleotide and deduced amino acid sequences of the *Drosophila Pah* gene. Nucleotide sequence is numbered with the first base of the ATG initiation codon designated as position +1 (nucleotides within introns are not numbered). Exon/intron boundary sequences are given (exons in uppercase letters; introns in lowercase). Sizes of both exons and introns are indicated between brackets. Polyadenylation signal is underlined. Putative TATA-box detected in 5′-untranscribed region (5′-UTR) and Cap addition site are doubly underlined. Putative ATG start codon observed in intron 1 and the corresponding amino acid sequence are boxed. The herein deduced ORF was coincident with the previously reported (6), except for a conservative change in amino acid 332 (Val→Leu) and for the last C-terminal amino acids (italics) due to the presence of an extra nt (position 1357). This extra nucleotide was previously reported (5).

perfect conservation extended to other members of the aromatic amino acid hydroxylase gene family (2,4). This was consistent with a common evolutionary origin for the members of this family. The conserved exon/intron boundaries were placed in the C-terminal two-thirds of the gene, the catalytic domain.

Potential promoter elements were identified within the 931 bp upstream from the ATG translation start codon. A putative TATA-box, characteristically surrounded by GC-rich sequences, was appropriately located 34 bp upstream from the longer cDNA isolated (5). The sequence TATTTAA perfectly matched with the putative promoter of the *Drosophila* tyrosine hydroxylase gene (*Tyh*) (13) as well as with the alcohol dehydrogenase gene promoter active during the adult stage (14). In addition, a perfect Cap-site sequence at 32 bp downstream from this putative promoter could be identified, suggesting that it could be a possible transcription start site of the *Pah* gene.

Intron 1 was longer than the average length of the *Drosophila* introns (12). Long introns have been related to the presence of functional elements other than those needed for splicing. Some putative TATA boxes have been identified within intron 1. An in-frame putative start codon ATG was also located 15 bp upstream from exon 2 junction. These observations suggested the possibility that an alternative transcript initiated in intron 1 could be produced. To test whether such transcript occurred, RT-PCR experiments were performed. Using I1 and E5 (shown in Fig. 1), oligo(dT)-primed cDNAs from total RNA of adult heads gave a fragment of around 1300 bp. Sequence analysis demonstrated that this fragment corresponded to a *Pah* mRNA that included intron 1 sequences (Fig. 1B). We referred to this novel transcript as mRNA type II *versus* the previously reported (5,6) referred as *Pah* mRNA type I. The usage of alternative promoters seemed to be the origin of this alternative expression. The two *Pah* transcripts would code two isoform proteins of 452 (type I) and 447 (type II) amino which only differed in the first N-terminal amino acids acids (see Fig. 2).

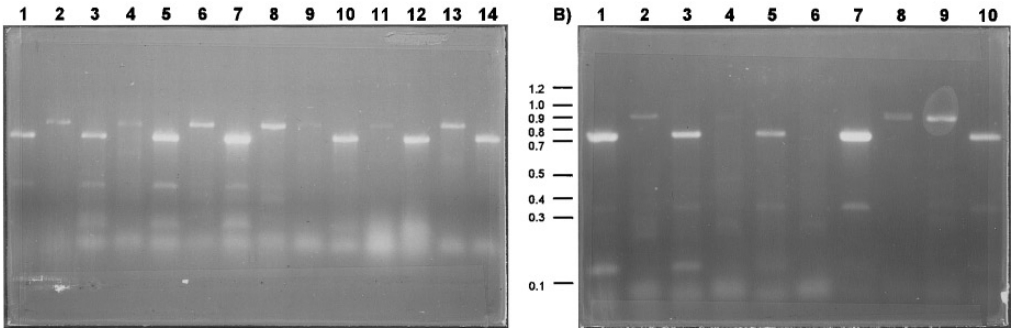


**FIG. 3.** RT-PCR analysis during *Drosophila* development. Detection of mRNA type I was carried out with primers E1 and E23 (A, lanes 1, 3, 5, 7, 10, 12, and 14; and B, lanes 1, 3, 5, 7 and 10). Detection of mRNA type II was carried out with primers I1 and E23 (A, lanes 2, 4, 6, 8, 9, 11, and 13; and B, lanes 2, 4, 6, 8, and 9). Size of the amplified fragments was 732 and 885 pb, respectively. (A) 0-6 hours old embryos (lanes 1, 2), 6-12 h embryos (3, 4), 12-18 h embryos (7, 8), first instar larvae (9, 10), second instar larvae (11, 12), and third instar larvae (13, 14). (B) White pupa (lanes 1, 2), two-day-old pupae (3, 4), three-day-old pupae (5, 6), four-day-old pupae (7, 8), and 0-48 h adult heads (9, 10).

To test the presence of both *Pah* transcripts throughout development, RT-PCR was carried out. Total RNA from different stages was reverse transcribed using E23 as primer. cDNAs were amplified using either E1 or I1 (specific for mRNA type I and type II, respectively) in combination with E23 (it did not anneal to genomic DNA as it spanned exons 2 and 3). While transcript type I appeared in all the analyzed stages, type II was not detected in two- and three-day-old pupae and was scarcely presented in the first and second larval instars (Fig. 3). This indicated a regulation of the *Pah* expression along development. Quantitative analyses will provide a more accurate interpretation.

In several organisms alternative transcription of the aromatic amino acid hydroxylase genes has been reported. Two rat *Pah* mRNAs derived from alternative polyadenylation have been detected (15). The presence of alternative transcripts for the *Tyh* gene was also reported in several species (13,16). In the case of the *Drosophila Tyh*, the three detected transcripts would code for two isoform proteins, one predominant in central nervous system while the other in the hypoderm. This tissue-specific distribution would be related with the two physiological functions proposed for tyrosine hydroxylase: neurotransmission and cuticle formation (13). The occurrence of alternative transcription of the *Drosophila Pah* gene herein described could be related to the three proposed functions of the protein: (i) Phe hydroxylation, required for Phe concentration control, and as a source of endogenous tyrosine for the sclerotization process (17); (ii) Trp hydroxylation, required for serotonin synthesis (6,9); (iii) Tetrahydropterin oxidation, required for eye-pigment synthesis (7). Further analyses will be required to confirm this hypothesis regarding the effect of the alternative transcripts on PAH activity and regulation, and the biological relevance of these observations.

## ACKNOWLEDGMENTS

We are grateful to Dr. C. Alonso, (Universidad Autónoma de Madrid, Spain) for the gift of the *Pah* cDNA and to Dr. M. Pérez-Alonso (Universitat de València, Spain) for the gift of the *Canton-S* strain genomic library. Support for M. Moulard and P. Ruiz-Vázquez was provided by grants from the Human Capital and Mobility Network program and the Ministerio de Educación y Ciencia, respectively. This work has been supported by the Commission of the European Communities (Contract ERBCHRXCT93-0243).

## REFERENCES

1. Hufton, S. E., Jennings, I. G., and Cotton, R. G. H. (1995) *Biochem. J.* **311,** 353–366.
2. Grenett, H. E., Ledley, F. D., Reed, L. L., and Woo, S. L. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 5530–5534.

3. Ledley, F. D., DiLella, A. G., Kwok, S. C. M., and Woo, S. L. C. (1985) *Biochemistry* **24,** 3389–3394.
4. Stoll, J., and Goldman, D. (1991) *J. Neurosci. Res.* **28,** 457–465.
5. Morales, G., Requena, J. M., Jiménez-Ruiz, A., Lopez, M. C., Ugarte, M., and Alonso, C. (1990) *Gene* **93,** 213–219.
6. Neckameyer, W. S., and White, K. (1992) *J. Biol. Chem.* **267,** 4199–4206.
7. Alcañiz, S., Ruiz-Vázquez, P., and Silva, F. J. (1995) *Pteridines* **6,** 135–137.
8. Silva, F. J., Bel, Y., Botella, L. M., Cotton, R. G. H., and Ferré, J. (1992) *Biochem. J.* **287,** 85–89.
9. Alcañiz, S., and Silva, F. J., in preparation.
10. Ruiz-Vázquez, P., and Silva, F. J., in preparation.
11. Davis, M. D., and Kaufman, S. (1989) *J. Biol. Chem.* **264,** 8585–8596.
12. Mount, S. M., Burks, C., Hertz, G., Stormo, G. D., White, O., and Fields, C. (1992) *Nucl. Acids Res.* **20,** 4255–4262.
13. Birman, S., Morgan, B., Anzivino, M., and Hirsh, J. (1994) *J. Biol. Chem.* **269,** 26559–26567.
14. Benyajati, C., and Dray, J. F. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1701–1705.
15. Dahl, H. H., and Mercer, J. F. (1986) *J. Biol. Chem.* **261,** 4148–4153.
16. Nagatsu, T., and Ichinose, H. (1991) *Comp. Biochem. Physiol.* **98C,** 203–210.
17. Piedrafita, A. C., Martínez-Ramírez, A., and Silva, F. J. (1994) *Insect Biochem. Molec. Biol.* **24,** 581–588.