EVIDENCE FOR EVOLUTIONARY DUPLICATION OF GENES IN THE DOPA DECARBOXYLASE REGION OF DROSOPHILA

DAVID D. EVELETH AND J. LAWRENCE MARSH

Developmental Biology Center and Department of Developmental and Cell Biology, University of California, Irvine, California 92717

Manuscript received December 31, 1986
Revised copy accepted June 14, 1986

ABSTRACT

The region surrounding the dopa decarboxylase gene (Ddc) of Drosophila contains a cluster of genes, many of which appear to be functionally related by virtue of their effects on cuticle development and/or catecholamine metabolism. In this report we describe evidence that the Ddc gene and the closely linked alpha-methyldopa hypersensitive (amd) gene share extensive sequence homology and are the products of a gene duplication event. The two genes are transcribed convergently and are separated by 2.4 kb. A gene located between Ddc and amd expresses a 2.0-kb mRNA and appears to partially overlap the Ddc gene. The organization of these transcripts implies a complex series of events giving rise to the present pattern. The patterns of expression of these genes do not support a model of coordinate regulation, but are more consistent with a pattern of duplication and divergence to various related metabolic subspecialties. These data provide the first evidence for structural relationships among genes in the 37C cluster.

The dopa decarboxylase (Ddc) region of Drosophila (i.e., Df(2L)TW130, Wright, Hodgetts and Sherald 1976) contains a cluster of 18 genes, of which at least 14 are thought to be functionally related by virtue of their effects on cuticle development and catecholamine metabolism (e.g., Wright et al. 1981; Pentz and Wright 1986); yet, Gilbert, Hirsh and Wright (1984) were unable to detect any DNA sequence homology within this region, and they concluded that the genes were not structurally related.

We have focused our attention on the Ddc and closely linked (0.002 cM) alpha-methyldopa hypersensitive gene (l(2)amd, abbreviated amd). Several observations suggest that the gene product of the amd locus is related to the DDC enzyme and involved in catecholamine metabolism (Wright, Bewley and Sherald 1976; Marsh and Wright 1986). First, the recessive lethal phase of both Ddc and amd is embryonic hatching, with both amd and Ddc embryos showing abnormal cuticles (Sparrow and Wright 1974; Wright 1977). Second, the amd+ gene product confers resistance to dietary administration of dopa analogues, specifically alpha-methyldopa (Marsh and Wright 1986),

FIGURE 1.—Organization of transcripts in the Ddc region. Transcripts discussed in this communication are placed on the genomic map. We have used the HpaI site near the terminus of Ddc as a fixed reference point at 0 kb. This is different than the arbitrary zero point used by other authors (e.g., Gilbert, Hirsh and Wright 1984), which would lie at a point slightly greater than 860 bp on our map. The amd transcript is described by Marsh, Erfle and Leeds (1986). The arrow represents the 2-kb transcript described by Spencer, Geitz and Hodgetts (1986). The Ddc transcript is positioned according to Eveleth et al. (1986). The Cc transcript is described in Eveleth and Marsh (1986). Solid bars indicate exon material included in the mature mRNA or cDNA. Arrows indicate direction of transcription from 5' to 3'. In this figure, the centromere is to the right.

whereas DDC activity is inhibited in cell-free extracts by the same dopa analogues (Sparrow and Wright 1974). Third, temperature-sensitive mutants of Ddc shifted to restrictive temperature during the larval stages and larvae fed dopa analogues die at the larval molts or at pupariation and exhibit abnormal pupal cuticles (Wright 1977; Wright et al. 1982). Fourth, lethal mutations of the amd gene can be rescued by dietary supplements of dopa, tyramine and octopamine and pyridoxal-5'-phosphate (PLP) (P. D. L. Gibbs and J. L. Marsh, unpublished results). Taken together, these observations strongly suggest that the Ddc and amd genes both encode enzymes involved in the catecholamine biosynthetic pathway and that these gene products both recognize substrates of similar structure.

In addition to the Ddc and amd genes, several other genes affecting cuticle formation or catecholamine metabolism are located in the Df(2L)TW130 region including the diphenol oxidase gene (Dox-A2) (Pentz and Wright 1986). These and other observations have led to the hypothesis that Ddc, amd and possibly other genes in this cluster may be structurally and evolutionarily related (Marsh and Wright 1979, 1986). We have investigated the origin and physical organization of the transcription units surrounding the Ddc gene by direct sequence analysis. This has revealed that the 12-kb interval including the Ddc gene also includes three other transcripts (Figure 1). These transcripts are extremely closely spaced, and the Ddc and Cs transcripts (Spencer, Geitz and Hodgetts 1986) actually overlap. In this report, we show that the Ddc and amd genes are structurally and evolutionarily related. Based on these observations, we propose that the Ddc and amd genes represent a structurally related gene pair (a paralogous set) that arose by duplication and divergence. These data provide the first evidence for genes structurally related to the Ddc
gene in any organism and document that genes in the Ddc region of Drosophila are structurally related by gene duplication. These observations support the interpretation of genetic studies that this region contains a cluster of functionally (and now structurally) related genes.

MATERIALS AND METHODS

The sequences used in this analysis are derived from clones originating from the MANIATIS library (MANIATIS et al. 1978). The restriction pattern of the cloned DNA resembles the Canton-S haplotype (MARSH and WRIGHT 1986). The sequence of the amd gene is described in the accompanying paper (MARSH, ERFLE and LEEDS 1986). The sequence of the l(2)37Cc gene is described by EVELETH and MARSH (1986), and the Ddc gene structure and sequence by EVELETH et al. (1986). The sequence of the region between Ddc and amd that includes the Cs transcript (SPENCER, GEITZ and HODGETTS 1986) was determined by D. D. EVELETH and J. L. MARSH (unpublished observations). Computer analyses of DNA sequences were performed on an IBM-PC employing a series of programs modified from SCHWINDINGER and WARNER (1984) by A. GOLDIN (California Institute of Technology). Dot matrix homology searches employed a set of programs written by B. WARD and G. GUTMAN (Microbiology and Molecular Genetics Department, University of California, Irvine).

RESULTS

Four genes have been mapped within the 12-kb area surrounding the Ddc gene. These include the amd gene (MARSH, ERFLE and LEEDS, 1986), the Ddc gene (HIRSH and DAVIDSON 1981; EVELETH et al. 1986), the l(2)37Cc (Cc) gene (EVELETH and MARSH 1986) and an additional transcript (designated Cs) mapping to the region between amd and Ddc (SPENCER, GEITZ and HODGETTS 1986). The structure and genomic organization of these transcripts is shown in Figure 1.

Ddc and amd show sequence homology: We sought to determine whether any of the genes in the region were structurally related by direct sequence comparison. Using computerized dot matrix analysis, we compared the sequences of each gene to the other three in both orientations. A striking homology was found when the Ddc and amd gene regions were compared in their respective directions of transcription. Figure 2 presents a dot matrix analysis comparing the amd gene to a portion of the Ddc gene that includes exons I1 and I11 and the second intron of Ddc. The sequences were scanned for 67% homology with a window size of 15 bases and required match of ten. Regions of sequence similarity that meet this criteria are seen as diagonal rows of dots on the matrix (each dot represents a block of 15 bases, ten of which match). Clear regions of sequence homology (seen as diagonal lines in Figure 2) are seen in the major exons of both genes, showing that the Ddc and amd genes are structurally related and form a gene pair. Substantial sequence similarity is seen even when the search criteria are raised to 90% (matrix not shown).

Intron sequences and positions are not conserved: Interestingly, sharp discontinuities in the alignment are seen at the intron exon borders of both genes. We have detected no homology between the intron sequences of either amd or Ddc and any portion of the sequenced region. If one examines the homology between Ddc and the sequences flanking the amd intron borders, there is clear
homology with an uninterrupted portion of the Ddc coding region even when examined at higher stringencies, such as at the 73% homology level (11 of 15 match) (matrix not shown). The major regions of homology with the Ddc coding region define the borders of the amd intron exactly.

The Ddc gene produces several transcripts that share a major 1031-base pair (bp) intron (Geitz and Hodggetts 1985; Eveleth et al. 1986). A comparison of the sequences flanking this intron reveals that the 5' untranslated region of the early embryonic amd transcript shows extensive homology to the sequences that border this major Ddc intron. This is illustrated by an expanded dot matrix comparison at the 73% homology level (11 of 15 match) between the amd cDNA sequence and the spliced Ddc sequence (Figure 3) that shows a line of homology extending straight through the splice junction of the Ddc intron.
FIGURE 3.—Alignment around the Ddc introns. The alignment between the amd sequence and the sequences around the borders of the Ddc introns is shown. The matrix is as described in legend of Figure 2, except that the scan was performed at a higher stringency requiring that 11 of 15 bases (73%) match. The amd sequence includes the 5′ untranslated leader and the beginning of the coding region of the early embryonic amd transcript. The Ddc sequence represents a partially processed transcript with the 1031-bp intron removed (indicated by a wedge), but the region of the alternative exons is shown unprocessed. This permits comparison of both forms of the Ddc mRNA (EVELETH et al. 1986). The regions of homology that were seen in the lower stringency scan of Figure 2 are designated. Regions ① and ②, which were split by the intron in Figure 2, are now juxtaposed. While much of the background has been filtered out in this higher stringency scan, the region of homology with the 2.0-kb Ddc transcript, which is difficult to resolve in Figure 2, is expanded and indicated by the bracket labeled ③ in both figures.

(i.e., the broken diagonal line indicated by arrows in Figure 2 is continuous in Figure 3).

Recent analysis of the Ddc gene (EVELETH et al. 1986) has revealed two possible splicing modes near the second exon that may lead to two isoforms of the DDC enzyme. These splicing alternatives are shown graphically on the axes of Figures 2 and 3. One of the splicing modes involves the excision of a 77-base intron. Close inspection of the Ddc and amd sequences at high stringency (i.e., 73% in Figure 3) reveals only limited homology to the 77-base intron of Ddc (Figure 3), even though this is thought to encode one of the DDC isoforms. Regions of limited but significant homology between the coding regions on either side of this intron and the 5′ untranslated leader of amd are apparent.

Sequence homology is greatest in two regions: In Table 1 the amd cDNA sequence is aligned with portions of the Ddc genomic sequence. For purposes of illustration, we have used a partially processed Ddc sequence that has the
The nucleotide sequence of the *amd* cDNA is aligned with that of the partially processed *Ddc* sequence (see legend to Figure 3). The upper line shows the *Ddc* sequence. In the lower lines, nucleotides of the *amd* sequence that match are indicated by a dash (−). Where bases fail to match, the corresponding base is printed. Some gaps have been inserted for sequence alignment. The locations of the 480-bp *amd* intron and the 103-bp *Ddc* intron are indicated by wedges. The boundaries of the optional 77-bp *Ddc* intron are delimited by flags pointing into the intron. There is another 5′ untranslated exon of *Ddc* (not shown) that lies 770 bp upstream.
1031-bp intron removed but the region containing the alternative splicing events intact, thus permitting comparison of both forms of Ddc mRNA. This alignment illustrates the degree of local variation in the homology between these two genes. Overall, approximately 55% of the bases match between the two sequences (excluding intron regions); however, two areas of more extensive homology are apparent. One area beginning near the second exon of amd is over 80% homologous over 100 bp, and a second run of 124 bp (700 bp from the 3′ end) is approximately 90% homologous. Although more liberal use of gaps would permit additional alignments, we have attempted to minimize the use of gapping. The locations of introns are indicated in the alignment, thus permitting visualization of the exact nucleotide alignment in the intron regions.

Protein sequence and structure are conserved: In light of the DNA sequence homology, we compared the deduced amino acid sequence of the amd and Ddc proteins. The alignment shown in Table 2 shows two blocks of considerable amino acid sequence conservation. To ascertain whether these proteins might have retained regions of secondary structure that were not apparent from the amino acid alignment, we compared the proteins for common structural features. We have used the method of GARNIER, OSGUTHORPE and ROBSON (1978) to predict the distribution of potential alpha helices and beta sheets. Regions of structural similarity are observed corresponding primarily to the regions of conserved amino acid sequence. However, a high density of potential alpha helical structures was noted near the carboxy terminus of both proteins in a region of only limited amino acid conservation. Thus, the amd and Ddc proteins appear to have retained considerable structural, as well as sequence, similarity.

Homology among other genes: Extensive dot matrix analysis of the region surrounding Ddc does not reveal any similarity between the Ddc or amd genes and the Cc or Cs transcripts, nor is any relationship between the Cc and Cs transcripts seen. While structural relationships that are not detected by our methods may exist between these genes, these relationships must be limited in length or be much lower in homology than are the relationships between Ddc and amd.

Homology with other published sequences: We have scanned the National
The deduced amino acid sequence of the *amd* protein is compared to that of the DDC protein. The upper line shows the DDC sequence. In the lower line, amino acids of the *amd* protein that match are indicated by a dash (−). Where residues fail to match, the amino acid is printed below. Some gaps have been inserted for sequence alignment. We have used the sequence of DDC isoform I protein (Eveleth et al. 1986), although the differences between these two potential isoforms are restricted to the region that does not overlap the *amd* protein; thus, choice of isoform does not affect the alignment.

Institutes of Health DNA sequence data base for genes related to *Ddc* and *amd*. We have scanned both with the entire protein coding sequence, as well as with two strings of approximately 200 bp from the two major regions of sequence homology between these two genes. Using the default parameters of the Bionet IFIND program based on the algorithm of Wilbur and Lipman (1983), we identified no significant homologies among the vertebrate or Drosophila sequences. The similarity score for the *amd* and *Ddc* genes compared with each other is 150, whereas the next highest score on any scan was 20 (a score of 8–14 is expected between random sequences). This, perhaps, is not unexpected since this is the first *Ddc* gene sequence available.

The amino acid sequence of the PLP binding domain of pig DDC has been determined (Bossa et al. 1977). The Drosophila DDC protein contains a precisely homologous domain (amino acids 335–342). The corresponding region of the *amd* gene is highly diverged (Table 2). Analysis of PLP binding domains in a series of PLP enzymes that act on different substrates reveals little conservation (each has a PLP binding lysine) (Tanase, Koijma and Morino 1979). However, the observation that dietary PLP can rescue lethal alleles of *amd* (P. D. L. Gibbs and J. L. Marsh, unpublished observations) suggests that the *amd* sequence is compared to that of the *Ddc* protein. The upper line shows the *Ddc* sequence. In the lower line, amino acids of the *amd* protein that match are indicated by a dash (−). Where residues fail to match, the amino acid is printed below. Some gaps have been inserted for sequence alignment. We have used the sequence of *Ddc* isoform I protein (Eveleth et al. 1986), although the differences between these two potential isoforms are restricted to the region that does not overlap the *amd* protein; thus, choice of isoform does not affect the alignment.

The amino acid sequence of the PLP binding domain of pig DDC has been determined (Bossa et al. 1977). The Drosophila DDC protein contains a precisely homologous domain (amino acids 335–342). The corresponding region of the *amd* gene is highly diverged (Table 2). Analysis of PLP binding domains in a series of PLP enzymes that act on different substrates reveals little conservation (each has a PLP binding lysine) (Tanase, Koijma and Morino 1979). However, the observation that dietary PLP can rescue lethal alleles of *amd* (P. D. L. Gibbs and J. L. Marsh, unpublished observations) suggests that the *amd* sequence is compared to that of the *Ddc* protein. The upper line shows the *Ddc* sequence. In the lower line, amino acids of the *amd* protein that match are indicated by a dash (−). Where residues fail to match, the amino acid is printed below. Some gaps have been inserted for sequence alignment. We have used the sequence of *Ddc* isoform I protein (Eveleth et al. 1986), although the differences between these two potential isoforms are restricted to the region that does not overlap the *amd* protein; thus, choice of isoform does not affect the alignment.
gene product may bind PLP. Thus, if the amd product fulfills a diverged biosynthetic function, such as an amino transferase activity, the PLP binding domain might be quite different.

**DISCUSSION**

**Relationship of Ddc and amd:** The Ddc and amd genes are thought to be functionally related by virtue of their interaction with structural analogues of dopa, thus leading to the speculation that the amd gene may encode a catalytically active product that recognizes substrates similar to DDC (MARSH and WRIGHT 1979, 1986). Enzymes that recognize similar substrates might be expected to have similar active sites. Evolution of enzymes within a metabolic pathway could occur by gene duplication, followed by divergence to metabolic subspecialties [e.g., paralogous gene families as in bacteria (YEH and ORNSTON 1980) or globins (EFSTRATIADIS et al. 1980)]. The comparison of the Ddc and amd sequences presented here clearly documents extensive homology between these genes.

Detailed alignment of the DNA sequence (Table 1) shows two areas of extensive homology between Ddc and amd. One area beginning near the second exon of amd is over 80% homologous over 100 bp, and a second run of 124 bp (700 bp from the 3' end) is approximately 90% homologous. These regions of strong homology reflect regions of conserved amino acid sequence and structure and may reflect functionally conserved portions of the peptides. The DDC and amd proteins do share at least two features. The DDC enzyme functions as a homodimer (CLARK et al. 1978), and genetic evidence suggests that the amd protein may also function as a dimer (MARSH and WRIGHT 1986). Thus, one of the conserved protein domains may mediate dimerization, whereas the other may form part of the active site.

An amino acid comparison of the regions of the Ddc and amd genes that are homologous shows that only 170 of 428 (38%) amino acids are identical. At this level of divergence, the application of molecular clocks is dubious, particularly as the Ddc and amd genes are a paralogous (fulfilling different enzymatic functions), rather than an orthologous (fulfilling identical functions), set.

The intron of the amd gene exhibits no evidence of sequence similarity with any part of the Ddc gene; yet, the regions of homology with Ddc extend to the precise borders of this intron on both the donor and acceptor side. The almost perfect alignment of the Ddc coding sequence across the amd intron suggests that this intron was added to the amd gene (or removed from Ddc) after the duplication event. Similarly, no portion of the major 1031-base Ddc intron shows any sequence similarity with any portion of the amd gene; yet, the 5’ untranslated sequence of amd shows strong homology to the protein coding regions on either side of this Ddc intron. This finding suggests that this intron was also added after the duplication event that gave rise to these two genes. Although the homology is limited, the lack of substantial homology to the 77-base intron of the 2.3-kb transcript of Ddc (EVELETH et al. 1986), while
regions flanking this intron show greater similarity, suggests that this intron may also have been added to Ddc (or removed from amd) after the duplication.

One can propose two models for the genesis of this pattern: (1) Different introns were added to the ancestral duplicates of these genes, but in different locations; (2) the original duplicated gene had all the introns of both genes (and maybe more), and the amd intron was lost giving rise to the present Ddc gene, whereas the 1031-base intron of Ddc (and possibly the 77-base intron) were lost in giving rise to the amd gene. Although intron positions in homologous genes are usually conserved [e.g., globins (Maniatis et al. 1980), vitellogenin genes (Wahl et al. 1980), ovalbumin X-Y (Royal et al. 1979)], some examples have been noted in which introns occur at new positions [e.g., actin (Fyrberg et al. 1981)] or are deleted (Gilbert 1985). The complete lack of homology between the intron borders of either the amd or Ddc introns and any portion of the corresponding gene argues strongly against intron migration. Although precise deletion of introns has been documented (Perler et al. 1980) and is a formal possibility for the origin of different intron patterns (Crabtree et al. 1985), an intron deletion model implies that the primordial genes must have had a very large number of introns (at least all those found in contemporary genes) (Sharp 1985). Thus, we favor the view that the introns of Ddc and amd have been added to these genes after the duplication event.

It has been proposed (e.g., Gilbert 1978; Duester, Jornvall and Hatfield 1986) that introns separate functional domains of proteins, thus enhancing the shuffling of exons to produce novel gene products. In the case of the Ddc and amd genes, no evidence of exon shuffling in the body of either gene is present. Although information regarding the boundaries of functional domains in the Ddc and amd proteins is slight, there is no evidence that the existing introns delineate such domains.

Time of expression: The genes in this cluster exhibit different patterns of expression. The amd transcript begins accumulating on polysomes after gastrulation and reaches a maximum in the later stages of embryogenesis, from approximately 12 to 16 hr (Marsh, Erfle and Leeds 1986). Very low levels of the 2.0-kb amd mRNA are observed in adults, and little or none of the 2.0-kb mRNA can be detected in third instar larvae. The Cs transcript (Spencer, Geitz and Hodgetts 1986) is transcribed in the same orientation as amd; it is most abundant in adult males and very early embryos and decreases in amount during the first 8 hr of embryogenesis, thus exhibiting almost the opposite pattern of expression as the amd transcript. The Ddc transcripts begin accumulating only after 12 hr of embryogenesis and reach a maximum at about 18 hr (Geitz and Hodgetts 1985). Transcription is reinitiated at pupariation and just before adult eclosion (it is also assumed to peak at each of the molts). The Cc transcript upstream of Ddc is present in both early and late stages of embryogenesis, as well as at pupariation and in adults (Eveleth and Marsh 1986). Thus, the transcripts in this region do not represent a cluster of coordinately expressed genes, as might have been expected from a chromosomal domain hypothesis of expression (Weintraub 1985). Rather, the Ddc and amd
genes appear to serve different metabolic subspecialties of catecholamine metabolism and to have evolved separate control features.

The four transcripts discussed here lie within 12 kb of genomic DNA, and the primary transcripts from these genes account for over 10.5 kb after processing for addition of the poly-A tail. If transcription termination in Drosophila occurs as much as 1 kb downstream of the poly-A addition processing site (e.g., CITRON et al. 1984), the gene arrangement seen here implies that the primary transcripts from these genes actually overlap one another, and in fact, the poly-A addition signals for the Ddc and Cs transcripts overlap by about 80 bp (D. D. EVELETH and J. L. MARSH, unpublished observations). If the Ddc gene is located within the large 37C1 salivary gland chromosome band (WRIGHT et al. 1981; GILBERT, HIRSH and WRIGHT 1984), our data imply that at least four genes are contained within this band. We would find it surprising, in the absence of selection, for each of these genes to remain functionally intact through the chromosomal rearrangements that gave rise to the present arrangement. Thus, we think it likely that all four of these genes are subject to some positive selection. This is clearly the case for Ddc, amd and Cs, and this argument suggests that the Cs transcript may also represent a vital gene locus.

**Origin of the gene arrangement:** A simple duplication event does not readily account for the arrangement and orientation of the amd, Ddc and Cs genes (Figure 1). Several models can be envisioned for the generation of this arrangement. One could assume that the "ancestral state" was two closely linked genes (Cs and a Ddc/amd precursor) and that a transposition event created an inverted copy of the Ddc/amd ancestor. This model does not readily account for the overlap of the Ddc and Cs genes. A single-step scheme (two breaks) giving rise to the Ddc/amd duplication could be imagined, but this might be expected to result in sequence homology (symmetry around the duplication breakpoint). No internal homology within the Cs gene can be detected that would support this model. Unfortunately, the domains outside of the coding regions of the two genes are quite diverged, raising the possibility that such symmetry might have diverged to undetectability in the Cs gene. A more complex postulate is that Ddc and amd were duplicated in a transposition event (minimum of three breaks) and that the Cs gene then evolved by other means, either a second transposition event or evolution from sequences already in situ. Either of these models predicts that the present form of the Cs gene postdates the Ddc/amd duplication event. Since the region between Ddc and amd is almost completely occupied by the Cs gene, which, in fact, overlaps the Ddc transcript, any preexisting proto-Cs gene would be severely disrupted by the event giving rise to the Ddc/amd gene pair. Thus, we favor the interpretation that the Cs gene has arisen either by insertion or by more gradual mechanisms after the Ddc/amd duplication event. It has been noted (SNYDER et al. 1982) that inverted orientation of closely spaced homologous genes will suppress gene duplication by unequal crossing over and will prevent gene correction mechanisms; thus leaving the gene sequences free to diverge.

An alternative explanation for these data is that the structural similarity between Ddc and amd is the result of convergent evolution. We feel that this
is unlikely because the degree of similarity between \( Ddc \) and \( amd \) is greater than that required to generate similar functions from independent precursors. For example, the ADH protein of humans shares only 25% amino acid homology with yeast ADH (Duester, Jornvall and Hatfield 1986). Since \( Ddc \) and \( amd \) do not share the same enzymatic function, somewhat less homology may be required to satisfy a convergence hypothesis.

Although many examples of structural gene duplication exist, precedent for duplication of catalytic genes within a pathway is less common. One example is the \( \beta \)-ketoadipate pathway of bacteria, in which several genes that share a common evolutionary history now catalyze a variety of reactions in this catabolic process (Yeh and Ornston 1980). A second example is found among the genes of catecholamine metabolism in vertebrates, in which recent studies suggest that tyrosine hydroxylase (TH), dopamine-beta-hydroxylase (DBH) and phenylethanolamine-\( N \)-methyltransferase (PNMT) are antigenically related. In addition, cDNA clones for DBH and PNMT appear to cross-hybridize, and DNA blotting experiments suggest that DBH and PNMT may be tightly linked within approximately 4 kb. Joh et al. (1983) conclude that TH, DBH and PNMT share a common evolutionary history. They have not yet detected any gene products related to DDC.

Our analysis provides the first demonstration of structural homology between \( Ddc \) and any other gene in any organism. Our data also provide the first evidence for structural relationships among the genes of the 37C cluster in Drosophila and suggest that the genetic observations indicating functional relationships among the genes in the \( Ddc \) region may be a consequence of the fact that at least some of the genes in this region share common evolutionary histories. This conclusion is in disagreement with those of Gilbert, Hirsh and Wright (1984), who concluded that this region contains no extensive DNA sequence homology and that the genes in this region are not evolutionarily related. We would speculate that additional analysis may reveal a region of active gene duplication and divergence leading to a cluster of structurally related genes functionally related to catecholamine metabolism.

This work was supported by a National Institutes of Health grant (GM28972) and a National Science Foundation grant (8316485) to J.L.M. D.D.E. gratefully acknowledges the support of a National Institutes of Health training grant (T32 CA09054). The authors are grateful to R. B. Hodgetts and C. A. Spencer and to T. R. F. Wright and E. S. Pentz for sharing their manuscripts before publication and for permission to quote their unpublished work; they are also grateful to M. P. Erefle and C. A. Leeds for their valuable contributions to this work. J.L.M. is indebted to Howard A. Schneiderman and the Monsanto Company for their generous support in providing the facilities to carry out these studies. We appreciate the valuable discussions and constructive comments of D. S. Haymer, P. D. L. Gibbs, K. Konrad, P. Bryant, J. Manning, C. Greer, R. Davis and others.

LITERATURE CITED


Citron, B., E. Falck-Pederson, M. Salditt-Georgieff and J. E. Darnell, 1984 Transcription
termination occurs within a 1000 base pair region downstream from the poly(A) site of the mouse beta-globin (major) gene. Nucleic Acids Res. 12: 8723–8731.


Communicating editor: V. G. FINNERTY