Title
MOLECULAR LOCALIZATION, DEVELOPMENTAL EXPRESSION AND NUCLEOTIDE SEQUENCE OF THE ALPHA-METHYLDOPA HYPERSENSITIVE GENE OF DROSOPHILA

Journal
Genetics, 114(2)

ISSN
0016-6731

Authors
Marsh, J Lawrence
Erfle, Mary P
Leeds, Carol A

Publication Date
1986-10-01

DOI
10.1093/genetics/114.2.453

MOLECULAR LOCALIZATION, DEVELOPMENTAL EXPRESSION AND NUCLEOTIDE SEQUENCE OF THE *ALPHA-METHYLDOPA HYPERSENSITIVE* GENE OF DROSOPHILA

J. LAWRENCE MARSH, MARY P. ERFLE AND CAROL A. LEEDS

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 of Drosophila contains a cluster of functionally related genes, many of which affect cuticle development and/or catecholamine metabolism. In this report we describe the molecular mapping and sequencing of a full-length cDNA copy of a transcript that maps to the *alpha-methyldopa hypersensitive* region. Developmental RNA blots show stage-specific patterns of transcription and multiple RNA transcripts. A 2-kb transcript is most abundant at about 12 hr of embryogenesis, and lower levels are detected throughout most of embryogenesis. Lower levels of this transcript are also detected in adults. Smaller stage-specific transcripts are detected in late third-instar larvae. This pattern of transcription is consistent with the known lethal phases and phenotypes of the *alpha-methyldopa hypersensitive* gene (*amd*). Based on map position, pattern of transcription, homology with dopa decarboxylase and association with altered DNA in mutants, we conclude that this transcript represents the *amd* gene.

THE *alpha-methyldopa hypersensitive* gene of Drosophila, *l(2)amd* (abbreviated *amd*) (SPARROW and WRIGHT 1974), maps immediately adjacent to the dopa decarboxylase gene (*Ddc*) (0.002 cM) (WRIGHT *et al.* 1982) and participates in catecholamine metabolism in an unknown way (MARSH and WRIGHT 1979, 1986). In insects, catecholamines serve as cross-linking agents for cuticle sclerotization (hardening), as neurotransmitters in the central nervous system, and as precursors for melanin formation (WRIGHT 1977). Mutations of *amd* exhibit a dual phenotype of recessive lethality and dominant conditional lethality. The lethal phase of the recessive lethal phenotype is embryonic hatching, with dying embryos showing abnormal cuticles as evidenced by thin friable cuticles and necrotic anal organs (SPARROW and WRIGHT 1974). The dominant phenotype is manifested by the lower LD_{50} of *amd*/+ heterozygotes reared on food containing dopa analogues, (*e.g.*, alpha-methyldopa) (SPARROW and WRIGHT 1974). Resistance to dietary alpha-methyldopa is determined primarily by the level of amd^+ gene product (MARSH and WRIGHT 1986). The observa-

tion that larvae fed alpha-methyldopa die at the larval molts and at pupariation exhibiting a thin friable pupal cuticle, whereas adult females fed dopa analogues become sterile and lay eggs that cannot complete embryogenesis, supports the conclusion that the *amd* locus plays a vital role in cuticle development (MARSH and WRIGHT 1979). Thus, the *amd* locus is thought to be functionally related to the nearby *Ddc* locus (MARSH and WRIGHT 1979, 1986).

Analysis of deletion breakpoints indicates that the genes in this region are very densely clustered on a molecular scale (*e.g.*, at least six genes within 17.9–23 kb, GILBERT, HIRSH and WRIGHT 1984). Although this physical organization has potential significance for regulation of these genes, the only gene that has been physically mapped in this region is the *Ddc* gene. We wished to determine the physical organization of the transcription units surrounding the *Ddc* gene and to identify the *amd* gene. In this communication, we report the isolation of a new chromosomal deficiency [*Df(2L)NST*] that uniquely localizes the *Bg* locus (WRIGHT *et al.* 1981) and leaves only the *amd* and *l(2)Bd* loci in the interval immediately distal to *Ddc*. By cDNA cloning and RNA blotting, we have identified a transcription unit mapping in the 1- to 4-kb interval distal to *Ddc*, and we report the precise physical mapping and nucleotide sequence of a complete cDNA for this transcript and the underlying genomic DNA. The location and pattern of expression of this gene is consistent with the known location, lethal phase and phenotype of *amd*. The gene maps only 2.5 kb from *Ddc*, shows extensive structural homology with *Ddc* (EVELETH and MARSH 1986) and contains DNA alterations associated with recently recovered *amd* mutations (E. S. PENTZ and T. R. F. WRIGHT, personal communication). Based on these observations, we assign this transcription unit to the *amd* gene.

MATERIALS AND METHODS

Preparation of plasmid DNA, restriction analysis and subcloning were as described by MANIATIS, FRITSCH and SAMBROOK (1982). Embryonic (1–5.5 hr) and pupal (1-day pupae) cDNA libraries were a kind gift of D. HOGNESS and M. GOLDSCHMIDT-CLERMONT (Stanford University) and were prepared in the lambda gt10 vector. Screening of the libraries was as described by HUYNH, YOUNG and DAVIS (1985), except that we grew phage on a *RecA*$^-$ host because of early instability problems with some isolates.

DNA sequence was determined by the dideoxy chain termination method of SANGER, NICKLEN and COULSEN (1977). Growing and manipulating the phage was performed as described by MESSING (1983). Reactions were performed using nucleotides from P-L Biochemicals, following the manufacturers protocol with minor modifications. Reactions were run on 0.20 mm × 40 cm urea gels at 50 watts. After running, the gel was dried by bonding directly to one of the glass plates which had been treated with 3(trimethoxysilyl)propyl methacrylate (Aldrich) and exposed to X-ray film.

RNA was prepared by aqueous phenol extraction and oligo-dT chromatography (GEITZ and HODGETTS 1985). RNA was separated on 1.3% agarose gels containing formaldehyde and blotted to nitrocellulose. Radiolabeled strand-specific probes (approximately 10^9 cpm/μg) were prepared by *in vitro* transcription of the cloned cDNA using T7 RNA polymerase (U.S. Biochemicals) according to the manufacturers directions. Molecular sizes were determined relative to RNA standards synthesized as above. Hybridization was in 50% formamide, 0.8 M NaCl, 0.1 M PIPES, pH8, 0.01% SDS, 5 × Denhardt's and 100 μg/ml salmon sperm DNA at 65° and washing 3× at 70° in 50 mM NaCl, 20 mM sodium phosphate, pH6.5, and 1 mM EDTA. The amount of intact RNA transferred was confirmed and adjusted relative to the signal obtained with actin

probes (FYRBERG *et al.* 1983). Quantitation of signal was accomplished using an LKB Ultroscanner.

The RNA blots shown in Figure 3A contain approximately 5 μg of poly-A$^+$ RNA, except for the 0- to 4-hr lane, which contains a fivefold excess relative to the adjacent lanes in order to visualize the weak band at 2.0 kb. The 20- to 24-hr adult lanes were exposed 1.7 times longer than the earlier lanes, again in order to better visualize the weak signal. The lanes in Figure 3B contain 10 μg of poly-A$^+$ RNA and were judged to have equivalent amounts of intact RNA based on hybridization with actin. The 5′ and 3′ specific *amd* probes used were 5′ probe = from −270 (cDNA start) to the *Pst*I site at 666, less the intron sequences and 3′ probe = 1313 to cDNA end at 1980.

Analysis of DNA sequences was performed on an IBM-PC or Compaq computer employing a series of programs modified from SCHWINDINGER and WARNER (1984) by ALAN GOLDIN (California Institute of Technology).

RESULTS

Isolation of cDNA clones: Genetic mapping predicts that *amd* mutants lie in the 1- to 4-kb interval distal to *Ddc* (MARSH and WRIGHT 1979; see DISCUSSION). Initial screening of embryonic RNA blots, using a probe that covered the 1- to 10.5-kb region distal to *Ddc*, detected transcript(s) of approximately 2 kb from this region. We used this probe to screen two cDNA libraries made from 1- to 5.5-hr embryo and pupariating larval RNA, respectively. Four clones isolated from the embryonic library appear to represent various portions of the same transcription unit based initially on their common restriction patterns. By using these cDNAs to probe blots of cloned genomic DNA, we were able to localize the cDNA transcripts on the genomic map, as well as confirm their overlapping patterns of hybridization. The overlapping cDNA clones all derived from the interval 2.5–4.7 kb distal to the *Ddc* gene. One clone had an insert of approximately 1.8 kb and was chosen for subsequent analysis. The other clones were judged to represent subfragments of this larger clone by restriction pattern and hybridization. We also analyzed one cDNA, recovered from the late third-instar library, that is homologous to the 3′ portion of the larger transcript and that has the same poly-A addition site.

Sequence analysis: To map this transcript more precisely and to determine its structure in detail, we sequenced the largest cDNA and the underlying genomic DNA by the dideoxy chain termination method. We first prepared a detailed restriction map of both the genomic and cDNA clones. Clones for sequencing were constructed using defined start points chosen to lie every 2–300 base pairs (bp) apart. Several isolates of each clone were checked for identity by C-test (MESSING 1983) before sequencing. The sequencing strategy is shown in Figure 1. The complete genomic sequence is given in Table 1, with the regions corresponding to the mRNA and the deduced translation indicated. All six base restriction enzyme sites that were predicted by the sequence or known from restriction mapping were confirmed.

The 3′ end of the gene was identified by the presence of A residues at the end of the cDNA sequence and the presence of the *Eco*RI linker sequence at the end of the A's. We sequenced the terminus of the late third-instar cDNA clone and found that it was identical to the embryonic cDNA clone and had the same poly-A addition site as the early embryonic cDNA clone. The se-

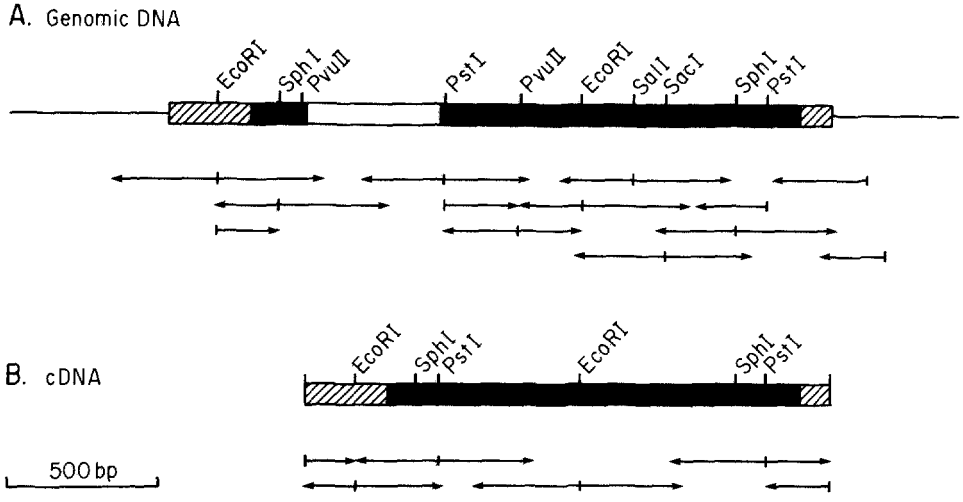

FIGURE 1.—Sequencing strategy. A, Map of the genomic DNA, with the regions of the cDNA transcript indicated by boxes. ■, protein coding regions, ▨, untranslated portions of the mRNA; □, introns; → and ← indicate extent of individual sequencing reactions. B, Map of the cDNA clone. Shading, as in (A), and arrows represent the extent of individual sequencing reactions.

quence of the 1.8-kb cDNA was identical to the genomic sequence over most of its length except for an intron located near the 5′ end of the gene. The sequence of the intron borders conform to the GT··AG rule of 5′ donor 3′ acceptor splice sites (MOUNT 1982) and are in agreement with the direction of transcription deduced from the poly-A tail. The intron is 483 bp in length.

The first 11 bases at the 5′ end of the cDNA did not match the genomic sequence, although they did read as a perfect inverted complement of the adjacent 11 bp of 5′ genomic sequence (Table 2). Within this inverted complement we noted a small inverted repeat (*i.e.*, 5′--CCTC---GAGG--3′). We postulate that this repeat provided the hairpin loop for reverse strand priming during synthesis of the cDNA and that the actual RNA template extended to at least the position indicated in Table 2. The sequence TATAAAA appears 30 bp upstream of this postulated transcription initiation point; this conforms to the spacing and sequence of the consensus eukaryotic promoter sequence (reviewed by BREATHNACH and CHAMBON 1981).

Analysis of the DNA sequence predicts a mature mRNA transcript of 1779 bp after removal of a 483-bp intron (this does not include the poly-A tail). A single, long open-reading-frame of 1395 bp is found that begins at a methionine codon 284 bp from the 5′ end and extends to within 100 bp of the 3′ end, leaving a long 5′ untranslated leader sequence and a smaller 3′ untranslated region. Conceptual translation of the 1395-bp open-reading-frame predicts a protein product of 50,481 kilodaltons (kD) with a slight negative charge (excess of four Asp and Glu residues). The codon usage and distribution of amino acids are given in Table 3.

Genomic map: Genetic analysis places the *B* gene series distal and the *C* gene series proximal to *Ddc* (WRIGHT *et al.* 1981). Recovery of additional

TABLE 1

Sequence of the *amd* gene region

```
-439
TCAAGCTAA ATTAGTTAGA TCAAAGAATA AACAAGTCAG TTGCGCCGTT TTAATGATTC TCAAAACTAG CCAGATTGGC TGAACCGACA GCTCTGGAGG
                                                                          inv."caat"
-340
CTGTCCAGAG AAGTCGGAGT ATAAAAGGCC AGTCACCGGC GATCGGTTTC AGAGTGAACC TCAGGCAACT TGGAGGAGCA TCAACGGATC GGGAACTGAA
              "tata"                                    \mRNA start----------------------------------------
-240
ATCGAGTTGG GCAAACAAAT CAAAAACGAA AACGGGGAAA TAAAACCAAA ACAAACAGAA CGTAAAAAGT GCAAATAGAA AACGATATCG CAACATTGTC
----------------------------------------------------------------------------------------------------------
-140
AGCGGTATGG ATGCCAAGGA GTTTCGGGAA TTCCGAAGGC CGCCATTGAC TACATAGCCG ACTATCTGGA GAATATTCGG GATGACGACG TACTGCCCAA
----------------------------------------------------------------------------------------------------------
-40                                  1                                  30
TCTGGAGCCA GGCTATCTGT TGGACCTGCT GCCCACAGAG ATG CCG GAA GAG CCC GAA GCG TGG AAG GAT GTC CTC GGC GAC ATT AGT CGC
------------------------------------------ Met Pro Glu Glu Pro Glu Ala Trp Lys Asp Val Leu Gly Asp Ile Ser Arg

          60                                 90                                 120
GTC ATC AAG CCG GGA CTG ACC CAC TCG GAG TCG CCT CAC ATG CAT GCC TAC TAC CCC ACC AGC ACC TCG TAT CCC TCC ATT GTG
Val Ile Lys Pro Gly Leu Thr His Ser Glu Ser Pro His Met His Ala Tyr Tyr Pro Thr Ser Thr Ser Tyr Pro Ser Ile Val

                150                                180                                210
GGC GAG ATG CTG GCC AGC GGG TTC GGC GTC ATC GGA TTC AGC TGG         GTATGTTGGT TTATGGTGAA ATCTGCTGCT GCTGCTGCTG
Gly Glu Met Leu Ala Ser Gly Phe Gly Val Ile Gly Phe Ser Trp| splice donor

                                                                                   320
CTGCTTCTGG CCCATTGTTT TGGCCGGCTG AATGGGCGCT CATTGTGCCG GGTGGGTGAA AGTGAATCCA AGAACTCGAC AAACAGGTTG CCACTGCACC

                                                                                   420
GGACCGAAGA GAGTTGTTCA CACAAATCAA TCGGCAATTG TCACCAAATA AAAGCAATAA AATTGGGCAG CAGACTCACC TTAAAGGCAT ACAAATAAAT

                                                                                   520
TTAAATGTAT ATAGAATTGA TTAGCCAAAC CCAAACTAGA CGTTTCGCAA ACAATATTTG TCATTGCGAA CAAAGAAGTT ACCACCGAAC AAAAACTTAG

                                                                                   620
TGAAATAAAC CCTAGTTTAA ATTATAATAT ATTTGTAAAA ATTACTATAT GTATGTATTC CGGATTTAAT AGTGTATTAC AAACGGATGG AGTTATCTTC

                                   660   BglII                          690
AATGCATAAT TTCTTACATA ATAATTCGTA TAATCCCCCA CAG ATC TGC AGT CCC GCC TGC ACA GAA CTG GAG GTG GTG GTC ATG GAC TGG
                              splice acceptor |Ile Cys Ser Pro Ala Cys Thr Glu Leu Glu Val Val Val Met Asp Trp

        720                                750                                780
CTG GCC AAG TTC CTG AAG CTG CCC GCA CAC TTC CAG CAC GCC AGC GAT GGA CCA GGA GGC GGG GTG ATC CAG GGA TCA GCT AGC
Leu Ala Lys Phe Leu Lys Leu Pro Ala His Phe Gln His Ala Ser Asp Gly Pro Gly Gly Gly Val Ile Gln Gly Ser Ala Ser

                810                                840                                870
GAG GCT GTG TTG GTG GCT GTC CTA GCT GCC AGG GAA CAA GCT GTG GCC AAC TAC AGG GAA TCG CAT CCG GAG CTG AGC GAA AGT
Glu Ala Val Leu Val Ala Val Leu Ala Ala Arg Glu Gln Ala Val Ala Asn Tyr Arg Glu Ser His Pro Glu Leu Ser Glu Ser

                900                                930                                960
GAG GTG CGT GGC CGC TTG GTG GCC TAC TCC TCG GAC CAG AGT AAC AGC TGC ATT GAG AAG GCT GGA GTC CTG GCT GCC ATG CCG
Glu Val Arg Gly Arg Leu Val Ala Tyr Ser Ser Asp Gln Ser Asn Ser Cys Ile Glu Lys Ala Gly Val Leu Ala Ala Met Pro

                990                               1020
ATT CGA TTG CTG CCG GCT GGA GAG GAT TTC GTA CTT AGA GGC GAT ACA CTG AGA GGA GCC ATC GAG GAG GAC GTG GCA GCG GGC
Ile Arg Leu Leu Pro Ala Gly Glu Asp Phe Val Leu Arg Gly Asp Thr Leu Arg Gly Ala Ile Glu Glu Asp Val Ala Ala Gly

               1080                               1110
AGG ATT CCG GTG ATC TGC GTT GCC ACT CTG GGC ACC ACG GGC ACT TGT GCC TAT GAC GAT ATT GAA TCC CTG TCC GCT GTC TGC
Arg Ile Pro Val Ile Cys Val Ala Thr Leu Gly Thr Thr Gly Thr Cys Ala Tyr Asp Asp Ile Glu Ser Leu Ser Ala Val Cys

      1140                               1170                               1200
GAG GAA TTC AAG TGT GGC TCC ATG TTG ATG CCG CGT ATG CGG TGG AGC CTT TGC TCT GGA GGA ATG TTC GGA TTT GCG AAA GGG
Glu Glu Phe Lys Cys Gly Ser Met Leu Met Pro Arg Met Arg Trp Ser Leu Cys Ser Gly Gly Met Phe Gly Phe Ala Lys Gly

               1230                               1260                               1290
ATT GGA TCG CGT GGA CTC GCT AAA CTT CAA CCT GCA CAA GTT CAT GCT GGT CAA CTT CGA TTG CTC GGC CAT GTG GCT AAG GGA
Ile Gly Ser Arg Gly Leu Ala Lys Leu Gln Pro Ala Gln Val His Ala Gly Gln Leu Arg Leu Leu Gly His Val Ala Lys Gly

                    1320                               1350                               1380
TGC CAA CAA GGT GGT CGA CAG CTT CAA TGT GGA TCG CAT CTA TCT GAA GCA CAA GCA CGA GGG TCA GTC GCA AAT TCC AGA CTT
Cys Gln Gln Gly Gly Arg Gln Leu Gln Cys Gly Ser His Leu Ser Glu Ala Gln Ala Arg Gly Ser Val Ala Asn Ser Arg Leu

                         1410                               1440
CCG TCA TTG GCA AAT CCC CTG GGT CGC CGC TTC CGA GCT CTA AAA GTC TGG ATC ACA TTC CGC ACT CTG GAA GCC GAG GGA TTG
Pro Ser Leu Ala Asn Pro Leu Gly Arg Arg Phe Arg Ala Leu Lys Val Trp Ile Thr Phe Arg Thr Leu Glu Ala Glu Gly Leu

                               1500                               1530
CGA AAC CAT GTC GCG AAG CAC ATC GAG TTG GCC AAA CAG TTT GAG CAG CTT GTG CTC AAG GAT TCG CGA TTC GAG CTG GTG GCT
Arg Asn His Val Ala Lys His Ile Glu Leu Ala Lys Gln Phe Glu Gln Leu Val Leu Lys Asp Ser Arg Phe Glu Leu Val Ala

      1560                               1590                               1620
CCT CGT GCC CTG GGA CTG GTT TGT TTC CGG CCC AAA GGT GAC AAT GAG ATT ACC ACC CAG TTG CTG CAA CGG CTT ATG GAT CGA
Pro Arg Ala Leu Gly Leu Val Cys Phe Arg Pro Lys Gly Asp Asn Glu Ile Thr Thr Gln Leu Leu Gln Arg Leu Met Asp Arg

   BglII  1650                               1680                               1710
AAG AAG ATC TAC ATG GTT AAG GCC GAG CAT GCG GGT CGT CAC TTT CTG CGA TTC GTC GTA TGC GGC ATG GAC ACC AAA GCC TCC
Lys Lys Ile Tyr Met Val Lys Ala Glu His Ala Gly Arg Gln Phe Leu Arg Phe Val Val Cys Gly Met Asp Thr Lys Ala Ser


                    1740                               1770                               1800
GAT ATT GAT TTC GCC TGG CAG GAG ATC GAG TCT CAA CTG ACG GAC CTG CAG GCG GAC GAA TCC TTG GTG GCC CGC AAA TCG GGA
Asp Ile Asp Phe Ala Trp Gln Glu Ile Glu Ser Gln Leu Thr Asp Leu Gln Ala Asp Glu Ser Leu Val Ala Arg Lys Ser Gly
```

The TATA sequence at −320 is overlined and identified by "tata"; the mRNA start site is indicated by a slash at −284; the ribosome binding sequence of HAGENBUCHLE *et al.* (1978) is underlined at −110; the protein sequence deduced from the cDNA is indicated; the poly-A addition signal is at 1957 and the site of poly-A addition is indicated by \An at position 1977. The inverted "CAAT-like" sequence is underlined and identified by "inv.caat" at −368. Underlining denotes the 5′ and 3′ untranslated sequences found in the cDNA (mature mRNA).

TABLE 1—*Continued*

```
                                  1830                                    1860                    1875
AAC GTC GGC GAT CTT GCG CAC GAC TTC CAG ATC CAT CTG AGC ACC GAA AAT GCA ACG CAC GAG AAA TCT CAG TGA GA
Asn Val Gly Asp Leu Ala His Asp Phe Gln Ile His Leu Ser Thr Glu Asn Ala Thr His Glu Lys Ser Gln *** --

                                                                                                    1980
AAAACGGATA AACTATTTAT GTTTAGGGAC AACCTAGTTA GTTTGCGATG TAGTTITTAA CTTTCCACAT GTTTATAAAT AAAGTGAAAT TAAATGTACG
-------------------------------------------------------------------------------------------------\An

                                                                                                    2080
ATCATTTGGC ACGTTTTCTA TAAAGGTAGA GTGGTTTTTC CCTGTCATTT TTTTTTGTCG AAAAAGTTCC CAACATTCTC TGTTAAACTT TCTGCCGAGG

                                                                                                    2180
CTTTAGTTTT TTAAGCATTA CAAATATCGT CGACTTTTAT TTTAAAATTT AAAACCAAAA TTTTCGCGGC TTAGTGTGAC TGCATTTGGT TATGAATCGA

                                                                                                    2280
TACACTTCTT CATCGCCCTT CGATAAGTTC GCCAAGGTCT ATCGTCATGT GCCGATCCGC AGGGCAAACA GCTGTTTCTC CCAATTGGGA CCACCTGATA

                                                                                                    2380
TCGGTTAAAT AACAAAGTAT AAACAAAACA AAAATATCTG TTTCCCTTAA TTCAATATTT TGATTAGCTT TGAATAGCGT TTAGTGCTAT TTCTCATAAA

                                  2420
TATAGAATAG AAGCAGCCGC GGCTCGCCTT TGTACGAATT C
```

TABLE 2

The 5′ end of the cDNA sequence

A.	5′—AACCTCAGGCAACTTGGAGGAGCAT—3′ 3′—TTGGAGTCCGTTGAACCTCCTCGTA—5′	Genomic sequence
B.	5′—AACCTCAGGCAACTTGGAGGAGCAT—3′	Proposed mRNA sequence
C.	TGAACCTCCTCGTA—5′ S1 → TGCCTGAGGTT—	cDNA copy Second strand synthesis
D.	5′----------**GAATTCC**TGCCTGAGGTTCAT—3′	cDNA sequence

The sequence of the genomic DNA and cDNA clones are given in A and D, respectively. The sequences derived from the linker used to construct the cDNA clones are highlighted in D. The postulated priming loop and S1 cleavage site are indicated in C. Paired bases that might stabilize the loop are underlined. The first 11 bp of the cDNA are a perfect inverted complement of the 11 bp of adjacent genomic sequence. The proposed extent of the mRNA is given in B.

chromosomal deficiencies (BRITTNACHER and GANETZKY 1983; GILBERT, HIRSH and WRIGHT 1984) places three of these genes in the 50-kb region immediately distal to *Ddc*; namely, *amd*, *Bd* and *Bg*. We have recovered a new deficiency, *DF(2L)NST*, which fails to complement alleles of *l(2)Bd*, *amd*, *Ddc* and all of the *C* genes (*i.e.*, *l(2L)37Ca,Cb,Cc,Cd,Ce,Cf,Cg*) (P. D. L. GIBBS and J. L. MARSH, unpublished observations). This deficiency extends to a region approximately 30 kb distal to the *Ddc* gene (PENTZ and WRIGHT 1986) and, in conjunction with others already described (GILBERT, HIRSH and WRIGHT 1984), provides a unique genetic localization of the *Bg* gene and identifies the *Bd* and *amd* genes as the only known complementation groups in the cytogenetic interval immediately distal to *Ddc*.

Our cDNA clones map to the same cytogenetic interval as *amd* and *Bd*. SPENCER, GEITZ and HODGETTS (1986) have recently mapped an additional transcript to this region that we shall designate *Cs*, following the suggestion of WRIGHT *et al.* (1981). The genomic organization of these transcripts is shown in Figure 2. The transcript that we have sequenced lies in the interval 2.5–4.7 kb distal to the 3′ end of the *Ddc* transcript and is transcribed from the opposite strand as *Ddc* but from the same strand as the *Cs* transcript. PENTZ and WRIGHT examined seven radiation and EMS-induced *amd* mutations by DNA blotting (E. S. PENTZ and T. R. F. WRIGHT, personal communication)

TABLE 3

Codon usage and amino acid content of the deduced *amd* protein

Amino acid	Codon usage								Overall amino acid content		Frequency of occurrence (%)
	Codon	No.	Codon	No.	Codon	No.	Codon	No.			
Ala	(GCT)	14	(GCC)	19	(GCA)	8	(GCG)	7	Ala	48	10.34
Arg	(CGT)	5	(CGC)	6	(CGA)	9	(CGG)	3			
	(AGA)	3	(AGG)	3					Arg	29	6.25
Asn	(AAT)	4	(AAC)	4					Asn	8	1.72
Asp	(GAT)	10	(GAC)	10					Asp	20	4.31
Cys	(TGT)	4	(TGC)	8					Cys	12	2.59
Gln	(CAA)	10	(CAG)	12					Gln	22	4.74
Glu	(GAA)	12	(GAG)	21					Glu	33	7.11
Gly	(GGT)	6	(GGC)	13	(GGA)	18	(GGG)	4	Gly	41	8.84
His	(CAT)	8	(CAC)	7					His	15	3.23
Ile	(ATT)	9	(ATC)	11	(ATA)	0			Ile	20	4.31
Leu	(TTA)	0	(TTG)	10	(CTT)	9	(CTC)	4			
	(CTA)	3	(CTG)	22					Leu	48	10.34
Lys	(AAA)	8	(AAG)	12					Lys	20	4.31
Met	(ATG)	12							Met	12	2.59
Phe	(TTT)	3	(TTC)	14					Phe	17	3.66
Pro	(CCT)	3	(CCC)	7	(CCA)	1	(CCG)	8	Pro	19	4.09
Ser	(TCT)	4	(TCC)	8	(TCA)	3	(TCG)	9			
	(AGT)	4	(AGC)	9					Ser	37	7.97
Thr	(ACT)	3	(ACC)	8	(ACA)	3	(ACG)	3	Thr	17	3.66
Trp	(TGG)	6							Trp	6	1.29
Tyr	(TAT)	2	(TAC)	5					Tyr	7	1.51
Val	(GTT)	4	(GTC)	12	(GTA)	2	(GTG)	15	Val	33	7.11
END	(TAA)	0	(TAG)	0	(TGA)	1			END	1	0.22
Molecular weight = 50481											

Codons are given in parentheses and are followed by the number of times each codon is used. On the right, the overall amino acid content is given with the percent occurrence listed. There are 464 amino acids with 49 Arg and Lys and 53 Asp and Glu residues, respectively.

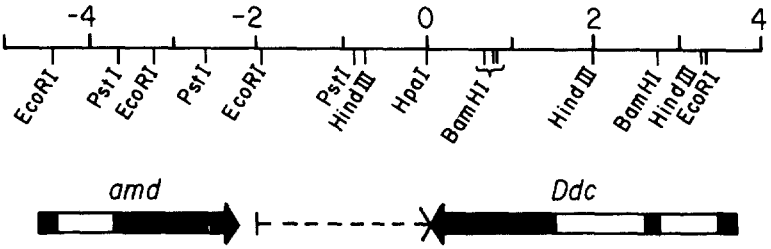

FIGURE 2.—Organization of transcripts in the *Ddc* region. Transcripts discussed in this communication are placed on the genomic map. We have used the *Hpa*I site near the terminus of the *Ddc* as a fixed reference point at 0 kb. This is different from the arbitrary zero point that is used by other authors (*e.g.*, GILBERT, HIRSH and WRIGHT 1984) which would lie at a point slightly greater than 860 bp on our map. The position of the *amd* transcript derives from the sequence data reported here. The dashed arrow represents the 2-kb transcript described by SPENCER, GEITZ and HODGETTS (1986). Solid bars indicate exons included in the mature RNA or cDNA. Arrows indicate direction of transcription from 5′ to 3′. In this figure the centromere is to the right. Restriction enzymes are represented by E(*Eco*RI), B(*Bam*HI), P(*Pst*I), H(*Hin*dIII).

and found two *amd* alleles that showed DNA alterations within the 1-kb *Bgl*II fragment that is internal to the coding region of the *amd* transcript (662 to 1639 in Table 1), thus providing additional and compelling evidence that this transcription unit represents the *amd* gene.

Developmental expression: Total cellular and polysomal poly-A^+ RNA isolated from several stages of development was subjected to electrophoresis on denaturing agarose gels, transferred to nitrocellulose or nylon membranes and probed with labeled single-stranded RNA complementary to the mRNA. Transfer and quantitation of intact RNA was confirmed by probing with actin probes. The *amd* cDNA strand-specific probes revealed a 2.0-kb *amd* transcript, which was first detectable early in embryogenesis and reached a maximum in mid-embryogenesis (*i.e.*, 12–16 hr, Figure 3A). Very weak signal could be detected even in 0- to 4-hr embryonic RNA when a fivefold excess of this RNA was loaded, as in Figure 3A. Whether this represents carry-over of oocyte transcripts or new synthesis cannot be determined from these data. We did observe the 2.0-kb transcript in 3- to 6-hr embryo polysomal RNA (data not shown). Allowing for a poly-A tail, this transcript size of 2.0 kb agrees well with the size of 1780 nucleotides predicted from the cDNA, and it supports our interpretation of the origin of transcription deduced from the sequence analysis. Low levels of the 2.0-kb transcript were also observed in RNA extracted from adults (Figure 3A). *In situ* hybridization to ovaries shows *amd* transcripts concentrated in approximately stage 8–9 oocytes (K. KONRAD and J. L. MARSH, unpublished observations). Transcripts were also detected on blots of third-instar larval RNA, although the bulk of the hybridization is due to small RNA species present at this stage. RNA blots probed with subcloned portions of the *amd* cDNA indicate that the small RNAs hybridize to the 5′ half of the gene and not to the 3′ half (Figure 3B). To determine whether the smaller stage-specific transcripts originate from the *amd* gene itself or from a closely related gene, we probed DNA blots at elevated stringency. Even in 1 M Na^+, 60% formamide 65° (equivalent to approximately 113° 1 × SSC) the small RNAs hybridize, suggesting that they either share a considerable stretch of strong homology with *amd* or they are *bona fide* transcripts from the *amd* gene. To determine whether the structural homology between *Ddc* and *amd* (EVELETH and MARSH 1986) is sufficient to cause *amd* probes to react with *Ddc* transcripts, we performed a reconstruction experiment in which *amd* probes were hybridized to *in vitro* synthesized full-length *amd* and *Ddc* transcripts at various stringencies. The hybridization signal with *Ddc* transcripts was less than 10% of that with *amd* and was not detectable at the RNA concentrations found on the Northern blots. Thus, *Ddc* transcripts are not contributing significantly to the signals observed in these studies.

DISCUSSION

Genomic map: Sixteen lethal genes and the visible gene *hk* have been identified in the *Ddc* region as defined by *Df(2L)TW130* (WRIGHT *et al.* 1981). The chromosomal deficiencies described here in conjunction with previously described deficiencies establish the following chromosomal order in this region:

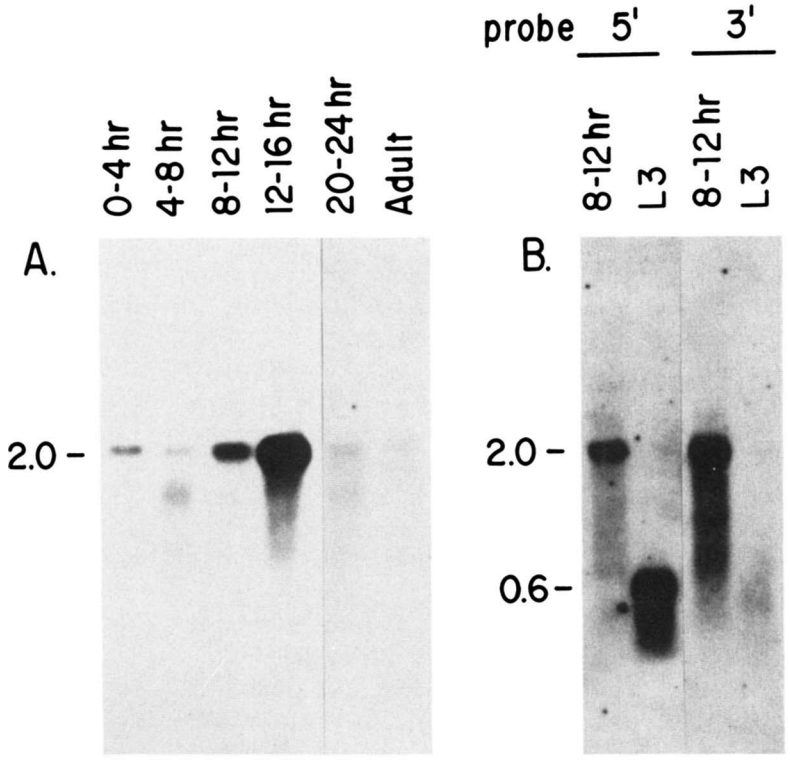

FIGURE 3.—Developmental RNA blots. Poly-A$^+$ RNA was extracted and run on denaturing agarose gels and was probed, as described in MATERIALS and METHODS. Approximate developmental stages are given above each lane. A, Each lane contains an equivalent amount of poly-A$^+$ RNA. The 0- to 4-hr lane contains fivefold excess of RNA to permit visualization of the weak signal. The 20- to 24-hr and adult lanes are exposed 1.7 times longer than are the others. Molecular weights are given in kilobases. B, Each lane contains 10 μg of poly-A$^+$ RNA probed with 5′ and 3′ specific probes, as described in MATERIALS and METHODS. Transfer of intact RNA was confirmed by probing with actin.

[*hk*] [*Ba Bc Be*] [*Bd Bb*] [*Bg*] [*Bd amd*] [*Ddc*] [*Cc*] [*Cb*] [*Ca Cd Ce Cg*] [*Cf*]. The *Bd* and *amd* genes are the only two genetic units identified in the approximately 30-kb cytogenetic interval immediately distal to *Ddc*. The *amd* gene maps 0.002 cM distal to the *Ddc* gene (WRIGHT *et al.* 1981), which would place it approximately 1.2 (GELBART, MCCARRON and CHOVNICK 1976) to 4 kb (BOSSY, HALL and SPIERER 1984) distal to the *Ddc* gene. The 3′ end of the cDNA transcript we have sequenced lies approximately 2.5 kb from the 3′ terminus of the *Ddc* gene (SPENCER, GEITZ and HODGETTS 1986; D. D. EVELETH, J. L. MARSH and R. B. HODGETTS, unpublished observations) on the opposite strand. SPENCER, GEITZ and HODGETTS (1986) describe another 2.0-kb RNA (herein designated as *Cs*) that maps between *Ddc* and *amd* but its pattern of transcription is inconsistent with the *amd* phenotype.

Dense clustering: Placing a 2-kb transcript between *Ddc* and the *amd* cDNA would place the 5′ start of the *Cs* transcript only 450–650 bp from the 3′ terminus and the *amd* gene, without including any introns but allowing for a

poly-A tail. As seen in Figure 2, three transcripts are contained within about 8.5 kb, and the unspliced transcripts from these genes account for almost 8 kb of this region. Thus, the genes in this interval are very densely packed. This dense clustering of genes has complicated the analysis of transcripts in this region, because probes used in other studies have later been found to overlap more than one transcription unit (*e.g.*, SPENCER, GEITZ and HODGETTS 1986). The use of strand-specific probes derived from cDNA clones and the detailed sequence analysis and mapping described here ensure that we are analyzing only a single transcription unit. If transcription termination in Drosophila (PRICE and PARKER 1984) occurs as much as 1 kb downstream of the poly-A addition processing site, as has been observed in a number of cases (CITRON *et al.* 1984; HAGENBUCHLE *et al.* 1984; BIRNSTEIL, BUSSLINGER and STRUB 1985), the gene arrangement seen here implies that the primary transcripts from these genes overlap one another. The regulatory consequences of this fact (if any) await functional studies.

Developmental expression: The major 2.0-kb transcript is weakly detectable during early embryogenesis, reaches a maximum in mid-embryogenesis at approximately 12–16 hr and is relatively weak again at embryonic hatching. Developmentally, mid-embryogenesis is a period of active cuticle deposition (POODRY 1980), which correlates well with the observation that embryos homozygous for *amd* mutations show cuticular defects and die at the embryonic/larval boundary, exhibiting a necrotic anal organ and thin friable cuticles (SPARROW and WRIGHT 1974). In the absence of conditional mutations, there is no direct evidence that *amd* function is required in adults. However, the effects of dietary alpha-methyldopa on adults are specific to females who, after 2–3 days exposure, lay eggs which do not complete embryogenesis (MARSH and WRIGHT 1979). The appearance of low levels of the 2.0-kb mRNA in adults, and the appearance of *amd* transcripts in nurse cells of stage 8–9 oocytes, suggests a possible maternal requirement for *amd* function. The pattern of transcription in third instar larvae is more complex. Smaller, stage-specific transcripts are observed that suggest developmentally specific modes of transcription or processing. Whether the smaller transcripts originate from a closely related gene or from alternate promoters or alternative RNA processing events of the *amd* gene is presently under study.

Correlation with *amd*: The transcript described by SPENCER, GEITZ and HODGETTS (1986) (*Cs*) is most abundant in adult males and early embryos and decreases in amount during the first 8 hr of embryogenesis. This pattern of expression is inconsistent with the known phenotype and lethal phases of *amd*. The location of our transcript to the region 1.2–4 kb distal to *Ddc* and the appearance of the 2.0-kb transcript, which correlates well with the embryonic lethal phase of *amd*, are consistent with the genetic observations of the *amd* gene. The observation that two of seven EMS- and gamma-ray-induced *amd* mutations show altered restriction patterns that map to the 1-kb *Bgl*II fragment contained within the cDNA sequence (E. S. PENTZ and T. R. F. WRIGHT, personal communication) further extends this correlation. We interpret these observations to indicate that this transcript defines the *amd* gene.

Structure of the gene: The eukaryotic promoter sequence TATAAAA occurs 30 bp upstream of the proposed 5′ end of the RNA transcript. This conforms nicely to the consensus sequence (5′-TATA[A/T]A[A/T]-3′) and position (3′ terminal A or T at − 24 ± 4 bp) derived from 60 examples of TATA boxes (BREATHNACH and CHAMBON 1981). We have placed the origin of transcription at the last base sequenced from the cDNA (Table 2), although it is quite likely that the mature RNA would extend another few bases, thus placing the TATA box closer to the canonical interval for this structure. It is interesting to note that the TATA homology used in the *amd* gene is identical to the corresponding sequences in three of four cuticle protein genes analyzed by SNYDER *et al.* (1982). The consensus sequence (5′-GG[C/T]CAATCT-3′) is often found approximately 50 bp upstream of the TATA box homology (BREATHNACH and CHAMBON 1981). The sequence AGATTGGC is found 45 bp upstream of the TATA homology in our sequence and reads as an inverted "CAAT box," which matches the consensus in eight of nine positions and, as such, may serve as a "CAAT box." No strong homology was noted to the GGGCGG sequence implicated in some promoter regions (DYNAN and TIJAN 1985).

We also note the occurrence of sequences of noteworthy structure because of their symmetry or repetition. One copy of the sequence AATAAACAA is found at position −414 that is similar to the decamer AATAAACAAA found repeated three times in regions flanking the hormonally regulated SGS4 gene sequence (MUSKAVITCH and HOGNESS 1982). In addition, the sequence GGCTGAAccGAC beginning at position −363 is repeated in inverse at position −335 as CAGagAAGTCGG (lower case letters indicate mismatch), and the sequence GGCTG at −363 is repreated at −342, while the sequence TTGGCTG (−365) is repeated as the inverse complement with AACCGAC (−358). The significance of these sequences is unclear, but they are sufficiently uncommon to stand out in scans for regions of symmetry or repetition and are reminiscent of similar sequence motifs found in the regions flanking other genes, especially ecdysone-induced genes (MUSKAVITCH and HOGNESS 1982).

The sequences immediately surrounding the AUG initiation codon site do not match the generalized eukaryotic initiation consensus proposed by KOZAK (1984); namely, preference for C's and an A at −3 bp (ccAccAUG). The use of this initiation codon indicates that the *amd* gene has an unusually long 5′ untranslated leader (284 bp). In a compilation of 211 leader sequences from eukaryotic genes, KOZAK (1984) finds over 70% of the leaders clustered in the 20- to 80-nucleotide range and speculates that the structure of the longer leaders may participate in unusual modes of regulation. This long leader contains four premature AUG codons. HAGENBUCHLE *et al.* (1978) propose that ribosomal binding involves base-pairing of sequences in the untranslated leader sequences with sequences near the 3′ end of the 18 S rRNA. The sequence at −110, 5′-UUCCGAAG-3′ can base pair at seven of its eight positions with the conserved 3′-AAGGCGUC-5′ sequence located 7 bp from the 3′ end of the 18 S rRNA and, thus, might serve as a ribosome binding site. This proposition would still leave two "nonfunctional" AUG codons before the functional

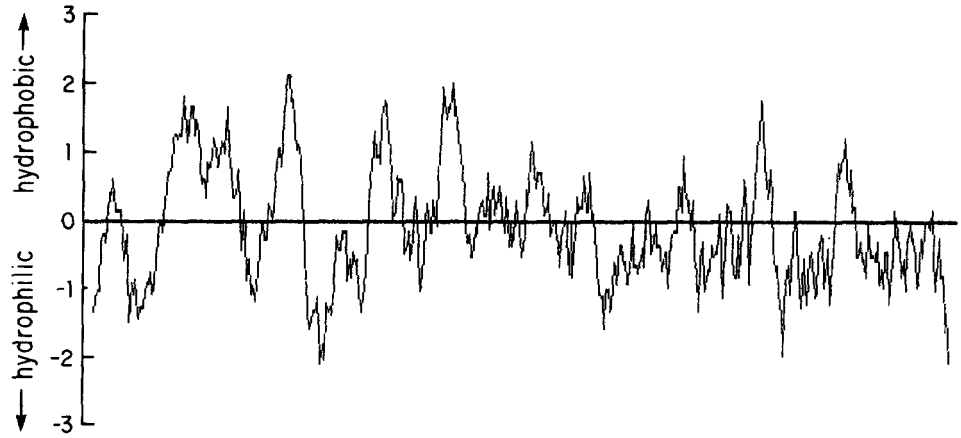

FIGURE 4.—Hydropathy plot of *amd* protein. The regional hydropathy values for the *amd* protein were calculated by the method of KYTE and DOOLITTLE (1982) using a window size of 11 amino acids. The plot begins with the amino terminus on the left.

AUG. We do not know the effect of these codons on translation, but it is possible that the long leader containing "nonfunctional" AUG codons may serve to regulate translation using false starts, as has been observed in some genes (HUNT 1985).

Examination of the predicted amino acid sequence does not reveal a leader peptide resembling a signal sequence (KREIL 1981). The predicted *amd* protein does show one broad lipophilic (hydrophobic) domain (Figure 4) that by attaining average hydropathy value of >+1.6 over about 19 residues, has a high probability of participating in membrane binding (KYTE and DOOLITTLE 1982). Four other lipophilic spikes with hydropathy values of >+1.6 occur, but they do not meet the length requirements described by KYTE and DOOLITTLE (1982) for membrane binding. A potential glycosylation site Asn-X-Thr (MCREYNOLDS *et al.* 1978) is located eight amino acids from the carboxy terminus of the protein. It is possible that these sequences play a role in the intracellular sorting of the *amd* protein. Immunocytochemical localization of the *amd* protein will provide insight into this possibility.

Sequence comparison of the cDNA clone and the genomic sequence indicated the presence of a 483-bp intervening sequence conforming to the GT··AG rule of 5′ donor 3′ acceptor splice sites. We compared the splice sequences of this gene with the tabulated frequencies of splice junction sequences from approximately 130 genes in a variety of species (MOUNT 1982). The overall agreement with these tabulated frequencies is remarkably good, with the Drosophila sequence showing a preference for C's near the splice acceptor where vertebrates show a slight preference for runs of T's. We also note the occurrence of the sequence 5′-ATAAT-3′ approximately 10, 20 and 30 bp 5′ of the splice acceptor site. This sequence matches in four of five positions with the proposed consensus 3′ splice signal thought to be necessary for lariat formation during splicing in Drosophila (KELLER and NOON 1985).

The sequence of the poly-A addition site conforms well to the consensus

sequence characteristic of vertebrate poly-A addition sites (BERGET 1984; BIRNSTEIL, BUSSLINGER and STRUB 1985). For example, the consensus hexanucleotide sequence AATAAA is found 14 bp upstream of the poly-A addition site that occurs at the T of a TA pair. In addition, the genomic sequences downstream of this region are GT rich, with long runs of blocks of contiguous T residues similar to the sequences seen in a number of vertebrate genes (BIRNSTEIL, BUSSLINGER and STRUB 1985).

The functional nature of the reading frame predicted by the sequence analysis was confirmed by fusing a major portion of the predicted frame to the *lacZ* region of pUC19 (MESSING 1983) and showing the production of a new polypeptide. Antiserum raised against this polypeptide cross reacts with a 52-kD protein in extracts of Drosophila embryos (K. D. KONRAD and J. L. MARSH, unpublished results), thus corroborating the conclusion based on the polysomal location of the mRNA that the functional reading frame described here is actually expressed as an mRNA and is translated to protein.

Although the metabolic function of the *amd* gene product is unknown, we know that it does not encode phenol oxidase or dopamine acetyl transferase, the prior and next enzymes in the insect sclerotization pathway (SPARROW and WRIGHT 1974; PENTZ, BLACK and WRIGHT 1986). Nevertheless the rescue of *amd* mutants with catecholamine metabolites (P. D. L. GIBBS and J. L. MARSH, unpublished observations), the ability of *amd* to confer resistance to dopa analogues, the effect of *amd* on cuticle development and its structural similarity to *Ddc* (EVELETH and MARSH 1986) all suggest a role in catecholamine metabolism. The catalytic activity is currently under investigation.

This work was supported by grants from the National Institutes of Health (GM28972) and the National Science Foundation (PCM 8316485) to J.L.M. We wish to thank R. B. HODGETTS, C. A. SPENCER, E. S. PENTZ, B. BLACK and T. R. F. WRIGHT for communicating their results and for permission to quote their work prior to publication. We are indebted to P. D. L. GIBBS for isolating and characterizing the deficiencies described here, to K. D. KONRAD for permission to quote his unpublished observations, to R. GANGULE for his kind gift of polysomal RNA and to D. HAYMER for his valued contributions to this work. We gratefully acknowledge R. ROBERTO for generating the M13 cDNA clones as part of his undergraduate research project. J.L.M. is indebted to M. P. ERFLE and C. A. LEEDS for their expert technical assistance and valuable and continuing contributions to this study, and to HOWARD SCHNEIDERMAN and the Monsanto Corporation for providing the facilities to conduct these studies.

LITERATURE CITED

BERGET, S. M., 1984 Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature **309:** 179–182.

BIRNSTIEL, M. L., M. BUSSLINGER and K. STRUB, 1985 Transcription termination and 3′ processing: the end is in site! Cell **41:** 349–359.

BOOSY, B., L. M. C. HALL and P. SPIERER, 1984 Genetic activity along 315 kb of the Drosophila chromosome. EMBO J **3:** 2537–2541.

BREATHNACH, R. and P. CHAMBON, 1981 Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50:** 349–383.

BRITTNACHER, J. C. and B. GANETZKY, 1983 On the components of segregation distortion in *Drosophila melanogaster*. II. deletion mapping and dosage analysis of the *SD* locus. Genetics **103:** 659–673.

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.

DYNAN, W. S. and R. TIJAN, 1985 Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature **316:** 774–778.

EVELETH, D. D. and J. L. MARSH, 1986 Evidence for evolutionary duplication of genes in the dopa decarboxylase region of Drosophila. Genetics **114:** 469–483.

FYRBERG, E. A., J. W. MAHAFFEY, B. J. BOND and N. DAVIDSON, 1983 Transcripts of the six Drosophila actin genes accumulate in a stage- and tissue-specific manner. Cell **33:** 115–123.

GEITZ, R. D. and R. B. HODGETTS, 1985 An analysis of dopa decarboxylase expression during embryogenesis in *Drosophila melanogaster*. Dev. Biol. **107:** 142–155.

GELBART, W., M. MCCARRON and A. CHOVNICK, 1976 Extension of the limits of the XDH structural element in *Drosophila melanogaster*. Genetics **84:** 211–232.

GILBERT, D., J. HIRSH and T. R. F. WRIGHT, 1984 Molecular mapping of a gene cluster flanking the Drosophila dopa decarboxylase gene. Genetics **106:** 679–694.

HAGENBUCHLE, O., M. SANTER, J. A. STEITZ and R. J. MANS, 1978 Conservation of the primary structure at the 3′ end of 18S rRNA from eukaryotic cells. Cell **13:** 551–563.

HAGENBUCHLE, O., P. K. WELLAUER, D. L. CRIBBS and U. SCHIBLER, 1984 Termination of transcription in the mouse α-amylase gene *Amy-2a* occurs at multiple sites downstream of the polyadenylation site. Cell **38:** 737–744.

HUNT, T., 1985 False starts in translational control of gene expression. Nature **316:** 580–581.

HUYNH, T. V., R. A. YOUNG and R. W. DAVIS, 1985 Constructing and screening cDNA libraries in lgt10 and lgt11. In: *DNA Cloning Techniques: A Practical Approach*, Edited by D. GLOVER. IRL Press, Oxford.

KELLER, E. B. and W. A. NOON, 1985 Intron splicing: a conserved internal signal in introns of Drosophila pre-mRNAs. Nucleic Acids Res. **13:** 4971–4990

KOZAK, M., 1984 Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12:** 857–872.

KREIL, G., 1981 Transfer of proteins across membranes. Annu. Rev. Biochem. **50:** 317–348.

KYTE, J. and R. F. DOOLITTLE, 1982 A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 105–132.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

MARSH, J. L. and T. R. F. WRIGHT, 1979 Control of dopa decarboxylase expression during development in Drosophila. pp. 183–194. In: *Eukaryotic Gene Regulation: ICN-UCLA Symposia on Molecular and Cell Biology*, Edited by R. AXEL, T. MANIATIS and C. F. FOX. Academic Press, New York.

MARSH, J. L. and T. R. F. WRIGHT, 1986 Evidence for regulatory variants of the dopa decarboxylase and alpha-methyldopa hypersensitive loci in Drosophila. Genetics **112:** 249–265.

MCREYNOLDS, L., B. W. O'MALLEY, A. D. NISBET, J. E. FOTHERGILL, D. GIVOL, S. FIELDS, M. ROBERTSON and G. G. BROWNLEE, 1978 Sequence of chicken ovalbumin mRNA. Nature **273:** 723–728.

MESSING, J., 1983 New M13 vectors for cloning. pp. 20–78. In: *Methods in Enzymology*, Vol. 101C, Edited by R. WU, L. GROSSMAN and K. MOLDAVE. Academic Press, New York.

MOUNT, S. M., 1982 A catalogue of splice junction sequences. Nucleic Acids Res. **10:** 459–472.

MUSKAVITCH, M. A. T. and D. S. HOGNESS, 1982 An expandable gene that encodes a Drosophila

glue protein is not expressed in variants lacking remote upstream sequences. Cell **29:** 1041–1051.

PENTZ, E. S., B. C. BLACK and T. R. F. WRIGHT, 1986 A diphenol oxidase gene is part of a cluster of genes involved in catecholamine metabolism and sclerotization in Drosophila. I. Identification of the biochemical defect in *Dox-A2* [*l(2)37Bf*] mutants. Genetics **112:** 823–841.

PENTZ, E. S. and T. R. F. WRIGHT, 1986 A diphenol oxidase gene is part of a cluster of genes involved in catecholamine metabolism and sclerotization in Drosophila. II. Molecular localization of the *Dox-A2* coding region. Genetics **112:** 843–859.

POODRY, C. A., 1980 Epidermis: morphology and development. pp. 443–448. In: *The Genetics and Biology of Drosophila*, Edited by M. ASHBURNER and T. R. F. WRIGHT. Academic Press, New York.

PRICE, D. H. and C. S. PARKER, 1984 The 3′ end of Drosophila histone H3 mRNA is produced by a processing activity in vitro. Cell **38:** 423–429.

SANGER, F., S. NICKLEN and A. R. COULSEN, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHWINDINGER, W. F. and J. R. WARNER, 1984 DNA sequence analysis on the IBM-PC. Nucleic Acids Res. **12:** 601–604.

SNYDER, M., M. HUNKAPILLER, D. YUEN, D. SILVERT, J. FRISTROM and N. DAVIDSON, 1982 Cuticle protein genes of Drosophila: structure, organization and evolution of four clustered genes. Cell **29:** 1027–1040.

SPARROW, J. C. and T. R. F. WRIGHT, 1974 The selection of mutants in *Drosophila melanogaster* hypersensitive to α-methyldopa, a dopa decarboxylase inhibitor. Mol. Gen. Genet. **130:** 127–141.

SPENCER, C. A., R. D. GEITZ and R. B. HODGETTS, 1986 Developmental expression of the transcription unit adjacent to the 3′ end of the dopa decarboxylase gene in *Drosophila melanogaster*. Dev. Biol. **114:** 260–264.

WRIGHT, T. R. F., 1977 The genetics of dopa decarboxylase and α-methyldopa sensitivity in *Drosophila melanogaster*. Am. Zool. **17:** 707–721.

WRIGHT, T. R. F., W. BEERMANN, J. L. MARSH, C. B. BISHOP, R. STEWARD, B. C. BLACK, A. D. TOMSETT and E. Y. WRIGHT, 1981 The genetics of dopa decarboxylase in *Drosophila melanogaster*. IV. The genetics and cytology of the 37B10-37D1 region. Chromosoma **83:** 45–58.

WRIGHT, T. R. F., B. C. BLACK, C. P. BISHOP, J. L. MARSH, E. S. PENTZ, R. STEWARD and E. Y. WRIGHT, 1982 The genetics of dopa decarboxylase in *Drosophila melanogaster*. V. *Ddc* and *l(2)amd* alleles: isolation, characterization and intragenic complementation. Mol. Gen. Genet. **188:** 18–26.

Communicating editor: V. G. FINNERTY