UC Irvine

UC Irvine Previously Published Works

Title

Sequence and structure of the dopa decarboxylase gene of Drosophila: evidence for novel RNA splicing variants.

Permalink

https://escholarship.org/uc/item/6588q2s2

Journal

The EMBO Journal, 5(10)

ISSN

0261-4189

Authors

Eveleth, DD
Gietz, RD
Spencer, CA
et al.

Publication Date

1986-10-01

DOI

10.1002/j.1460-2075.1986.tb04549.x

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

The EMBO Journal vol.5 no.10 pp.2663–2672, 1986

Sequence and structure of the dopa decarboxylase gene of *Drosophila*: evidence for novel RNA splicing variants

David D.Eveleth, R.Daniel Gietz[1], Charlotte A.Spencer[1], Frank E.Nargang[1], Ross B.Hodgetts[1] and J.Lawrence Marsh

Department of Developmental and Cell Biology and The Developmental Biology Center, University of California at Irvine, Irvine, CA 92717, USA, and [1]Department of Genetics, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Communicated by R.Miledi

In *Drosophila*, dopa decarboxylase (DDC) serves a dual role in neurotransmitter production and sclerotization of the cuticle. The *Ddc* gene is under complex hormonal and tissue-specific control and several sizes of *Ddc* RNA are observed at embryonic hatching, pupariation and adult eclosion. We present here the complete nucleotide sequence of the *Drosophila* dopa decarboxylase gene and the partial sequence of two corresponding *Ddc* cDNAs. The sequence allows us to account for the detailed structure of four of the five major *Ddc* RNA species observed. The cDNA sequence reveals the existence of previously undetected splicing events and provides evidence for two RNA splicing alternatives which appear to encode two protein isoforms. The structure, processing and developmental regulation of the *Ddc* transcripts and putative protein isoforms are discussed. Interestingly, the pyridoxal-binding peptide of porcine DDC matches the *Drosophila* sequence perfectly suggesting considerable selective pressure on at least portions of the sequence. This is the first available *Ddc* gene sequence from any organism and should serve as a basis of comparison for the related proteins of other species.

Key words: alternative mRNA processing/dopa decarboxylase/ecdysone/hormone response/*Drosophila*/catecholamine metabolism

Introduction

In *Drosophila*, dopa decarboxylase (DDC: EC 4.1.1.26), which catalyzes the conversion of dopa to dopamine, is encoded by a single gene (Wright *et al.*, 1982). Most of the DDC activity (>90%) is found in the epidermis where dopamine derivatives promote cuticle sclerotization (Lunan and Mitchell, 1969) but DDC is also found in the central nervous system where it produces neurotransmitter (Livingstone and Tempel, 1984), in the proventriculus where it participates in synthesizing the peritrophic membrane of the gut (Wright, 1977) and in the ovary where it is required for female fertility (Marsh and Wright, 1979; Wright *et al.*, 1981). During development, DDC enzyme activity in the epidermis peaks five times corresponding with each of the periods of molting (Marsh and Wright, 1980; Kraminsky *et al.*, 1980). At pupariation *Ddc* is induced by rising titers of the steriod hormone 20-hydroxyecdysone (Kraminsky *et al.*, 1980) but falling levels of hormone are required for *Ddc* expression during embryogenesis and imaginal disc evagination (Clark *et al.*, 1986) thus suggesting a dual mode of response to hormone (Hodgetts *et al.*, 1986). Conditional mutations of the *Ddc* locus result in death at periods of major cuticle maturation such as hatching, pupariation and eclosion (Wright *et al.*, 1981). Mutant adults carrying these alleles show learning disabilities under restrictive conditions (Tempel *et al.*, 1984) attesting to the functional requirement for DDC activity in the central nervous system.

RNA blots probed for *Ddc* transcripts reveal RNA species of 4.0, 3.0, 2.7, 2.3 and 2.0 kb (Beall and Hirsh, 1984; Gietz and Hodgetts, 1985). Based on a proposed gene structure determined by R-looping and blotting experiments, the 2.3-kb, the 2.7-kb and the 3.0-kb RNA species have all been interpreted as processing intermediates leading from a single 4.0-kb primary transcript to a single mature mRNA of 2.0 kb (Beall and Hirsh, 1984; Gietz and Hodgetts, 1985). Gene transfer experiments show that all the sequences necessary for proper developmental expression of *Ddc* function are contained within a 7.2-kb fragment of genomic DNA (Scholnick *et al.*, 1983; Marsh *et al.*, 1985). To understand the structure of the *Ddc* RNA transcripts, the ontogeny of their expression and to identify sequences which might function in the regulation of *Ddc* expression, we have sequenced most of the 7.2-kb fragment and portions of two *Ddc* cDNAs which allow us to deduce the protein coding portion of the gene. The sequence and hybridization properties of the longest cDNA indicate that it corresponds to the 2.3-kb RNA. We conclude that the 2.3-kb RNA is not a precursor of the 2.0-kb species as has been proposed (Beall and Hirsh, 1984) but rather is an alternative processing variant of *Ddc* mRNA. These results suggest a detailed model for the structure of the various *Ddc* RNAs and surprisingly offer evidence that translation of the 2.3-kb and 2.0-kb species may lead to the production of two distinct DDC protein isoforms.

Results

Isolation of cDNA clones

A cDNA library was prepared from 0–2 h old adult Canton-S poly(A)$^+$ RNA by reverse transcription, dG tailing and oligo(dC) priming of second strand synthesis (see Materials and methods). Before ligation of linkers half of the sample was treated with S1 nuclease to polish ends. From 10 μg of adult poly(A)$^+$ RNA, 4.6×10^6 p.f.u. were recovered from the S1-treated library and 1.7×10^6 p.f.u. from the non-S1-treated library. For screening, 250 000 p.f.u. were transfered to nitrocellulose (Benton and Davis, 1977) and probed with a nick-translated 3.6-kb *Hpa*I fragment which covers the entire *Ddc* gene. Of 24 positive plaques detected, five were shown to bear recombinant inserts three of which appeared to represent sizeable *Ddc* cDNA inserts. These observations suggest an RNA abundance of 1.2×10^{-3}% which is considerably lower than the 0.1–0.01% estimated by Hirsh and Davidson (1981). The source of this under-representation in cDNA libraries is unknown but has been observed in several libraries from various sources (J.L.Marsh *et al.*, unpublished observations). No striking secondary structure which might preclude reverse transcription is seen in the sequence.

Two cDNAs were analyzed from these screens. The longest

© IRL Press Limited, Oxford, England

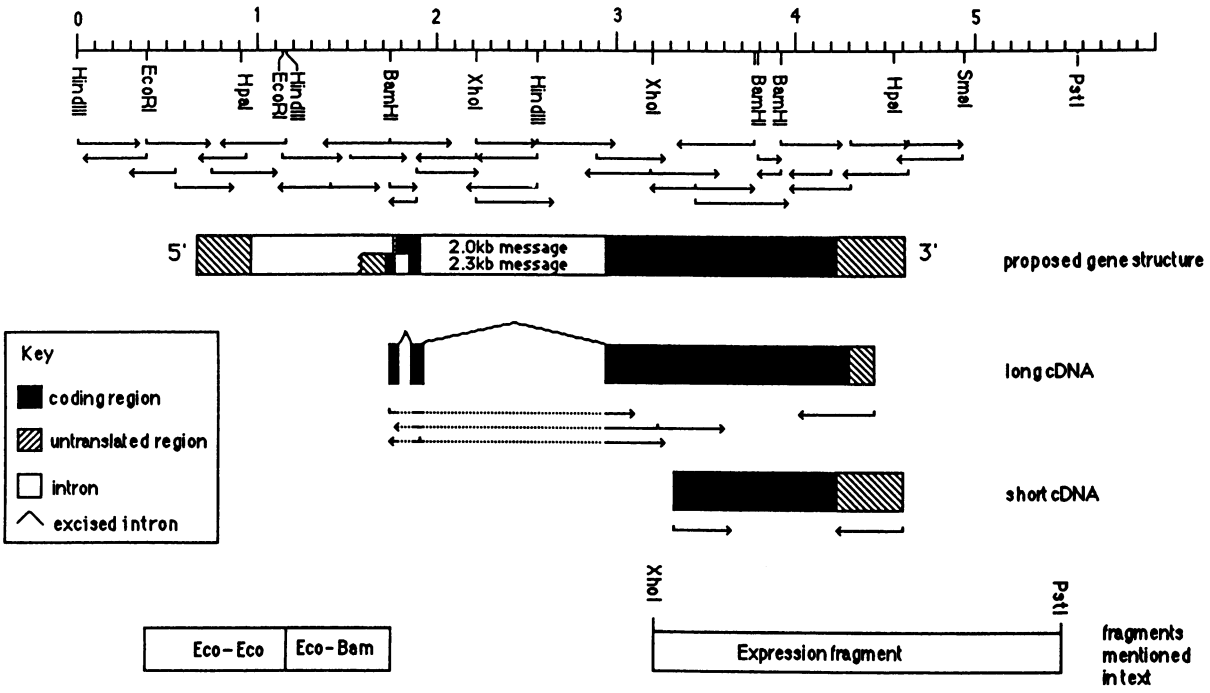

Fig. 1. Sequencing strategy and clone maps. The scale at the top is in kilobases and the numbering corresponds to the numbering used in the sequence (Figure 2) and all subsequent figures. Key restriction sites are shown. Arrows below indicate the extent of genomic sequence obtained from M13 clones. The 'proposed gene structure' is shown below to provide visual reference to the sequenced regions and the cDNAs. Solid boxes correspond to protein coding regions, hatched areas to untranslated regions of the mRNA and open boxes to introns. The split diagram in the central portion indicates that two alternative splicing pathways are used, the upper giving rise to a 2.0-kb mRNA and the lower to a 2.3-kb mRNA. The jagged 5′ edge of the second exon of the 2.3-kb message signifies our uncertainty as to the precise location of the splice acceptor. The extent of the long and short cDNAs is shown relative to the genomic map. Excised introns are shown by thin diagonal lines. Sequenced portions of the cDNAs are shown by thin solid lines (dotted portions indicate intron breaks in the cDNA). At the bottom, fragments referred to in the text are shown in relation to the genomic sequence and particular elements of the transcription unit. All figures are vertically in line.

is 1532 bases in length, spans two introns, contains a single long open reading frame (ORF) and defines the protein coding portion of the gene. The shorter (1200 bp) is identical to part of the longest clone but contains a poly(A) tail thus defining the site of poly(A) addition. Both cDNAs hybridized to probes from the major exon of the *Ddc* gene (i.e. *Xho*I/*Bam*HI and *Bam*HI/*Hpa*I; positions 3202−3662 and 3813−4526, respectively) while the longest hybridized to the *Eco*RI/*Bam*HI and *Bam*HI/*Xho*I fragments (1139−1732 and 1732−2197 on the genomic sequence, respectively, see below) (data not shown). Since the *Eco*RI/*Bam*HI fragment (Figure 1) hybridizes to the 2.3- but not the 2.0-kb *Ddc* RNA (Beall and Hirsh, 1984; Gietz and Hodgetts, 1985; Swiderski and O'Connor, personal communication; unpublished observations), we conclude that this cDNA represents the 2.3-kb RNA. Since all of the RNA species hybridize to the *Eco*RI/*Eco*RI fragment located between 387 and 1139 and shown in Figure 1 (Beall and Hirsh, 1984; Gietz and Hodgetts, 1985) while neither of the cDNAs hybridize to this fragment, we conclude that the small 5′ exon is not included in the cDNAs. We show below that this exon is not translated.

Sequence and structure of genomic and cDNA clones

We have determined the complete nucleotide sequence of a 4835-bp region of genomic DNA corresponding to the entire *Ddc* gene and flanking sequences. The genomic clones originate from the Maniatis library (Maniatis *et al.*, 1978) and the restriction pattern of these recombinant phage resembles the Canton-S haplotype of *Drosophila* (Marsh and Wright, 1986). A 7.2-kb *Pst*I fragment from these phage has been demonstrated by gene transfer to rescue lethal alleles of *Ddc* and to exhibit proper developmental regulation (Scholnick *et al.*, 1983; Marsh *et al.*, 1985). Alignment of the *Ddc* gene on the genomic sequence was facilitated by the recovery and partial sequencing of the two *Ddc* cDNA clones described above. The M13 clones used to generate the genomic and cDNA sequences are shown in Figure 1 while the genomic sequence itself is given in Figure 2. The reliability of the DNA sequence is demonstrated by the fact that the only ORF predicted by the sequence has been confirmed experimentally (see below) and all six-base restriction sites and some four-base restriction sites have been confirmed (data not shown).

Sequence analysis of the larger cDNA reveals a single ORF corresponding to the only large ORF seen in the genomic sequence. This ORF extends from the beginning of the cDNA through two splice junctions to position 4290 of the genomic sequence (Figures 1 and 2). Translational initiation for the RNA represented by the cDNA could occur either at the AUG at position 1832 (the first in-frame methionine codon contained within the cDNA) or at the in-frame AUG at position 1650 (26 bp upstream of the end of the large cDNA). There are no in-frame AUG codons and there are numerous stop codons upstream of the potential initiator AUG at position 1650. We postulate that the AUG at position 1650 is the most likely point of initiation of translation. Confirmation of this prediction will require direct analysis of the amino terminus of the native protein.

The major ORF is followed by multiple termination codons the first of which is located at 4290. Translation of the genomic DNA beginning at the methionine at 1650 and terminating at 4290 after excision of the introns revealed by the cDNAs would produce a protein of 56.7 kd which is in good agreement with the reported monomer mol. wt of 54 kd, for native DDC (Clark *et al.*, 1978).

Comparing the restriction pattern of the cDNA and genomic


HindIII
100
AAGCTTTCTA TAGTAATTTT GAATTTTTCA TCAGTGTATG CGGAACTGCC CGCTCAAAAG GCTCAACCTA GCCCACTTCC CCTAGCACAA TGCGAAAGGT
200
GAGGTGAGAG CATTGGATTA TTTGACGTCA CAATTCCATG AGCGGTTCAA AAAGCAACGT CATATGTGGT GCTCTAATAA CCGGTTTCCA AGAAGCGCGT
300
AAAGCTGCCA TTCCACGGCT TAATCAATTT CTTGTCTTTC CTACGAATAT AACTTTGTTT ACATTTTTTG CGTGATTTTT TCTTCGGGGA GTCCAAGAAA
Eco RI
400
AACCCTGTTT CGAGTGACTC ATAATTGGGG GATTCCTGAC GAGATCGCTC TCTTTCCACA AATTCGAGTT GGGAAGCACG TGAGCAGAAT TCAAAATGTT
500
TTGCTTGCTG TTTTAAATAT CACTAGGTTC TCAAACTAAT TACAAAAATA ATCAAATTAA GTTCACAGAG CTGGCAAATA AAATGTAATA GCTTGCATGT
600
ATGTATATAT ATATATTTTT TTAAATTCTA AATAAATCCA TGAAAAATAA TGCCTTTGAT ATCCAGTTAC TGATTCAGCG CCCAATTAAT GCATGTTCCA
700
AAAAAGTGTC AAAAAACGTG CACAAATCAA ACGAGAGCAG AATTTGTTTT TACGACAGCG GCTGCGATTC GAAGTTCAGC GGCTGCGGAC TGCGATTGAA
800
CCGGTCCTGC GGAATTGGCA CGCTGCTGGA CGGGCTTTAA AAGCCATGGC CAAGAGCGGG CAGCGCTCAG TTAAGAGGAG AAGCCAAGCG CACAGCAATC
"tata"
\mRNA start----------
900
AGCACCGAAA TATCAGCATC GAAATATCAG CAAATAAATA TTAGCTGTTC TAAACCAGGA GGGCAAACTC AACTTGGAGC AAAGATTTAG TTCGGAACGG
Hpa I
1000
AAGTAAAGCT CGGCAACAAG TGCAAACAAT TAAAAGCAGT TAACTAAAGT GCAACGGTGA GAGACGAAAG TGTGGCTCCT CAACAGCCTC AGCTGCCTGA
splice donor
1100
AGTGCTTGGC CAACATAATG AGTGCATGTG CATGCGAAAG ATTCATTTCG GGGCTAACGC TGCGTATACG TAATGTGTAT CTAAAACTGG GCATATACTA
Eco RI
HindIII
1200
TAGCCTTGCT TCGGTTCAAT TTGATATGTT CGGGCCCCGA ATTCTATAGT GCTTAAGCTT TTCTCGGCTT TCGGTATCTG CATGCTTTTG TGTATCTATT
1300
AAAATAAGAT TTTAGCTGGC AACAAGTCGT CGTCTCAATG CCAACTTGTT TACGTTGTTA AAATTGGAAT TTAGAAAAAA AAAAAATATA AAGCAGTCTT
1400
GATTAATGCA AGAATGCATT AAACATTCTA ATTACCATAC TAATTCACAG CCTATACTTA AGCAGCGCAC TCGATGGGAA AACGCTTTAA ACTATTAATA
1500
CCTTAATACC TTATTATTAT AACTATTATC ATCATCGTTT TGCCTATCAA GTAATTAGTA TTCAATGTCG TTCATTTGTC GTGTTTGCAG CGATAGCATT
1600
TTTTGTTTGA GCTGCTGCAC TAATTAGCAC TATCTTCAAA AACGCACTTC TATTAATAAC ACTTTCAATA ATCGCACATT CTTTCATATT AGCTCTAACC
1640
1650
1680
ATTCGAGTTC ATATCATTGC AAAAGTCAAA CGAAAAGTAA ATCTCTGAA ATG AGC CAC ATA CCC ATT AGT AAC ACA ATT CCA ACA AAA CAA
splice acceptor 1 <----------Met Ser His Ile Pro Ile Ser Asn Thr Ile Pro Thr Lys Gln
5'end of long cDNA
1710
BamHI
1750
ACT GAT GGT AAT GGT AAA GCT AAC ATT TCG CCG GAT AAG CTG GAT CCC AAG GTT TCG G TATGTCTATT GGGTTTAGGT ATAGAGCCAA
Thr Asp Gly Asn Gly Lys Ala Asn Ile Ser Pro Asp Lys Leu Asp Pro Lys Val Ser splice donor
Splice acceptor 2 ----------MetSerIle GlyPheArg TyrArgAlaAs
1800
1830
1860
CAATTATGCA CGTCTGATAA CTAAATACTT TTGCATCCAC ATCAAG ATC GAC ATG GAG GCG CCG GAG TTC AAG GAT TTT GCC AAG ACA ATG
Splice acceptor Ile Asp Met Glu Ala Pro Glu Phe Lys Asp Phe Ala Lys Thr Met
nAsnTyrAla ArgLeuIle ThrLysTyrPhe CysIleHis IleLys Ile Asp Met Glu Ala Pro Glu Phe Lys Asp Phe Ala Lys Thr Met

```
                    1890                                1920                                1950
GTC GAC TTT ATA GCC GAA TAT CTG GAG AAT ATA CGC GAA AG GTGAGCCA GATTTAGACT TCCTACTCAA CTTAGCTTGA ATTAAACTTA
Val Asp Phe Ile Ala Glu Tyr Leu Glu Asn Ile Arg Glu Arg¦ splice donor
Val Asp Phe Ile Ala Glu Tyr Leu Glu Asn Ile Arg Glu Arg¦
                                        2000                                                2050
ATTTAGCGTA TAAATTTCAT TTATATGGTA TCAGAATCAG TCGCTTGACC TCAGCATTTT ACGTTCGAAT CGAAAGTTCG TTCTGCTCGA TTCGAATCCC

                                        2100                                                2150
CGGGCAAGTG AATGACATTT CGCACACGTT TTGAGATTAG TCACGGGAAA GTCGCACCGA TCGGACATTT CCATTGCTAT ATATATCTAT ATATATATAT

                                      Xho I                                                 2250
CATTTTGTTT AGGGGGTTGA GGCGACCTTC CCATTAGCTC GAGGGCCAAG TACTTTCGCT GCTCTTGGGC CGAAAACTAA TTAATTAAAT GGCTTTGTTG

                                        2300                                                2350
AGTTGGCGTG TCAAGGTCGT TTTTCATGTA TACGAGTATA GATATAATTG CACTCGTAAC GCCTTGGCCA AAAGCAATTC GGGTATTTCA CTATTCTTGG

                                        2400                                                2450
GCAATTCTTC TAACGGCTTC GTTTCCATTA CCTTGAAAAT CAAAGTCAGC TAAGTAAACA ATTTTCTTAC TACAGCTGCT GAGTTTGTTT GCCCATCGAC

                                    HindIII2500                                             2550
AGTCGCTGAA ATTAATGGTT AATTGAAAAT CAAGCTTAAG TAGAGCGTAA TATAATAATT CATTTTGCTT TATTAAAGTT CCTTCGACAT TGAAGTTTCA

                                        2600                                                2650
AAACTATTTT CTTAGTTAGA TAACTTTTTA AACGAATCTT TGTTAATTGA AGATACATAT ATATAGAGAA ATTATCTTTT TATTTTCTTT TTTTCACCTC

                                        2700                                                2750
TTAGTAGTAC TTCCTTTTAA TTGAAAGGAT AGAACAATCC CACCATCATT ATCAGCATTG CCTCTCTATC TATATTCTGT TCCCATAGCA ATTTGCTACA

                                        2800                                                2850
TATTCGTATT GATTTTGCAT GGCAGTGGCA ACAAGTTGGG GGTGGTTGGA TGGGTGTCCT TCAACCCCAA TGATTCCTGA TGCCTTTGTT GGCTAACTGA

                                        2900                                                2950
GTTTCGCAGC CAATTAGCAA GGAGCTTTTA CTGAATGGGC GCCAAAATGC AATCAGAACG TAAACGCATT TTCGCAATTA CAG G CGC GTT CTG CCG
                                                                    splice acceptor ¦  Arg Val Leu Pro

                                            3000
GAA GTG AAG CCT GGC TAC CTG AAG CCA TTG ATT CCG GAT GCT GCG CCC GAG AAG CCG GAG AAG TGG CAG GAT GTG ATG CAG GAC
Glu Val Lys Pro Gly Tyr Leu Lys Pro Leu Ile Pro Asp Ala Ala Pro Glu Lys Pro Glu Lys Trp Gln Asp Val Met Gln Asp

        3050                                                3100
ATC GAG CGA GTC ATC ATG CCG GGC GTG ACA CAC TGG CAC AGT CCC AAG TTT CAT GCC TAC TTC CCC ACG GCC AAC TCG TAT CCA
Ile Glu Arg Val Ile Met Pro Gly Val Thr His Trp His Ser Pro Lys Phe His Ala Tyr Phe Pro Thr Ala Asn Ser Tyr Pro

                                3150                                                3200  Xho I
GCG TAC GTT GCG GAC ATG CTG AGT GGA GCG ATT GCC TGC ATC GGA TTC ACG TGG ATC GCC AGT CCC GCG TGC ACG GAA CTC GAG
Ala Tyr Val Ala Asp Met Leu Ser Gly Ala Ile Ala Cys Ile Gly Phe Thr Trp Ile Ala Ser Pro Ala Cys Thr Glu Leu Glu

                                            3250
GTG GTC ATG ATG GAT TGG CTG GGC AAG ATG CTG GAG CTG CCG GCA GAG TTC CTG GCC TGT TCG GGC GGC AAG GGT GGC GGT GTC
Val Val Met Met Asp Trp Leu Gly Lys Met Leu Glu Leu Pro Ala Glu Phe Leu Ala Cys Ser Gly Gly Lys Gly Gly Gly Val

      3300                                                3350
ATC CAG GGC ACG GCC AGT GAG TCC ACA CTG GTG GCC TCT GCT GGG AGC CAA GGC CAA GAA GTT GAA GGA GGT GAA GGA GCT CCA
Ile Gln Gly Thr Ala Ser Glu Ser Thr Leu Val Ala Ser Ala Gly Ser Gln Gly Gln Glu Val Glu Gly Gly Glu Gly Ala Pro

                            3400                                                3450
TCC GGA GTG GGA CTG GAG CAC ACC ATC TTG GGC AAG TTG GTG GGC TAC TGC TCG GAC CAG GCT CAC TCA TCC GTG GAG CGG GCT
Ser Gly Val Gly Leu Glu His Thr Ile Leu Gly Lys Leu Val Gly Tyr Cys Ser Asp Gln Ala His Ser Ser Val Glu Arg Ala

                                        3500
GGT CTT CTG GGC GGA GTA AAG CTC CGT TCC GTG CAG TCC GAG AAT CAC AGA ATG CGT GGT GCT GCC CTG GAA AAG GCC ATC GAA
Gly Leu Leu Gly Gly Val Lys Leu Arg Ser Val Gln Ser Glu Asn His Arg Met Arg Gly Ala Ala Leu Glu Lys Ala Ile Glu
```

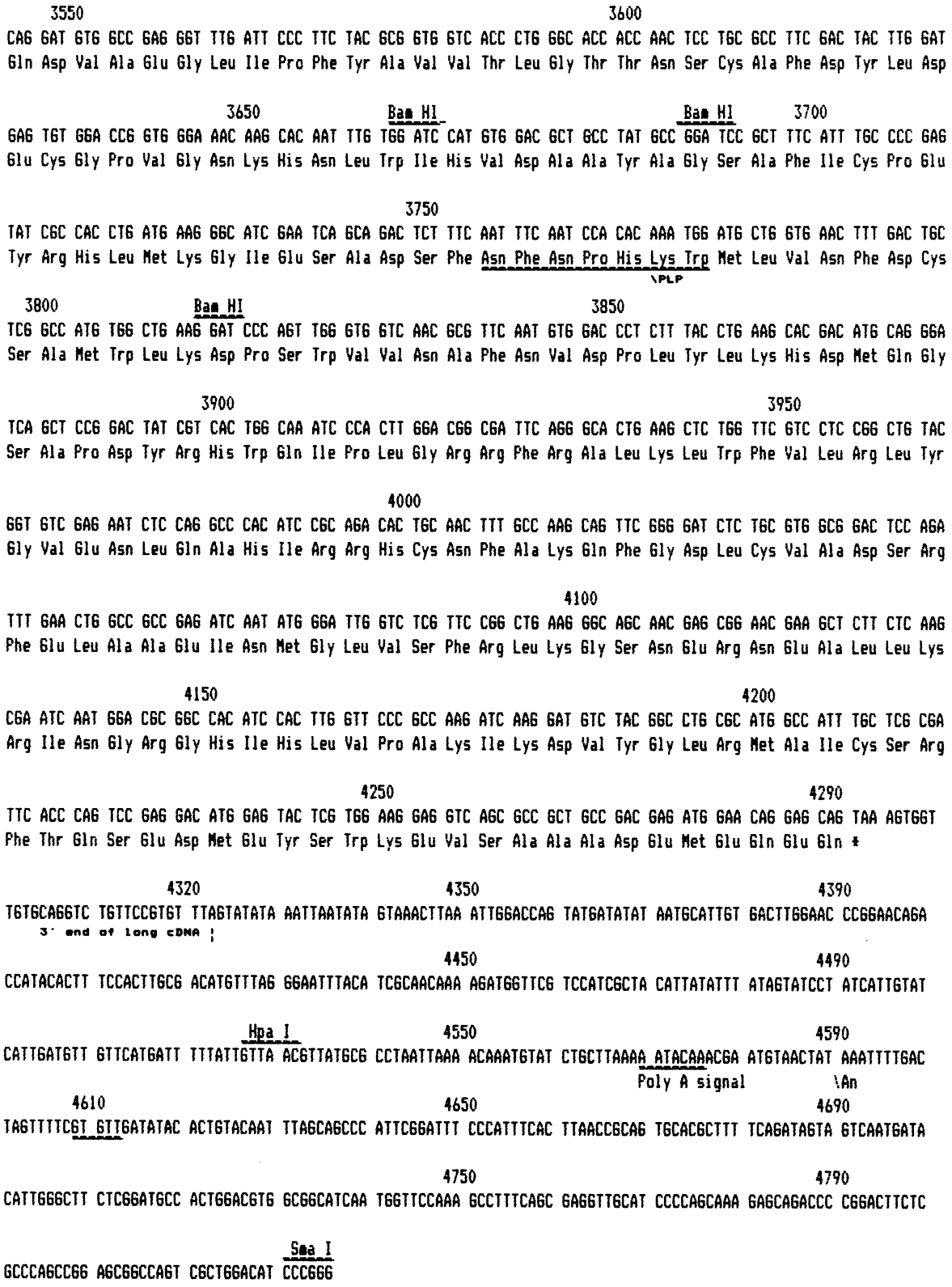

Fig. 2. Sequence of the gene. Splice junction sites are indicated by vertical lines. The deduced translation of the message is given underneath the sequence with the 2.3-kb message above and the 2.0-kb message below; dashes indicate untranslated regions of the mRNA. Since the translation of both messages is identical in the large third exon, only one is shown. The location of splice acceptor 1 is shown with an arrow to indicate that the precise location of this splice is upstream of the point indicated. Sequence motifs thought to be involved in initiation and polyadenylation are underlined (for details see text). The points of initiation and polyadenylation are indicated by slashes. The longer cDNA begins at position 1677 and ends at 4293 while the shorter cDNA begins at 3438 and ends at 4589. The 7-amino acid PLP binding domain which is identical to pig DDC is underlined and the PLP binding lysine indicated.

clones shows that they are identical in the region of the long ORF (data not shown) suggesting that no undetected introns are present in this region. In order to confirm the reading frame predicted by the sequence and to prove that no undetected interruptions to the reading frame (introns) occur within the large third exon, we devised an expression vector test of the predicted reading

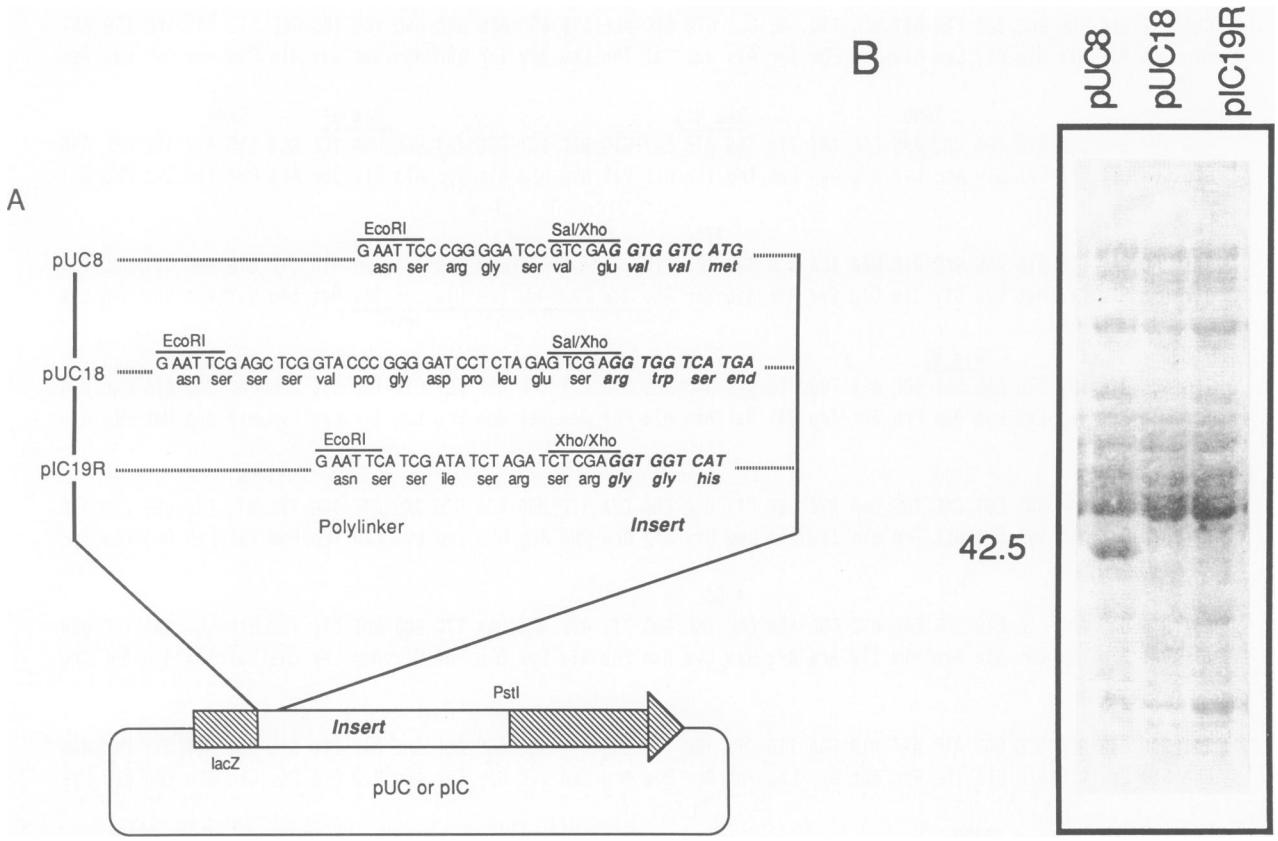

Fig. 3. *Ddc* expression vector. **A**, construction of plasmids. The sequence of the polylinker and the first few codons of the insert are shown with the translation underneath. The fragments were joined using complementary *Sal*I(vector)/*Xho*I(insert) ends in pUC8 and pUC18 or homologous *Xho*I ends in pIC19R. **B**, SDS electrophoresis of proteins from each construct. The novel protein expressed in the pUC8 construct is indicated.

frame using genomic DNA. A genomic fragment containing most of the third exon (i.e. the 'expression fragment' shown in Figure 1 beginning at the *Xho*I site at 3203 and extending ~1100 bp beyond the presumed stop codon) was inserted into the plasmids pUC8, pUC18 (Messing, 1983) and pIC19R (Marsh *et al.*, 1984), so as to give readthrough transcription from the *lac*Z promoter in all three reading frames. Each construct was tested for the production of a new protein by SDS−PAGE. Only the pUC8 construct produced a new protein (Figure 3). The reading frame in this plasmid is that predicted from the sequence and the size of the protein produced is in good agreement with the predictions of the sequence (i.e. 41.4 kd predicted versus 42.5 kd observed). This test provides experimental confirmation of the ORF and also defines the register of the reading frame at the *Xho*I site (at 3202). We conclude that the sequence data are highly accurate and that no undetected stop codons or small introns not multiples of 3 bp in length are present in this region.

Although the larger cDNA did not contain a string of As at its 3′ end, the smaller cDNA isolated from the same library did contain a string of As with the first A of the poly(A) tail at position 4590. This is in excellent agreement with the studies of Spencer *et al.* (1986) which show that *Ddc* RNAs from mature larvae and young adults terminate near the *Hpa*I site at position 4526. The polyadenylation signal AAUACA (underlined in Figure 2) is located 14 bp upstream from the polyadenylation site. This sequence functions as a polyadenylation signal in several genes (Wickens and Stephenson, 1984; Birnstiel *et al.*, 1985). The genomic sequences downstream of this region are GT rich similar to sequences seen in a number of vertebrate genes (Birnstiel *et al.*, 1985) and the sequence GTGTT occurs at position 4608 (underlined in Figure 2) and is thought to be required for efficient polyadenylation (McLaughlan, 1985).

Structure of the Ddc transcripts

The hybridization of the longest cDNA to the *Eco*RI/*Bam*HI fragment shown in Figure 1 indicates that it represents the 2.3-kb RNA. The cDNA contains three exons, one of at least 80 bases separated from a second exon of similar size by a previously undetected intron of only 77 bases, and a large third exon separated from the first two by an intron of 1031 bases. Comparison of the cDNA and genomic sequences suggests the structure for the 2.3-kb RNA shown in Figure 4. Excision of the 77-base intron and the 1031-base intron generates a single ORF, the majority of which comprises the large 3′ exon.

Beall and Hirsh (1984) have proposed that the 2.3-kb RNA is a processing intermediate leading to a mature 2.0-kb message and have proposed a structure for the 2.0-kb message on the basis of RNA blotting and R-loop analysis. We disagree with this interpretation because both the 2.0- and 2.3-kb RNA species are found on polysomes in approximately equal proportions (Gietz and Hodgetts, 1985), suggesting that both are translated, and because the cDNA indicates that the 2.3-kb mRNA exhibits all the elements of an mRNA. Accordingly, we propose that the 2.3-kb RNA is an alternative mature mRNA from the *Ddc* gene.

While we have not yet recovered a cDNA which would position the boundaries of the 2.0-kb message with precision, Beall and Hirsh (1984) have assigned an approximate structure based on R-looping experiments. Enough is known about RNA processing elements to propose a more precise version of the 2.0-kb RNA structure in light of the genomic sequence. The proposed

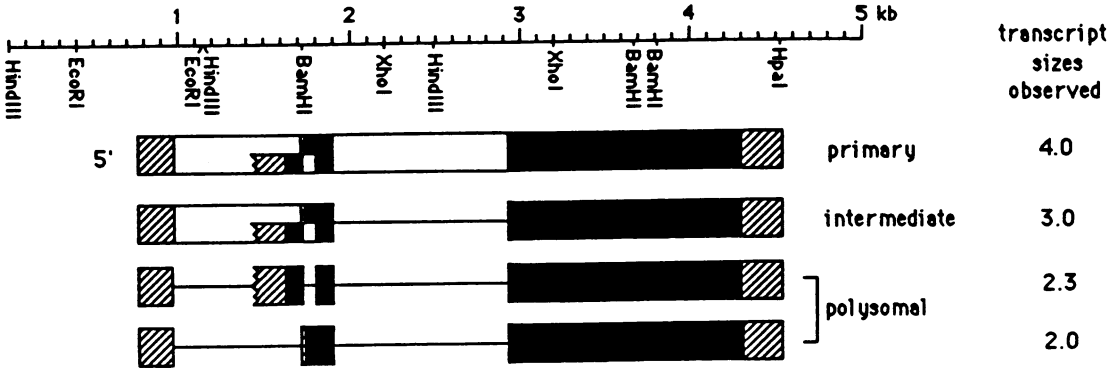

Fig. 4. Summary of RNA structures. The scale at the top is the same as in Figure 1 and is keyed to the sequence. Solid boxes correspond to coding regions, hatched areas represent 5'- and 3'-untranslated regions of the RNA, open boxes correspond to unexcised introns and thin lines represent introns excised in the RNA. The jagged 5' edge of the second exon of the 2.3-kb message signifies our uncertainty as to the precise location of the splice acceptor. Some of the 3.0-kb intermediate RNA species may have had the 77-base intron removed.

three exon, two intron structure can be reconciled with the genomic sequence by postulating that the 2.0-kb RNA shares the same 5' and 3' exons as the 2.3-kb species but differs in the region of the two small exons identified by the cDNA. Gietz and Hodgetts (1985) have shown that the second exon of the 2.0-kb RNA does not extend upstream of the *Bam*HI site (1732) (nor downstream of the *Xho*I site at 2197). These data and the best fit to the R-loop data of Beall and Hirsh (1984) place the second exon splice acceptor site of the 2.0-kb RNA near position 1742 on our sequence. There are two splice junction sites near 1742 which conform to the consensus for splice acceptors (Mount, 1982). A structure for the 2.0-kb RNA in which the first intron splice acceptor is located at 1742 (the 77-base intron of the 2.3-kb is not removed) and the large 3' intron is identical to that of the 2.3-kb mRNA would result in a 169-nucleotide second exon and a 1.5-kb terminal exon (Gietz and Hodgetts, 1985) whose properties are completely consistent with the R-loop and hybridization data. This structure would generate an intact ORF of 1509 bases. This model is also consistent with several lines of evidence which indicate that the splice junctions of the large 3' intron of both the 2.3-kb and 2.0-kb RNAs are similar or coincident. For instance, R-loop analysis (Beall and Hirsh, 1984) reveals only one splice junction in this region. Secondly, the incompletely processed 3.0-kb *Ddc* RNA species does not show size heterogeneity when both the 2.3-kb and 2.0-kb RNAs are present, which is consistent with the removal of only one large 3' intron. Finally, the only ORF in the sequence capable of encoding DDC is the one contained in the 1.5-kb exon defined by the cDNA. Taken together these observations strongly suggest that both species of mature *Ddc* message share the same 1031-base intron. Thus we propose that the 2.3-kb and 2.0-kb RNAs share similar or identical 5' and 3' ends but that they differ in the region of the second exon of each message.

The 2.0-kb *Ddc* message structure suggested here would predict that the AUG codon at position 1750 is used to initiate translation of DDC from this message as it is the first AUG codon in the message. This would result in a DDC protein which is slightly different at the N terminus compared with the translation of the 2.3-kb message but still in agreement with the measured size for the DDC enzyme (i.e. 56.2 and 56.7 kd for the 2.0- and 2.3-kb transcripts, respectively versus 54 kd for the native DDC monomer). Since the majority of the protein would be the same, both isoforms are expected to encode DDC enzyme activity.

Initiation of transcription

RNA blotting experiments show that all *Ddc* RNAs at all stages hybridize to the *Eco*RI/*Eco*RI fragment (positions 387−1139 on our sequence) (e.g. Gietz and Hodgetts, 1985; Beall and Hirsh, 1984). In addition, R-loop experiments suggest a small 5' *Ddc* exon completely contained within this fragment (Beall and Hirsh, 1984). Since neither of the cDNAs recovered extend into this region, we have positioned the RNA cap site using the S1 protection data of Scholnick *et al.* (1983) who have shown that late larval *Ddc* RNA contains a 200-base exon which is initiated at a site 370 bp upstream from the *Eco*RI site (1139). The sequence TTTAAAA is found 28 bp upstream of the proposed initiation site at 769 and conforms in six of seven positions to the eukaryotic promoter consensus sequence [5'-TATA(A/T)A(/T)-3'] (Breathnach and Chambon, 1981) and we postulate that this serves as the TATA portion of the promoter for this gene. No strong homology was noted to the consensus sequence (5'-GG[C/T]-CAATCT-3') often found ~50 bp upstream of the TATA box homology (Breathnach and Chambon, 1981) nor with the GGGCGG sequence implicated in a number of eukaryotic promoters (Dynan and Tjian, 1985). The transcription initiation site at 769 (1139−370) agrees with the consensus of Corden *et al.* (1980) while positioning the splice donor site at 970 provides the best fit to the extended splice donor consensus sequence of *Drosophila* (i.e. AG:GTnnnT, Keller and Noon, 1985). Placing the RNA cap at 769 and the 3' border of the first exon at 970 is entirely consistent with the S1 protection and R-loop results (Scholnick *et al.*, 1983; Beall and Hirsh, 1984). Since we have detected no AUG codons and no ORFs within this exon, we conclude that it does not contribute to the protein coding portion of either message.

A summary of the transcript structures consistent with the genomic and cDNA sequences is shown in Figure 4. Placement of the cap at 769 and the poly(A) tail at 4590 and inclusion of ~200 bases of poly(A) tail predicts a primary transcript of ~4.0 kb which is in excellent agreement with the observed 4.0-kb RNA species. Excision of the 1031-base 3' intron would generate an RNA species of 3.0 kb, also in excellent agreement with the observed 3.0-kb RNA. The 3.0-kb intermediate species can then be further processed into either the 2.3-kb or 2.0-kb mRNA structures. These two mature message structures are differentiated on the basis of hybridization to the *Eco*RI/*Bam*HI fragment.

	Donor		Acceptor	
970	AGACGAAAGTgtggctcctc	----------	ggatcccaagGTTTCGGTAT	1742
1747	CAAGGTTTCGgtatgtctat	----------	ccacatcaagATCGACATGG	1825
1910	TACGCGAAAGgtgagccaga	----------	gcaattacagGCGCGTTCTG	2942

Fig. 5. Splice junction sequences. The DNA sequence surrounding the proposed splice junctions of the various *Ddc* transcripts are shown. Capital letters indicate bases present in exons while small letters indicate bases contained in introns. The numbers next to the sequences refer to the position of the junction in the sequence presented in Figure 2. The 970 and 1910 splice donors and the 2942 splice acceptor are used in both forms of proposed messages.

Splice junctions are conserved

Since the *Ddc* message is differentially spliced and substantial quantities of immature message appear to accumulate, the splice junctions of the *Ddc* gene deserve special attention. The *Ddc* splice junctions are compared in Figure 5. It is interesting to note that the splice donor sequences of the two large exons are remarkably similar and match in 10 of 13 positions around the splice site. This may be related to the fact that these splice junctions are shared by both forms of mature *Ddc* message.

Discussion

Developmental regulation

During development both the 2.3-kb and 2.0-kb *Ddc* mRNAs appear in late embryos, evaginating discs and eclosing adults while only the 2.0-kb mRNA appears at pupariation (message precursors are present at all these stages). Since we propose that the 2.3-kb species is a mature mRNA, a reinterpretation of the data of Beall and Hirsh (1984), Gietz and Hodgetts (1985) and Clark *et al.* (1986) leads to the following conclusion. Two mature messages are expressed in late embryos, evaginating discs and adults while in the late third instar larval stage, only a single 2.0-kb mature messgae is expressed. Late larvae are unique in that the rise in DDC activity and mRNA is triggered by rising ecdysteroid titers while the appearance of DDC at other stages is correlated with falling hormone titers. These observations suggest that both the 2.3-kb and 2.0-kb messages respond to falling ecdysterone titers but that the 2.0-kb message is also produced in response to the rising ecdysterone titers at pupariation. Precedence for both a positive and negative response to ecdysterone is found in *Manduca* where Hiruma and Riddiford (1985) found that 20-OH ecdysone acts directly on the larval epidermis to inhibit DDC synthesis and concluded that DDC activity is expressed in response to falling ecdysterone titers at the larval molts with the amplitude of the response determined hours earlier by the levels of juvenile hormone. In contrast, the appearance of DDC at pupariation responds directly to ecdysterone if the conversion to pupal committment of the larval epidermal cells has been potentiated by an earlier ecdysone pulse in the absence of juvenile hormone (e.g. Riddiford, 1981). Measurements of juvenile hormone in *Drosophila* suggest a similar pattern (Sliter *et al.*, 1986). The interplay of juvenile hormone and ecdysterone may provide the basis for this apparently complex pattern of response.

Since the *Ddc* gene responds to ecdysterone at pupariation, we have searched for sequences similar to those upstream of other hormonally regulated *Drosophila* genes. A sequence beginning at 540 (ATGAAAAATAATGCCTTT) bears a strong resemblance to similarly located sequences in the hormonally regulated *Sgs-4* gene (ATGGAAA--TACCTTT) (Muskavitch and Hogness, 1982) and 74EF genes (ATGGAAC) (Moritz *et al.*, 1984). This sequence bears considerable homology to the canonical enhancer sequence of mammalian viruses (Weiher *et al.*, 1983) [(G)TGGAAA(G)] and is altered in the *Ddc*$^{+4}$ overproducer variant of Estelle and Hodgetts (1984a, 1984b) to more closely resemble the canonical enhancer (Hodgetts *et al.*, 1986). The *Ddc* sequence does contain one perfect example of the MMTV glucocorticoid receptor binding hexamer (TGTTCT at position 846) found by Scheidereit *et al.*, (1983) but does not contain any regions homologous to the progesterone receptor binding site in the ovalbumin gene (Mulvihill *et al.*, 1982).

The proposed *Ddc* transcripts both exhibit an unusually long 5′-untranslated leader (200 bases) which is interrupted by an intron of ~775 bases. In a compilation of 211 leader sequences from eukaryotic genes, Kozak (1981) finds >70% of the leaders clustered in the 20−80 nucleotide range and speculates that the longer leaders may participate in unusual modes of regulation. The transcript structures we have proposed do not contain any premature AUG codons, thus eliminating the possibility that one function of the long leader is to regulate translation using false starts (Hunt, 1985).

The Ddc gene may encode two proteins

The DNA sequence suggests that the two alternative forms of mature mRNA encode two polypeptides which may differ by as much as 25 amino acids in the amino terminus of the DDC enzyme. If the 2.3-kb message initiates translation at the AUG codon at 1650, it would encode a protein of 56.7 kd while initiation at the first AUG codon of the 2.0-kb message we have proposed (i.e. 1749) would encode a protein of 56.2 kd. Alternatively, either or both messages could ignore these AUGs and initiate at the AUG at position 1832. However in the absence of direct evidence identifying the amino terminus of the DDC protein, we must assume that the first available AUG in the mRNA initiates translation. Although double diffusion immunoprecipitation revealing only a single precipitin line using embryo, larval and adult extracts has been cited as evidence for a single molecular form of DDC (Clark *et al.*, 1978), a difference of 1% in protein mol. wt would not be detected in that assay. Additional biochemical studies will be required to directly demonstrate the presence of these two putative DDC isoforms.

The relative abundance of the 2.0- and 2.3-kb *Ddc* mRNAs suggest that the two hypothetical DDC isoforms may both be expressed in the epidermis where the majority of DDC activity is found. The significance of the variation in amino terminus between the two forms in mediating the potentially different roles of DDC awaits determination of their inter- and intra-cellular distribution.

All decarboxylases employ a pyridoxal co-factor linked to the ϵ amino group of a lysine residue. The pyridoxal phosphate (PLP) binding peptides for a few amino acid decarboxylases and other PLP enzymes are known (e.g. Tanase *et al.*, 1979) and we have compared these with the *Drosophila* DDC sequence. The seven-amino acid PLP binding peptide of pig DDC (Bossa *et al.*, 1977) matches perfectly with a 7-amino acid stretch of *Drosophila* DDC (i.e. nucleotides 3754−3771 in Figure 2). No strong homology with the PLP binding region of other amino acid decarboxylases was found. This homology provides strong evidence that the lysine at 3769 is the PLP binding lysine in *Drosophila*. The perfect homology between pig and fly DDC and lack of homology with other amino acid decarboxylases suggests that the amino acid sequence surrounding the PLP lysine may be substrate specific and suggests further that considerable sequence homology may exist between vertebrate and invertebrate DDCs. We have also observed that the *Ddc* gene exhibits strong sequence homology with

the adjacent α methyl dopa hypersensitive gene in *Drosophila* (Marsh *et al.*, 1986; Eveleth and Marsh, 1986). These studies suggest that the *Ddc* gene represents one member of a clustered family of genes involved in catecholamine biosynthesis in *Drosophila*.

The DNA sequence reported here is the first sequence of a dopa decarboxylase gene from any species. The DDC enzymes in insects and mammals are remarkably similar in subunit structure, mol. wt, kinetic properties and substrate specificities (Maneckjee and Baylin, 1983). The sequence presented here will provide the basis for a comparison of this fundamentally important enzyme from insects and mammals.

Materials and methods

General

Preparation of plasmid DNA, restriction analysis and subcloning were as described by Maniatis *et al.* (1982). Enzymes were obtained either from Boehringer Mannheim or P-L Biochemicals with the exception of reverse transcriptase (Bio-Rad).

DNA sequence was determined by the dideoxy chain termination method of Sanger *et al.* (1977). Growing and manipulating the phage was performed as described by Messing (1983). Reactions were performed using nucleotides from P-L Biochemicals following the manufacturer's protocol with minor modifications. Reactions were run on 0.25 mm × 40 cm urea gels at 50 W. After running, the gel was dried by bonding directly to one of the glass plates which had been treated with 3-(trimethylsilyl)propyl methacrylate (Aldrich). The gel was then exposed to X-ray film. DNA sequences were assembled and analysed using a series of computer programs modified from Schwindinger and Warner (1984) by Dr.Al Goldin (California Institute of Technology).

A cDNA library was prepared from 0–2 h old adult Canton-S poly(A)$^+$ RNA. RNA extraction and oligo(dT) chromatography were carried out following Gietz and Hodgetts (1985). The protocol used in preparing the library follows that of St. John *et al.* (personal communication). The first cDNA strand was made with reverse transcriptase. The mRNA:DNA hybrids were sized on Sepharose 4B and hybrids >500 bp were pooled and tailed with dGTP using terminal transferase. RNA was removed by digestion with RNase A and oligo(dC) was annealed to the tailed first strand. Second strand synthesis was performed using T4 DNA polymerase following the method of Burd and Wells (1974). Internal *Eco*RI sites were protected with *Eco*RI methylase.

The double-stranded cDNA was divided into equal aliquots and one was digested with limiting amounts of S1 nuclease following the methods of Vogt (1980). Both DNA samples were spermine precipitated (Hoopes and McClure, 1981) prior to ligation with *Eco*RI decanucleotide linkers using T4 DNA ligase. After *Eco*RI digestion, the cDNAs were fractionated again on Sepharose 4B and concentrated using spermine precipitation. The cDNA was ligated into *Eco*RI-digested λgt10 DNA and packaged *in vitro* (Promega Biotech).

Gene fusions were constructed using an *Xho*I–*Pst*I fragment of pP10 (described in Marsh *et al.*, 1985) containing most of the third exon of *Ddc* (i.e. 1089 bp of coding region beginning at the *Xho*I site at 3202; the fragment extends 1100 bp beyond the proposed translation termination site). This fragment, illustrated as 'expression fragment' in Figure 1, was ligated into *Sal*I–*Pst*I cut pUC8, *Sal*I–*Pst*I cut pUC18 (Messing, 1983) and *Xho*I–*Pst*I cut pIC19R (Marsh *et al.*, 1984). This leads to readthrough transcription from the *lacZ* promoter in each of three reading frames. The identity of each construct was confirmed by restriction analysis (data not shown). An example of each construct was grown overnight in 10 ml of B-broth to avoid catabolite repression of *lacZ* (B-broth, Messing, 1983). 1 ml of overnight culture was diluted into 10 ml of fresh B-broth and grown for 4–6 h (the optimal time of growth for maximal expression was determined empirically). Cells were harvested from 10 ml by centrifugation and proteins extracted by boiling the cell pellet in 1.0 ml of 20% glycerol, 10% β-mercaptoethanol, 6% SDS, 10 mM Tris pH 8.0. Debris was removed by centrifugation at 12 000 *g* for 15 min and 25 μl of the supernatant was loaded onto a discontinuous polyacrylamide slab gel (Laemmli, 1970).

Acknowledgements

We are indebted to M.P.Erfle, C.A.Leeds, S.S.Ogle and W.C.Clark for their expert technical assistance; to Dr R.Steele, whose comments were seminal in developing the expression assay and to Dr K.Konrad for his help in analysing the expressed proteins. We appreciate the constructive comments of Drs P.Bryant, R.Davis, C.Greer, J.Manning and others. 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. This work was supported by an NIH grant GM28972 and an NSF grant PCM 8316485 to J.L.M. and by grant #A6477 from the NSERC of Canada to RBH. C.A.S. and R.D.G. gratefully acknowledge the support of the Alberta Heritage Foundation for Medical Research and D.D.E. gratefully acknowledges support from PHS predoctoral training grant T32 CA09054.

References

Beall,C.J. and Hirsh,J. (1984) *Mol. Cell Biol.*, **4**, 1669–1674.
Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180–182.
Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349–359.
Bossa,F., Martini,F., Barra,D., Borri Voltattorni,C., Minelli,A. and Turano,C. (1977) *Biochem. Biophys. Res. Commun.*, **78**, 177–184.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.
Burd,J.F. and Wells,R.D. (1974) *J. Biol. Chem.*, **249**, 7094–7101.
Clark,W.C., Pass,P.S., Venkataraman,B. and Hodgetts,R.B. (1978) *Mol. Gen. Genet.*, **162**, 287–297.
Clark,W.C., Doctor,J., Fristrom,J.W., and Hodgetts,R.B. (1986) *Dev. Biol.*, **114**, 141–150.
Corden,J., Wasylyk,B., Buchwalder,A., Sassone-Corsi,P., Kedinger,C. and Chambon,P. (1980) *Science*, **209**, 1406–1411.
Dynan,W.S. and Tjian,R. (1985) *Nature*, **316**, 774–778.
Estelle,M.A. and Hodgetts,R.B. (1984a) *Mol. Gen. Genet.*, **195**, 434–441.
Estelle,M.A. and Hodgetts,R.B. (1984b) *Mol. Gen. Genet.*, **195**, 442–451.
Eveleth,D.D. and Marsh,J.L. (1986) *Genetics*, in press.
Gietz,R.D. and Hodgetts,R.B. (1985) *Dev. Biol.*, **107**, 142–155.
Hirsh,J. and Davidson,N. (1981) *Mol. Cell. Biol.*, **1**, 475–485.
Hiruma,K. and Riddiford,L.M. (1985) *Dev. Biol.*, **110**, 509–513.
Hodgetts,R.B., Clark,W.C., Eveleth,D.D.Jr, Gietz,R.D., Spencer,C.A. and Marsh,J.L. (1986) In Slavkin,H.C. (ed.), *Progress in Developmental Biology Part A.* Allan R.Liss, NY, pp. 221–234.
Hoopes,B.C. and McClure,W.R. (1981) *Nucleic Acids Res.*, **9**, 5493–5505.
Hunt,T. (1985) *Nature*, **316**, 580–581.
Keller,E.B. and Noon,W.A. (1985) *Nucleic Acids Res.*, **13**, 4971–4981.
Kozak,M. (1981) *Nucleic Acids Res.*,**9**, 5233–5252.
Kraminsky,G.P., Clark,W.C., Estelle,M.A., Gietz,R.D., Sage,B.A., O'Connor, J.D. and Hodgetts,R.B. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 4175–4179.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Livingstone,M.S. and Tempel,B.L. (1984) *Nature*, **303**, 67–70.
Lunan,K.D., and Mitchell,H.K. (1969) *Arch. Biochem. Biophys.*, **132**, 450–456.
Maneckjee,R. and Baylin,S.B. (1983) *Biochemistry*, **22**, 6058–6063.
Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,G.K. and Efstradiatis,A. (1978) *Cell*, **15**, 687–701.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual.* Cold Spring Harbor Laboratory Press, NY.
Marsh,J.L. and Wright,T.R.F. (1979) In Axel,R. and Maniatis,T. (eds), *Eukaryotic Gene Regulation, ICN-UCLA Symposium on Molecular and Cell Biology.* Vol. **XIV**, pp. 183–194.
Marsh,J.L. and Wright,T.R.F. (1980) *Dev. Biol.*, **80**, 379–387.
Marsh,J.L. and Wright,T.R.F. (1986) *Genetics*, **112**, 249–265.
Marsh,J.L., Erfle,M. and Wykes,E.J. (1984) *Gene*, **32**, 481–485.
Marsh,J.L., Gibbs,P.D.L. and Timmons,P.M. (1985) *Mol. Gen. Genet.*, **198**, 393–403.
Marsh,J.L., Erfle,M.P. and Leeds,C.A. (1986) *Genetics*, in press.
McLaughlan,J., Gaffney,D., Whitton,J.L. and Clements,J.B. (1985) *Nucleic Acids Res.*, **13**, 1347.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.
Moritz,T.H., Edstrom,J.E. and Pongs,O. (1984) *EMBO J.*, **3**, 289–295.
Mount,S.M. (1982) *Nucleic Acid Res.*, **10**, 459–472.
Mulvihill,E.R., LePennec, J.-P. and Chambon,P. (1982) *EMBO J.*, **24**, 621–632.
Muskavitch,M.A.T. and Hogness,D.S. (1982) *Cell*, **29**, 1041–1051.
Riddiford,L.M. (1981) *Am. Zool.*, **21**, 751–762.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Scheidereit,C., Geisse,S., Westphal,H.M. and Beato,M. (1983) *Nature*, **304**, 749–752.
Scholnick,S.B., Morgan,B.A. and Hirsh,J. (1983) *Cell*, **34**, 37–45.
Schwindinger,W.F. and Warner,J.R. (1984) *Nucleic Acid Res.*, **12**, 601–604.
Sliter,T.J., Sedlak,B.J., Baker,F.C. and Schooley,D.A. (1986) *Insect Biochem.*, in press.
Spencer,C.A., Gietz,R.D. and Hodgetts,R.B. (1986) *Dev. Biol.*, **114**, 260–264.
Tanase,S., Kojima,H. and Morino,Y. (1979) *Biochemistry*, **18**, 3002–3007.
Tempel,B.L., Livingstone,M.S. and Quinn,W.G. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3577–3581.
Vogt,V.M. (1980) *Methods Enzymol.*, **65**, 248–255.
Weiher,H., Konig,M. and Gruss,P. (1983) *Science*, **219**, 626–631.
Wickens,M. and Stephenson,P. (1984) *Science*, **226**, 1045–1051.

Wright,T.R.F. (1977) *Am. Zool.*, **17**, 707–721.
Wright,T.R.F., Steward,R., Bentley,K.W., and Adler,P.N. (1981) *Dev. Genet.*, **2**, 223–235.
Wright,T.R.F., Black,B.C., Bishop,C.P., Marsh,J.L., Pentz,E.S., Steward,R. and Wright,E.Y. (1982) *Mol. Gen. Genet.*, **188**, 18–26.

Received on 28 May 1986; revised on 21 July 1986