Drosophila Arginase Is Produced from a Nonvital Gene That Contains the elav Locus within Its Third Intron*

Received for publication, February 18, 2000, and in revised form, June 28, 2000
Published, JBC Papers in Press, June 30, 2000, DOI 10.1074/jbc.M001346200

Marie-Laure Samson‡
From the Department of Biochemistry and Molecular Biology, University of Nebraska Medical Center, Omaha, Nebraska 68198-4525 and the Laboratoire d’Embryologie Moléculaire et Expérimentale, UPRES-A 8080, Université Paris Sud, 91405 Orsay, CEDEX France

A Drosophila gene encoding a 351-amino acid-long predicted arginase (40% identity with vertebrate arginases) is reported. Interestingly, the third intron of the arginase gene includes the elav locus, whose coding sequence is on the complementary DNA strand to that of the arginase. Terrestrial vertebrates produce two arginases from duplicated genes. One form, essentially present in the liver, is a key enzyme of the urea cycle and eliminates excess ammonia through the excretion of urea. The function of the extrahepatic arginase, more ubiquitous, is not well understood. In macrophages, arginase competes with nitric-oxide synthase, which converts arginine into nitric oxide. Most organisms, including insects, produce only one type of arginase, whose function is not centered on ammonia detoxification. A Drosophila cDNA encoding a predicted arginase was isolated. It produces a 1.3-kilobase transcript present with highest levels toward the end of embryogenesis and thereafter. During embryogenesis, the arginase transcripts localize to the fat body. The first mutant allele of the Drosophila arginase gene was identified. It is predicted to produce a 199-amino acid-long C-terminally truncated protein, likely to be inactive. Preliminary characterization of the mutation shows that this recessive allele causes a developmental delay but does not affect viability.

Large scale eucaryotic genome sequencing and the generation of expressed sequence tags, including those from the yeast Saccharomyces cerevisiae (1), the nematode Caenorhabditis elegans (2, 3), and the fruit fly Drosophila melanogaster (4–6), are providing a great deal of information about gene organization and genome evolution. Twenty-three years ago, with the discovery of splicing, the basic idea of a gene as a "block" of information and genome evolution. Twenty-three years ago, with the discovery of splicing, the basic idea of a gene as a "block" of DNA was challenged. Today, exciting new observations are being made. What initially seemed peculiar oddities, such as operons in nematodes and genes nested within genes in fruit flies, are turning out to be more general phenomena. About 25% of C. elegans genes are in operons (for a review see Ref. 7), and one estimate suggests that 7% of D. melanogaster genes may be nested within others (5).

In this paper, I describe a Drosophila gene coding for an arginase that contains within its third intron the entire locus elav, a well characterized 13-kb-long gene that encodes an RNA-binding protein specifically present in all neurons (for a review see Ref. 8). The first identified so-called nested gene (Pcp) encodes a Drosophila pupal cuticle protein and maps within the intron of the adenosine 3 gene encoding guanine adenosine ribosyl transferase (9). Since then, more than 30 nested genes have been identified in Drosophila (10). The recent sequencing of 2.7 megabases of DNA in the Adh region revealed 17 new ones (5). Nested genes are not specific to Drosophila, and examples are known in humans (11, 12) and in mice (13). Among the 49 identified Drosophila nested genes, 37 are transcribed in the direction opposite to the direction of transcription of the gene into which they are inserted (5).

The gene that is described in this paper encodes a predicted polypeptide that shares 40–41% amino acid identity with human arginases (for a review see Ref. 14). It will be referred to as the arginase gene, or arg. Arginases convert arginine into urea and ornithine. Sequence analysis of arginase and arginase-like sequences suggests that arginase was probably present in the primordial ancestor, before the divergence of euca-

The effect of alteration in the activity of extrahepatic A-II

* This work was supported in part by Basic Research Grant FY96-0995/FY97-0605 from the March of Dimes Birth Defect Foundation. Sequencing was performed at the Eppley Cancer Center Molecular Biology Core Facility at the University of Nebraska Medical Center. Sequence analysis was accomplished through the use of the Genetic Sequence Analysis Facility at the University of Nebraska Medical Center and the Wisconsin Package software from the Genetics Computer Group. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Present address: Laboratoire d’Embryologie Moléculaire et Expérimentale, UPRES-A 8080, Bâtiment 445, Université Paris Sud, 91405 Orsay, CEDEX France. Tel.: 33-1-69-15-75-85; Fax: 33-1-69-15-68-02; E-mail: Marie-Laure.Samson@emex.u-psud.fr.

The abbreviations used are: elav, embryonic lethal abnormal visual system; kb, kilobase(s); A-I, arginase I (hepatic form); A-II, arginase II (nonhepatic form); PCR, polymerase chain reaction; ORF, open reading frame; EST, expressed sequence tag.
Drosophila Arginase

has not been described. It has been proposed to be involved in the synthesis of polyamines, amino acids, and neurotransmitters and competes with nitric-oxide synthase in macrophages, but its role remains poorly defined (for reviews see Refs. 14 and 17). It has been suggested that the mitochondrial A-II is the surviving form of the ancestral arginase, because the cytosolic A-I is restricted to a subset of more recently evolved species (14).

Aside from ureotelic organisms, most others produce only one form of arginase. In particular, in S. cerevisiae, arginase is encoded by the CAR1 gene, which has been extensively characterized (18, 19). CAR1 encodes a cytosolic arginase that converts arginine to urea and ornithine. Although S. cerevisiae produces all the enzymes required to complete the urea cycle, compartmental separation of metabolic pathways and sophisticated regulation lead to the absence of a functional urea cycle. Sequencing of the S. cerevisiae genome did not reveal any additional arginase-related sequences (1). In C. elegans, whose entire genome has also been sequenced (2, 3), only one gene showing significant relationship with known arginase genes has been identified (20). No biological information is available yet on the function of this gene.

Animals have evolved pathways adapted to their lifestyle for the excretion of urea, uric acid, or ammonia as the major nitrogenous end product. The excretion of urea produced through the urea cycle is specific for ureotelic organisms. Insects (uricotelic) convert most of their excess ammonia to uric acid, an oxidized purine, via an arginase-independent pathway. Thus, the arginase function in uricotelic organisms, and more generally in organisms that do not have a functional urea cycle, is not centered on ammonia detoxification and remains poorly understood. Analysis of the arginase gene of Drosophila, amenable to molecular and genetic analysis, should give insights into this other aspects of arginase activity.

**EXPERIMENTAL PROCEDURES**

**Screening of the cDNA Library**—A Drosophila head cDNA library in the agt11 vector (a gift from Paul Salvaterra; Ref. 21) was screened with a genomic probe corresponding to a mixture of two radiolabeled BstNI restriction fragments (kb 8–18 on the scale in Fig. 1) to obtain elav cDNAs clones. Among the cDNAs isolated in this screen, one clone (cDNA-30h) originated from a transcription unit different from elav. Its analysis is reported in this paper.

**Characterization of cDNA-30h**—cDNA-30h phage lysates were prepared on plates and used to generate PCR products corresponding to the cDNA insert. Amplification mixtures were as per Taq DNA polymerase manufacturer’s specifications (Promega), with 2 μl of lysate/25-μl reaction, using a pair of primers hybridizing to regions flanking the cDNA insert inagt11, respectively g11F (5′-GGTTGGCAGGGATCTTCGAGAC-3′) and g11R (5′-GACACGACAACTGTTAAT-3′). Amplification yield was not affected by the lysate titer in the 10^{3}–10^{5} plaque-forming unit/ml range. Taq polymerase addition to the PCR mixture was performed after the initial step of hot start (95 °C for 1 min and 80 °C for 15 min). Amplification was performed with 30 cycles of 95 °C for 1 min, 55 °C for 1.5 min, and 72 °C for 1.5 min. After purification on the Wizard PCR prep purification system (Promega), the PCR product was sequenced using g11 primers and subsequently primers hybridizing in the cDNA sequence.

**Isolation of a Head cDNA arginase Clone**—A head cDNA library was screened with a genomic probe corresponding to the presumptive 3′-untranslated region of the elav gene (kb 8–18 on the scale in Fig. 1). Sequencing of the cDNA clones recovered in the screen identified a new transcription unit. The cDNA corresponding to the new transcription unit (cDNA-30h) is 1170 base pairs long. It contains a 349-amino acid-long open reading frame (nucleotides 2–1051), which begins with a tryptophan, indicating that the 5′ region of the mRNA is truncated in cDNA-30h. Searches of protein sequence data bases with the ORF present in cDNA-30h revealed its similarity with arginases.

The arginase sequences of ureotelic (H. sapiens, M. musculus, and X. laevis) and nonureotelic (S. cerevisiae and C. elegans) organisms were compared with the Drosophila putative arginase sequence. Arginases contain about 300 amino acids. They include three conserved regions (arginase signatures 1, 2, and 3 from the N terminus to the C terminus) with charged residues involved in the binding of two manganese ions to histidines and aspartic acids in the motifs (26). Comparison of protein sequences highlights the presence of these three arginase family signatures in the Drosophila sequence (Fig. 2).

Drosophila Arginase.

molecular biology core facility on Applied Biosystems and Li-Cor sequencing was performed after the initial step of hot start (95 °C for 1 min). Amplification was performed with 30 cycles of 95 °C for 1 min, 55 °C for 1.5 min, and 72 °C for 1.5 min.

Sequence determination of phage and genomic PCR products were performed by the University of Nebraska Medical Center/Ripple mo...
motifs are present in all the compared sequences, and all the invariant residues of the signatures are present. However some positions do not match the defined signature sequences. First, in arginase signature 1, Cys is found in C. elegans where Leu, Ile, Val, Met, or Thr is expected, and Gly is found in D. melanogaster where Ser, Thr, Ala, or Val is expected. Second, in arginase signature 3, Ser is found in D. melanogaster and Val is found in C. elegans where Pro, Ala, or Gln is expected. Extensive alignments of arginase sequences from different species highlight invariant residues (14, 20). Sequences from 19 arginases from vertebrates and procaryotic/eucaryotic unicellular organisms share 37 invariant amino acids (20), with 36 of them present in the D. melanogaster arginase and 27 of them present in the C. elegans arginase-related sequence (Fig. 2).

Pairwise comparisons (Fig. 3) of vertebrate A-II show that their sequences are 71–85% identical. A-I arginases are 64–87% identical. However, comparisons between the A-I and A-II arginases shows a lower degree of similarity, with 54–67% identity. This is consistent with the model of an arginase gene duplication giving rise to the genes for A-I and A-II (15, 20). In contrast, the three proteins from invertebrates similarly resemble hepatic and nonhepatic vertebrate arginases, the identity levels with vertebrate arginases being, respectively, 39–40% for D. melanogaster arginase, 40–45% for S. cerevisiae, and 26–31% for C. elegans. This is consistent with a duplication of the arginase gene giving rise to the genes producing hepatic and extrahepatic arginases after divergence between the vertebrates and the invertebrates (15, 20).

The Arginase Gene Spans about 20 kb, and Its Third Intron Contains the elav Gene—Alignment of cDNA-30h with genomic sequence shows that it derives from regions that flank the elav gene (Fig. 1). Based upon the polarity of the arg ORF in cDNA-30h, complementary strands of DNA are transcribed to give rise to the elav and arg transcripts. Limited sequence information was available when the cDNA was first identified (kb 0–18 on scale in Fig. 1; GenBank[TM] Accession Number AF047180; Ref. 27) and the data suggested that the 3’ end of cDNA-30h derived from a region of DNA upstream of the elav gene. Genomic PCR of D. melanogaster DNA using primers ARG2 (designed according to cDNA-30h sequence) and ARG3 (designed according to the genomic sequence, see Fig. 1) was performed to demonstrate that elav is within an arg intron. A 1-kb-long PCR product was obtained and sequencing of its ends proved its specificity (not shown). Subsequent genomic sequencing of the entire region (6), with data from cosmid 171D11 (accession AL009147) and from cosmid 65F1 (accession AL022139) around position kb 14 of the scale in Fig. 1, is in agreement with these data.

The cDNA-30h clone is truncated at its 5’ end, because the ORF that it contains begins with a tryptophan at nucleotide 2 of the sequence. Four arg expressed sequence tags (ESTs) from Drosophila adult head tissue have been reported (4). ESTs are sequences of the 5’ or 3’ end of a cDNA that are generated to rapidly identify expressed genes in the genome. None of these ESTs overlaps the junction between exon 3 and 4, but they provide useful information about the 5’ and 3’ ends of the arg RNA. Two of the reported 5’ ESTs (GH02581 and GH02569) begin 99 nucleotides upstream of the first nucleotide of cDNA-30h, where their sequence is colinear with genomic DNA sequence. They are likely to define the actual 5’ end of the transcript, which generated CDNA-30h. Their structure was taken into account to predict the true N terminus of the arg ORF, which includes two amino acids (methionine initiation codon, then tryptophan) upstream of the cDNA-30h ORF, as shown in Fig. 1.

The ORF in cDNA-30h is followed by 119 nucleotides, ending with a stretch of eight A residues, likely corresponding to the complete 3’-untranslated region of the mRNA. This is consistent with the structure of the 481-base pair-long arg 3’ EST (GH02581), which identifies the same 3’ end for the arg RNA.

Developmental Expression of the Arginase Gene—The expression of arg was examined during the course of development using single-stranded RNA probes on Northern blots. A single 1.3-kb-long arg transcript was identified (Fig. 4) whose size is consistent with cDNA-30h sequencing and EST data. arg transcripts start accumulating during the last stages of embryogenesis (16-h-old embryos), quickly reach a plateau at 20 h of development, and are thereafter found at relatively high levels. Not surprisingly, the arg transcript is present, but not significantly enriched, in adult head RNA extracts (Fig. 4, lanes 10 and 11), consistent with the fact that cDNA-30h and the four arginase ESTs derive from head cDNA libraries.

The Structure of the Mature elav and arginase mRNAs Are Not Overlapping—The arg and elav mRNAs arise from transcription of the same genomic region in opposite directions. Based upon data from cDNA-30h, genomic DNA, and the arg ESTs, the structure of the mature arg mRNA is well defined. In contrast, multiple developmentally regulated elav mRNA are produced from the elav gene (8, 28), and their precise structure has not been determined. Thus, it is unclear how much overlap may exist between mature transcripts of arg and elav.

An attempt was made to re-evaluate the proposal (based upon Northern blot analysis using double-stranded DNA probes) that the 3’ region of one form of elav mRNA derives from the 2.3-kb BamHI genomic fragment (kb 16–18 on the scale Fig. 1; Refs. 25, 27, and 28) to where the 5’ region of arg maps. Northern blot analysis was performed using single-stranded RNA probes corresponding to the 2.3-kb BamHI genomic fragment. It shows that no transcripts are detected by the RNA probe colinear to antisense elav RNA (Fig. 5B), whereas two transcripts, respectively 1.3 and 5.5 kb long, are detected by the RNA probe colinear to antisense arg RNA (Fig. 5C). The 1.3-kb-long transcript detected in Fig. 5C is the expected arg transcript, similarly detected by the cDNA-30h antisense probe (Fig. 5A). The 5.5-kb-long transcript migrates similarly to one of the major elav transcripts (Fig. 5D) but is
produced from the DNA strand complementary to the one transcribed in elav. It must arise from a gene located 3' to the elav gene, transcribed from the opposite DNA strand, but different

Fig. 2. Protein sequence alignments between the virtual translation of cDNA-30h (D. melanogaster arginase) and representative arginase sequences from the data base. H.s., H. sapiens; Mm, M. musculus; X.l., X. laevis. The predicted N-terminal of the D. melanogaster arginase ORF includes two additional amino acids (M and W) upstream of cDNA-30h ORF (see text), shown in parentheses. Below the alignment, a consensus sequence (plurality 7) is shown, where residues present in all nine compared species are in bold type. The three black boxes frame the arginase family signatures. Residues are circled at the positions (in the D. melanogaster and C. elegans sequences) that do not match the signatures consensus. The 37 positions marked with asterisks are strictly conserved in 19 arginases from procaryotes and eucaryotes (20). Respectively 1 and 10 of these positions highlighted with small shadowed caps differ in D. melanogaster and C. elegans.

The vertical arrow separates the N-terminal region of the Drosophila arginase (encoded by sequences downstream of the elav transcription unit) from its C-terminal region (encoded by sequences upstream of the elav transcription unit).

Fig. 3. Matrix of the percentages of amino acid identity between the arginases aligned in Fig. 2. Percentages were calculated with the GAP program of GCG, as specified under "Experimental Procedures." Species names symbols are as defined in Fig. 1.

Fig. 4. Developmental Northern blot analysis of arginase expression in D. melanogaster using a cDNA-30h antisense RNA probe. About 5 μg of total RNA per lane was used. Lane 1, 0–4 h embryos; lane 2, 4–8 h embryos; lane 3, 8–12 h embryos; lane 4, 12–16 h embryos; lane 5, 16–20 h embryos; lane 6, 20–24 h embryos; lane 7, first instar larvae; lane 8, third instar larvae; lane 9, pupae; lane 10, adults; lane 11, adult heads. RP49 is a loading control. Sizes are in kb.
from the arg gene (Fig. 5). A good candidate is the silver gene, which maps upstream of the arg gene (10), whose largest transcript is estimated to be 6–7 kb (29). Thus, the proposal that the elav gene extends into the region constituting the 5′ of the arg gene is not supported by my analysis.

Arginase Transcripts Localize to the Fat Body in Embryos—Given the arrangement of the arg and elav genes, mutual regulation of the two loci could occur if transcription occurs in the same cells. This is a distinct possibility, because elav transcripts are exclusively present in all neurons (30), and arginase transcripts are found in Drosophila heads (Fig. 4). In rats, arginase activity has been reported in neuronal and glial brain cells (31). To determine whether the arginase gene is transcribed in neurons, we examined its expression in Drosophila embryos by in situ hybridization using an antisense arginase probe. Consistent with the developmental Northern, arg transcripts appear late during embryogenesis, first detected around stage 12 in dorso-lateral cells present in the most prosterior two-thirds of the embryo. As development proceeds, the labeling intensifies, spreads more anteriorly, and highlights a structure with clefts and holes in stage 16 embryos (see Fig. 6). At this stage, the pattern of arginase transcripts is similar to the pattern of expression of early fat body markers (32, 33). Note that the nervous system, where elav is specifically expressed (30), remains unlabeled by the arg antisense probe, as may be seen for instance in Fig. 6A, where the ventral chord of the nervous system is free of hybridization.

A Mutation Deleting the Region of the Arginase Gene Downstream of Its Third Intron Is Homozygous Viable—The elav4 lethal mutation is a chromosomal inversion with breakpoints near 1A6-1B1 and 1B5-9 (28). Because of the organization of the elav region (Fig. 7), it was likely that this elav allele was also an arg mutation. I reasoned that if this was the case, I could generate a fly stock specifically mutant for arg function by combining the elav4 allele with an elav transgene (350-83-1), which fully rescues elav function but which carries no sequences from the arg ORF (24). Such individuals (elav4/ + /elav4/Y; 350-83-1/350-83-1) are functionally normal for elav but mutant for arg. They develop as fertile adults. The arg transcripts produced in elav4 mutants were analyzed by Northern blot analysis, and RNA both from heterozygous flies (elav4/elav+) or from the arg mutant stock elav4/elav4/Y; 350-83-1/350-83-1 was examined. Consistent with the molecular structure of the arg mutation, truncated arg RNA is present in these flies (Fig. 8). It is predicted to generate a truncated arginase missing the 152 C-terminal amino acids (Fig. 2).

Another lethal elav allele (elav60) has been extensively used in the analysis of elav function. This allele was generated by the excision of a P element inserted 518 base pairs upstream of the elav transcription initiation site (approximately kb 2 on the scale Fig. 1), which removed the entire elav ORF (8). It seemed crucial to determine if arg expression was also affected in this elav mutant. Examination of the arg transcripts in elav60 mutants reveals that they are unaffected (Fig. 8), indicating that the P element excision does not remove sequences found in the arg mature transcripts. In addition, the alteration of arg intron 3 has no apparent effect on the processing of the transcripts. Thus the elav60 mutation, unlike elav4, is specific to elav and does not affect arg expression.

Flies Producing a C-terminally Truncated Form of Arginase Live but Develop More Slowly Than Normal—Individuals carrying the arg mutation (elav4/elav4/Y; 350-83-1/350-83-1) de-
The Drosophila Genome Contains One Arginase Gene—I identified a cDNA encoding a predicted arginase, based upon its similarity with arginases of other organisms, including the presence of conserved regions implicated in the binding of manganese ions typically found in arginases and required for their enzymatic activity. The putative arginase of *D. melanogaster* is similarly related to the terrestrial vertebrates arginases that were examined and to *S. cerevisiae* and *C. elegans* arginases, sharing 37–41% sequence identity with them. In addition, 36 of the 37 residues strictly conserved in arginases from 19 species of procaryotes and eucaryotes are conserved in the 351-amino acid-long *D. melanogaster* putative arginase. Clearly, the *D. melanogaster* cDNA clone that was identified encodes a *bona fide* arginase.

Arginase is a widely distributed enzyme in living organisms, found in bacteria, yeast, plants, invertebrates, and vertebrates. It catalyzes the hydrolysis of arginine into urea and ornithine. Most organisms that do not have a functional urea cycle (in contrast to terrestrial vertebrates) apparently produce only one arginase. The *arginase* gene duplication that gave rise to the 351-amino acid-long *D. melanogaster* putative arginase is detailed under “Experimental Procedures.” The cumulative number of adult progeny obtained from the cross was plotted as a function of the time of eclosion. Expected progeny numbers were calculated as a fraction (predicted by Mendelian rules) of the total progeny. Hence, arg− males were expected to be 3/15 of the total progeny (the *elav*/YW, *T(2;3)ap* males do not hatch), and each of the three other categories of flies were expected to represent 4/15 of the total progeny.

**DISCUSSION**

The *Drosophila* Genome Contains One Arginase Gene—I identified a cDNA encoding a predicted arginase, based upon its similarity with arginases of other organisms, including the presence of conserved regions implicated in the binding of manganese ions typically found in arginases and required for their enzymatic activity. The putative arginase of *D. melanogaster* is similarly related to the terrestrial vertebrates arginases that were examined and to *S. cerevisiae* and *C. elegans* arginases, sharing 37–41% sequence identity with them. In addition, 36 of the 37 residues strictly conserved in arginases from 19 species of procaryotes and eucaryotes are conserved in the 351-amino acid-long *D. melanogaster* putative arginase. Clearly, the *D. melanogaster* cDNA clone that was identified encodes a *bona fide* arginase.

Arginase is a widely distributed enzyme in living organisms, found in bacteria, yeast, plants, invertebrates, and vertebrates. It catalyzes the hydrolysis of arginine into urea and ornithine. Most organisms that do not have a functional urea cycle (in contrast to terrestrial vertebrates) apparently produce only one arginase. The *arginase* gene duplication that gave rise to the 351-amino acid-long *D. melanogaster* putative arginase is detailed under “Experimental Procedures.” The cumulative number of adult progeny obtained from the cross was plotted as a function of the time of eclosion. Expected progeny numbers were calculated as a fraction (predicted by Mendelian rules) of the total progeny. Hence, arg− males were expected to be 3/15 of the total progeny (the *elav*/YW, *T(2;3)ap* males do not hatch), and each of the three other categories of flies were expected to represent 4/15 of the total progeny.

**DISCUSSION**

The *Drosophila* Genome Contains One Arginase Gene—I identified a cDNA encoding a predicted arginase, based upon its similarity with arginases of other organisms, including the presence of conserved regions implicated in the binding of manganese ions typically found in arginases and required for their enzymatic activity. The putative arginase of *D. melanogaster* is similarly related to the terrestrial vertebrates arginases that were examined and to *S. cerevisiae* and *C. elegans* arginases, sharing 37–41% sequence identity with them. In addition, 36 of the 37 residues strictly conserved in arginases from 19 species of procaryotes and eucaryotes are conserved in the 351-amino acid-long *D. melanogaster* putative arginase. Clearly, the *D. melanogaster* cDNA clone that was identified encodes a *bona fide* arginase.

Arginase is a widely distributed enzyme in living organisms, found in bacteria, yeast, plants, invertebrates, and vertebrates. It catalyzes the hydrolysis of arginine into urea and ornithine. Most organisms that do not have a functional urea cycle (in contrast to terrestrial vertebrates) apparently produce only one arginase. The *arginase* gene duplication that gave rise to the 351-amino acid-long *D. melanogaster* putative arginase is detailed under “Experimental Procedures.” The cumulative number of adult progeny obtained from the cross was plotted as a function of the time of eclosion. Expected progeny numbers were calculated as a fraction (predicted by Mendelian rules) of the total progeny. Hence, arg− males were expected to be 3/15 of the total progeny (the *elav*/YW, *T(2;3)ap* males do not hatch), and each of the three other categories of flies were expected to represent 4/15 of the total progeny.

**DISCUSSION**

The *Drosophila* Genome Contains One Arginase Gene—I identified a cDNA encoding a predicted arginase, based upon its similarity with arginases of other organisms, including the presence of conserved regions implicated in the binding of manganese ions typically found in arginases and required for their enzymatic activity. The putative arginase of *D. melanogaster* is similarly related to the terrestrial vertebrates arginases that were examined and to *S. cerevisiae* and *C. elegans* arginases, sharing 37–41% sequence identity with them. In addition, 36 of the 37 residues strictly conserved in arginases from 19 species of procaryotes and eucaryotes are conserved in the 351-amino acid-long *D. melanogaster* putative arginase. Clearly, the *D. melanogaster* cDNA clone that was identified encodes a *bona fide* arginase.

Arginase is a widely distributed enzyme in living organisms, found in bacteria, yeast, plants, invertebrates, and vertebrates. It catalyzes the hydrolysis of arginine into urea and ornithine. Most organisms that do not have a functional urea cycle (in contrast to terrestrial vertebrates) apparently produce only one arginase. The *arginase* gene duplication that gave rise to the 351-amino acid-long *D. melanogaster* putative arginase is detailed under “Experimental Procedures.” The cumulative number of adult progeny obtained from the cross was plotted as a function of the time of eclosion. Expected progeny numbers were calculated as a fraction (predicted by Mendelian rules) of the total progeny. Hence, arg− males were expected to be 3/15 of the total progeny (the *elav*/YW, *T(2;3)ap* males do not hatch), and each of the three other categories of flies were expected to represent 4/15 of the total progeny.
The Third Intron of the Drosophila Arginase Gene Includes an Essential Locus—The arginase cDNA was identified during the course of screening for cDNAs corresponding to the elav gene. elav encodes an RNA binding protein present in the nucleus of all neurons throughout development and is essential for normal differentiation and maintenance of neurons (24, 27, 35). The selection of the arginase cDNA in my screen is due to the close proximity of the two genes and the use of a genomic probe that overlapped both of them. Indeed, this paper shows that the 13-kb elav locus is entirely included within the third intron of the arg gene, the two genes being transcribed from complementary DNA strands. Furthermore, the genomic sequences producing the mature elav and arg mRNAs are non-overlapping. Consistent with the new molecular data showing that elav mRNAs do not extend into a region corresponding to arg, elav minigenes corresponding to the 15.5-kb genomic fragment flanked by PstI and BamHI sites (approximate coordinates kb 1–16 on the scale Fig. 1) provide normal elav function (24). Finally, an elav CDNA whose polyadenylation site is included in the 15.5-kb genomic fragment has been identified. Together, the data demonstrate that the distance between the site of initiation of elav transcription and the 3′ splice site of arg intron 3 is 3284 base pairs and that 197 base pairs lie between the polyadenylation site of elav and the 5′ splice site of arg intron 3.

Conserved coding regions of the arg gene are split by an intron containing elav. This suggests that the elav gene, which encodes a function that appeared more recently than the arginase function during evolution, was somehow inserted into the arginase gene. This organization seems to be different in humans, because mapping of the two arginase genes and the four identified elav homologues indicates that these loci map to distinct regions of the genome (36). Although not a general feature of nested genes, conservation of the nested organization of the gene encoding a metalloproteinase inhibitor within the synapsin gene has been reported for Drosophila and humans (37). Understanding the phenomena that led to nested gene organization would be of interest in the context of genome evolution.

The Conserved Drosophila Arginase Gene Is Nonessential—A mutation of the arg gene predicted to generate a form of arginase missing the 152 C-terminal amino acids was identified and analyzed. Northern blot analysis confirms that truncated arg transcripts are produced from the mutant flies (elav/arg). Provided that the truncated RNA is actually translated, the mutant would produce a form of Drosophila arginase missing the 152 C-terminal amino acids. The determination of the crystal structure of rat hepatic arginase A-I reveals the presence of a 25-residue oligomerization motif at the C terminus (26). In humans, naturally occurring mutant A-I missing the most C-terminal 32 amino acids that include the trimerization domain provides only 0.6% of normal arginase activity (38). Thus, it seems likely that the mutant form of Drosophila arginase provides very low levels of enzymatic activity, if any. Flies carrying this mutation survive normally, indicating that the arginase gene of Drosophila does not provide a vital function. However, a significant developmental delay (16% increase) is associated with the mutation. Analysis of additional alleles of the arginase gene will be necessary to further the analysis, in particular to determine what processes are affected by the alteration of arginase activity.

Arginase is widely distributed in living organisms, indicating that it plays an important function. Accordingly, one might have expected that the arginase mutation described here would cause a severe phenotype. However, it has modest effects. Genome analysis has revealed that only 24–30% of the genes from D. melanogaster are vital (5, 39). Two models have been proposed to explain the lack of observable phenotype in an organism carrying a null allele of a gene. The traditional view suggests that the gene in question has a function in environmental conditions different from that of the laboratory, or in situations such as stress, that are normally not encountered in the laboratory. Recently, it has been proposed that this type of gene rather encodes a function that makes a small but significant contribution to the fitness of the organism, hence the “marginal benefit” hypothesis (40). Under this hypothesis, mutations would not lead to observable phenotypes but rather alter the efficiency/reliability of basic cellular processes, leading to selective maintenance of the normal allele of the gene. Experimental evidence that yeast strains carrying null mutations that lead to no apparent phenotype are counter-selected has been presented (40). Based upon my observations and consistent with the hypothesized roles for arginase (see below), it seems possible that arginase function in Drosophila is of the marginal benefit type, i.e. nonessential, but critical enough that it has been maintained during evolution. Many diverse roles have been proposed for extrahepatic arginases, such as participation in the metabolism of polyamines (small molecules implicated in many essential cellular processes), roles in neurotransmitter synthesis (glutamate and GABA), amino acid synthesis (glutamate and proline, which is a principal source of energy for insect flight metabolism), and function in nitric oxide synthesis (reviewed in Ref. 14).

Interest in arginases has increased with the demonstration of their influence on the synthesis of nitric oxide in macrophages, where arginase competes with nitric-oxide synthase, an enzyme of ancient evolutionary lineage (41) that converts arginine to nitric oxide and citrulline (reviewed in Ref. 17). Nitric oxide is a diffusible gas that serves as an intercellular effector modulating diverse aspects of mammalian physiology, such as vascular tone, neurotransmission, and immune system function, through the activation of macrophages (42). Induction of apoptosis by high levels of nitric oxide has been reported in human monocytes and in murine macrophage-like cells (43, 44).

Many aspects of the relationship between nitric oxide metabolism and arginases are under investigation. In Drosophila, nitric oxide has been shown to regulate cell proliferation versus differentiation as well as axon pathfinding (45–47). The analysis of possible genetic interactions with the nitric-oxide synthase gene from Drosophila (47) might thus prove useful for the study of potential competition between arginase and nitric-oxide synthase in cells other than vertebrate macrophages.

Acknowledgments—I thank Jean-Luc Boucher for helpful suggestions, Leonard Rabinow and Jean-Pierre Muller for critical reading of the manuscript, Kathy Matthews for providing stocks from the Bloomington Stock Center, Chad Price for help with manipulation of GCG, and Claudia Borgeon for excellent technical help.

REFERENCES
1. Mewes, H. W., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Maierl, A., Oliver, S. G., Pfeiffer, F., and Zollner, A. (1997) Nature 387, suppl. 7–65.
2. C. elegans Genome Sequencing Consortium (1999) Trends Genet. 15, 51–58.
3. C. elegans Genome Sequencing Consortium (1998) Science 282, 2012–2018.
4. Adams, M. D., Celiker, S. E., Holt, R. A., Evans, C. A., Goecayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., et al. (2000) Science 287, 2185–2195.
5. Ashburner, M., Miska, S., Roote, J., Lewis, S. E., Blazej, R., Davis, T., Doyle, C., Galle, R., George, R., Harris, N., Hartzell, G., Harvey, D., Hong, L., Houston, K., Hoskins, R., Johnson, G., Martin, C., Moshrefi, A., Palazzolo, M., Reese, M. G., Spradling, A., Tsang, G., Wan, K., Whitelaw, K., Kimmel, B., Celniker, S., and Rubin, G. M. (1999) Genetics 153, 179–219.
6. Rubin, G. M., Hong, L., Brokstein, P., Evans-Holm, M., Frise, E., Stapleton, M., and Harvey, D. A. (2000) Science 287, 2222–2224.
Drosophila Arginase

7. Blumenthal, T. (1998) Bioessays 20, 480–487
8. Yao, K.-M., Samson, M.-L., Reeves, R., and White, K. (1993) J. Neurobiol. 24, 723–739
9. Henikoff, S., Keene, M. A., Fechtel, K., and Fristrom, J. W. (1986) Cell 44, 33–42
10. FlyBase Consortium (1999) Nucleic Acid Res. 27, 85–88
11. Xu, G. F., O'Connell, P., Viskochil, D., Cawthon, R., Robertson, M., Culver, M., Dunn, D., Stevens, J., Gesteland, R., and White, R. (1990) Cell 62, 599–608
12. Levinson, B., Kenwrick, S., Lakich, D., Hammonds, G. J., and Gitschier, J. (1990) Genomics 7, 1–11
13. Cawthon, R. M., O'Connell, P., Buchberg, A. M., Viskochil, D., Weiss, R. B., Culver, M., Stevens, J., Jenkins, N. A., Copeland, N. G., and White, R. (1990) Genomics 7, 555–565
14. Jenkinson, C. P., Grody, W. W., and Cederbaum, S. D. (1996) Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 114, 107–132
15. Patterton, D., and Shi, Y. B. (1994) J. Biol. Chem. 269, 25328–25334
16. Ouzounis, C. A., and Kyrpides, N. C. (1994) J. Mol. Evol. 39, 101–104
17. Boucher, J. L., Moali, C., and Tenu, J. P. (1999) Cell Mol. Life Sci. 55, 1015–1028
18. Jauniaux, J. C., Dubois, E., Vissers, S., Crabeel, M., and Wiame, J. M. (1982) EMBO J. 1, 1125–1131
19. Green, S. M., Eisenstein, E., McPhie, P., and Hensley, P. (1990) J. Biol. Chem. 265, 1601–7
20. Perozich, J., Hempel, J., and Morris, S. M. J. (1998) Biochim. Biophys. Acta 15, 23–37
21. Itoh, N., Salvaterra, P., and Itakura, K. (1985) Drosophila Information Service 311, 425–431
22. Settle, S. H., Jr., Green, M. M., and Kurtis, K. C. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 9470–9474
23. Robinow, S., and White, K. (1991) J. Neurobiol. 22, 443–461
24. Braissant, O., Gotoh, T., Loup, M., Mori, M., and Bachmann, C. (1999) Brain Res. Mol. Brain Res. 70, 231–241
25. Zinke, I., Kirchner, C., Chao, L.C., Tetzlaff, M. T., and Pankratz, M. J. (1999) Development 126, 5275–5284
26. Riechmann, V., Rehorn, K.-P., Reuter, R., and Leptin, M. (1998) Development 125, 713–723
27. Robinow, S., and White, K. (1991) J. Neurobiol. 22, 443–461
28. Campos, A. R., Grossman, D., and White, K. (1985) J. Neurogenet. 2, 197–218
29. Maglott, D. R., Katz, R. S., Sicotte, H., and Pruitt, K. D. (2000) Nucleic Acids Res. 28, 126–128
30. Pohar, N., Godenschwege, T. A., and Buchner, E. (1999) Genomics 15, 293–296
31. Ash, D. E., Scolnick, L. R., Kyoto, Z. F., Vockley, J. G., Cederbaum, S. D., and Christianson, D. W. (1998) Mol. Genet. Metab. 64, 245–249
32. Miklow, G. L. G., and Rubin, G. M. (1996) Cell 86, 521–529
33. Green, S. M., Eisenstein, E., McPhie, P., and Hensley, P. (1990) J. Biol. Chem. 265, 1601–7
34. Robinow, S., and White, K. (1991) J. Neurobiol. 22, 443–461
35. Campos, A. R., Grossman, D., and White, K. (1985) J. Neurogenet. 2, 197–218
36. Maglott, D. R., Katz, R. S., Sicotte, H., and Pruitt, K. D. (2000) Nucleic Acids Res. 28, 126–128
37. Robinow, S., and White, K. (1991) J. Neurobiol. 22, 443–461
38. Campos, A. R., Grossman, D., and White, K. (1985) J. Neurogenet. 2, 197–218
39. Robinow, S., and White, K. (1991) J. Neurobiol. 22, 443–461
40. Robinow, S., and White, K. (1991) J. Neurobiol. 22, 443–461
41. Robinow, S., and White, K. (1991) J. Neurobiol. 22, 443–461
42. Robinow, S., and White, K. (1991) J. Neurobiol. 22, 443–461
43. Rouzaut, A., Subira, M. L., de Miguel, C., Domingo-de-Miguel, E., Gonzalez, A., Santiago, E., and Lopez-Moratalla, N. (1999) Biochim. Biophys. Acta 1451, 319–333
44. Ignarro, L., and Murad, F. (eds) (1995) Nitric Oxide: Biochemistry, Molecular Biology, and Therapeutic Implications, Academic Press, San Diego, CA
45. Rouzaut, A., Subira, M. L., de Miguel, C., Domingo-de-Miguel, E., Gonzalez, A., Santiago, E., and Lopez-Moratalla, N. (1999) Biochim. Biophys. Acta 1451, 319–333
46. Gotthard, T. and Mori, M. (1999) J. Cell Biol. 144, 427–434
47. Enikolopov, G., Banerji, J., and Kuzin, B. (1999) Cell Death Differ. 6, 956–963
48. Gibbs, S. M., and Truman, J. W. (1998) Neuron 20, 83–93
49. Kuzin, B., Roberts, I., Peunova, N., and Enikolopov, G. (1996) Cell 87, 639–649
50. Regulski, M., and Tully, T. (1995) Proc. Natl. Acad. U. S. A. 92, 9072–9076
51. Samson, M.-L., Lisbin, M. J., and White, K. (1995) Genetics 141, 1101–1111