Copyright © 1995 by the Genetics Society of America

Regulatory Autonomy and Molecular Characterization of the Drosophila *out at first* Gene

David E. Bergstrom,[1] Christopher A. Merli, Jennifer A. Cygan,[2] Richard Shelby[3] and Ronald K. Blackman

Department of Cell and Structural Biology, University of Illinois, Urbana, Illinois 61801

Manuscript received September 10, 1994
Accepted for publication December 2, 1994

ABSTRACT

Our previous work has shown that the expression of the Drosophila *decapentaplegic* (*dpp*) gene in imaginal disks is controlled by a 30 kb array of enhancers located 3′ of the *dpp* coding region. Here, we describe the cloning and characterization of *out at first* (*oaf*), a gene located near this enhancer region. Transcription of *oaf* results in three classes of alternatively polyadenylated RNAs whose expression is developmentally regulated. All *oaf* transcripts contain two adjacent open reading frames separated by a single UGA stop codon. Suppression of the UGA codon during translation, as seen previously in Drosophila, could lead to the production of different proteins from the same RNA. During oogenesis, *oaf* RNA is expressed in nurse cells of all ages and maternally contributed to the egg. During embryonic development, zygotic transcription of the gene occurs in small clusters of cells in most or all segments at the time of germband extension and subsequently in a segmentally repeated pattern in the developing central nervous system. The gene is also expressed in the embryonic, larval and adult gonads of both sexes. We also characterize an enhancer trap line with its transposon inserted within the *oaf* gene and use it to generate six recessive *oaf* mutations. All six cause death near the beginning of the first larval instar, with two characterized lines showing nervous system defects. Last, we discuss our data in light of the observation that the enhancers controlling *dpp* expression in the imaginal disks have no effect on the relatively nearby *oaf* gene.

THE *decapentaplegic* (*dpp*) gene of *Drosophila melanogaster* encodes a secreted polypeptide of the TGF-β family (PADGETT *et al.* 1987; KINGSLEY 1994) that plays several key roles in pattern formation during development. In its earliest zygotic function, the *dpp* protein acts as a morphogen in establishing the dorsal/ventral polarity of the early embryo (IRISH and GELBART 1987; FERGUSON and ANDERSON 1992; WHARTON *et al.* 1993). Later, *dpp* serves as a signaling molecule to specify cell fates within the developing midgut (IMMERGLÜCK *et al.* 1990; PANGANIBAN *et al.* 1990; REUTER *et al.* 1990). Expression of *dpp* is also required for the allocation and development of the imaginal disks, the groups of cells that proliferate during the larval stages and differentiate to form much of the adult (SPENCER *et al.* 1982; POSAKONY *et al.* 1991; COHEN 1993).

This variety of developmental requirements for *dpp* function is reflected in the complex molecular organization of the gene (ST. JOHNSTON *et al.* 1990; Figure 1). *dpp* spans 60 kb at polytene bands 22F1,2 on the second chromosome. The gene has been divided on the basis of molecular and genetic criteria into three major regions, called *shortvein* (*shv*), *Haplo-insufficient* (*Hin*), and *imaginal disk-specific* (*disk*) (ST. JOHNSTON *et al.* 1990). The *shv* and *Hin* regions contain the transcription units responsible for the *dpp* mRNAs as well as the regulatory information directing the gene's transcription in the embryonic epidermis and midgut (ST. JOHNSTON *et al.* 1990; HUANG *et al.* 1993; HURSH *et al.* 1993; MASUCCI and HOFFMANN 1993; CAPOVILLA *et al.* 1994; JACKSON and HOFFMANN 1994). Conversely, the *disk* region is not transcribed, but rather constitutes an expansive (30 kb) 3′ *cis*-regulatory region responsible for the gene's expression along the anterior/posterior compartment boundary in the imaginal disks (MASUCCI *et al.* 1990; ST. JOHNSTON et al. 1990; BLACKMAN *et al.* 1991; RAFTERY *et al.* 1991; BLAIR 1992). An ongoing analysis of the *disk* region has identified at least seven enhancers, each of which directs transcription in a subset of the overall *dpp* disk pattern (Figure 1; BLACKMAN *et al.* 1991; R. BLACKMAN and T. GILLEVET, unpublished data).

As shown in Figure 1, the regulatory information within the *dpp disk* region operates over great distances.

Corresponding author: Ronald K. Blackman, Department of Cell and Structural Biology, University of Illinois, 505 S. Goodwin Ave., Urbana, IL 61801. E-mail: ron_blackman@qms1.life.uiuc.edu

[1] *Present address:* The Jackson Laboratory, 600 Main St., Bar Harbor, ME 04609–0800.

[2] *Present address:* Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Ave., Cambridge, MA 02138.

[3] *Present address:* Scripps Research Institute, 10666 N. Torrey Pines Rd., La Jolla, CA 92037.

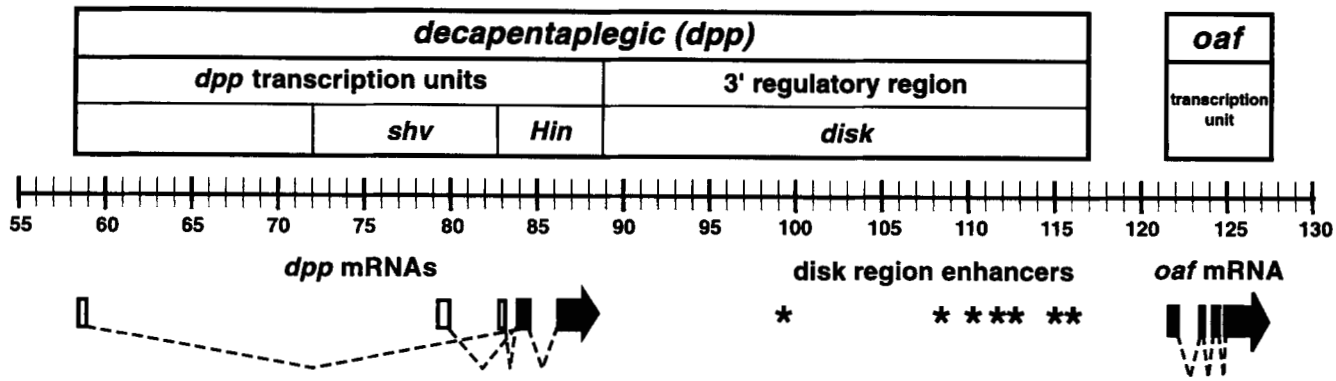

FIGURE 1.—Molecular organization of the *dpp* and *oaf* loci. Molecular map coordinates are in kilobases and are extended from the original map of ST. JOHNSTON *et al.* (1990). The *dpp* and *oaf* transcripts are shown as segmented arrows with exons (rectangles) connected by introns (– – –). Unfilled exons in *dpp* represent three alternatively used first exons (initiated at 58, 80 and 83) that are spliced onto the second and third exons common to all transcripts. These different RNAs initiate at promoters Pr-B, Pr-C and Pr-A, respectively. Other minor start sites in *dpp* are not represented. Imaginal disk-specific expression of *dpp* is regulated by enhancers (*) located within the 3′ *disk* region. The longest of the *oaf* mRNAs, with its four exons, is diagrammed.

Enhancers near the 3′ end of the region activate transcription from alternate *dpp* promoters in some cases >50 kb away. Further, as judged by *dpp*'s sensitivity to transvection, the *disk* region enhancers are even able to direct transcription from a *dpp* gene on another chromosome (its paired homolog) (GELBART and WU 1982; BENSON and PIRROTTA 1988; WU and GOLDBERG 1989). Given that the *dpp* enhancers can operate over such vast distances, we became interested in the mechanisms involved in constraining the action of these regulatory elements to their cognate promoters.

An important first step in this analysis is to determine whether the gene adjacent to the *disk* region is influenced by these enhancers. Although no such gene has been described molecularly, previous genetic studies have identified three lethal complementation groups [*l(2)ND1, l(2)ND2* and *l(2)ND3*] positioned just proximal (to the right in Figure 1) of *dpp* (SPENCER *et al.* 1982; LITTLETON and BELLEN 1994). Of these, *l(2)ND1* maps closest to *dpp* (SPENCER *et al.* 1982; R. BLACKMAN and W. GELBART, unpublished results). Because other nonessential genes could lie even closer, we decided to take a molecular approach to identify *dpp*'s nearest neighbor. Here, we report the cloning, molecular characterization, and expression patterns of a gene, *out at first (oaf)*, lying just 6 kb from the nearest *disk* region enhancer. This gene appears unresponsive to the influence of the *dpp* enhancers during disk development, suggesting that mechanisms are in place to maintain the regulatory independence of these adjacent genes.

MATERIALS AND METHODS

Fly stocks: Unless otherwise noted, all stocks used in this study are described in LINDSLEY and ZIMM (1992). The E-32 enhancer trap line was first described by BROOK *et al.* (1993). This line contains the *P*-element enhancer trap *P*[*lacZ*, *rosy*$^+$]PZ (JACOBS *et al.* 1989) within the 5′ untranslated region of *oaf* and is homozygous viable.

Flies were reared at 25° on a standard medium of yeast extract, sucrose, cornmeal and agar seeded with live Baker's yeast.

Generation and analysis of *oaf* mutant alleles: We induced *oaf* recessive lethal mutations by remobilizing the E-32 transposon (containing the *rosy*$^+$ gene) and scoring for the loss of rosy$^+$ activity. Some of these events were imprecise excisions leading to the loss of *oaf* sequences. To induce the mobilization of the transposon, we crossed flies containing the *P*[*ry*$^+\Delta 2$–3](99B) chromosome, which serves as a stable source of *P* transposase (ROBERTSON *et al.* 1988), to our *E-32/ CyO;ry* stock. One hundred *E-32/ CyO;ry/ SbΔ2–3* male progeny were then individually mated to $dpp^{d\text{-}blk}$ *cn;ry* females. A single rosy$^-$ fly of the genotype *E-32*/* $dpp^{d\text{-}blk}$ *cn;ry* (where *E-32** denotes the remobilized transposon chromosome) was selected from 70 of the crosses. None of these contained deletions that extended far enough leftward to uncover the $dpp^{d\text{-}blk}$ deficiency [106 to 111.4 on the molecular map (ST. JOHNSTON *et al.* 1990; Figure 1)] and produce the small eye phenotype associated with this *dpp* mutation. The selected flies were crossed to establish balanced stocks of the remobilized chromosomes over *CyO*. Seven of the lines failed to homozygose, suggesting that they carried newly induced lethal *oaf* mutations. These lines were then complementation tested with deficiency stocks encompassing the *oaf* region [*Df(2L)*dpp^{d14} and *Df(2L)*dpp^{d19}] and an EMS-induced allele of *l(2)ND1*. Lines which did not complement these mutations were outcrossed to *y w* flies to observe *oaf* phenotypes in the absence of the lethal *Cy* mutation. F_1 flies lacking the balancer were crossed *inter se* and their progeny scored for survival at 25°.

cDNA isolation: *oaf* cDNAs were recovered from plasmid libraries prepared from 0–4, 4–8, 8–12 and 12–24 hr embryonic RNA (BROWN and KAFATOS 1988). Approximately 300,000 colonies from each library were transferred to nitrocellulose and screened in duplicate with ^{32}P-labeled probes from the *oaf* region and vicinity. The filters were hybridized overnight at 65° in 5× SSCP, 5× Denhardt's solution, 300 μg/ml denatured salmon sperm DNA, 0.2% SDS and 10% dextran sulfate. The filters were washed for 1 hr at 65° in 2×

TABLE 1

cDNA clones recovered

Clone	Library	5′ extent	3′ extent	A_n[a]
Class 1 cDNAs				
NN	12–24 hr	+1	+3171	12
D	4–8 hr	+6	+3171	>12
QQ	12–24 hr	+841	+3171	>12
MM	12–24 hr	+1043	+3171	>12
Class 2 cDNAs				
25–3	0–4 hr	+466	+2586	>12
13–1	0–4 hr	+1370	+2586	>12
22	0–4 hr	+1608	+2586	>12
Class 3 cDNAs				
2	0–4 hr	+868	+1734	>12
Internally primed cDNAs				
A	4–8 hr	+1	+1590	12
G	4–8 hr	+1	+1590	12
OO	12–24 hr	+1	+1590	12
Y	8–12 hr	+289	+1590	12
21	0–4 hr	+466	+1590	12

The 5′ and 3′ extent of each clone is given using the nucleotide sequence of clone NN as the reference. The cDNA class designation is discussed in the text.

[a] A_n = Number of adenine residues at 3′ end of cDNA insert.

SSCP, 0.1% SDS, and exposed to XAR-5 film overnight at −70° with Lightning Plus intensifying screens. Approximately 100 positives were found and 13 of these were purified and characterized in detail.

Sequence analysis and plasmid clones: The 5′ and 3′ ends of the cDNAs were sequenced using primers which read in from the cloning vector's ends (Table 1). One of the cDNAs, clone NN, was selected for sequencing because it was likely to be full length. The *oaf* sequence was subcloned into pBluescript II KS and nested deletions were created from both ends of the cDNA using the Erase-a-Base System (Promega). Double-stranded DNA, purified by Magic Miniprep columns (Promega), or single-stranded DNA (SAMBROOK *et al.* 1989) was sequenced using Sequenase 2.0 (U.S. Biochemical Corp.) according to the manufacturer's protocol. Reaction products were resolved on 6% acrylamide, 7 M urea, 1× TBE or TTE gels, dried, and exposed to Kodak XAR-5 film. Sequences were aligned using AssemblyLIGN software and analyzed with MacVector and DNA Strider sequence analysis software. Searches of the NCBI databases were done using the BLAST program (ALTSCHUL *et al.* 1990).

While analyzing the 13 cDNAs, we found that one class of five clones (A, G, OO, Y, and 21) appeared to be artifactually terminated at its 3′ end. Each of the five ended at nucleotide 1590 and are followed by 12 adenines (Table 1). This is the same as the number of thymidines used in the primer to initiate cDNA synthesis during the original library construction (BROWN and KAFATOS 1988). Clones ending at true *in vivo* polyadenylation sites typically have A tracts longer than 12 nt, as evidenced by seven of the other cDNAs we analyzed (Table 1). Because the region from 1591 to 1600 in clone NN is adenine-rich, the oligo-dT primer is likely to have annealed here inappropriately and internally primed the cDNA synthesis.

Three of the cDNAs (NN, G, and OO) have nontemplated guanosines at the 5′ end of the cDNA insert (D. BERGSTROM, unpublished results). These are commonly found at the 5′ ends of cDNA inserts that have extended to the true initiation site (THUMMEL 1993). Each of the three clones extends to the identical nucleotide, which is the 5′ most present in any clone.

Genomic DNA sequence was generated from clone p119B, containing an 8.2-kb *Bam*HI fragment (map coordinates 119.2–127.4), and its subclone p119BS, with a 2.3-kb *Bam*HI/*Sal*I fragment (119.2–121.5). Both are inserted in pBluescript II KS. The Drosophila DNA is from a phage containing cloned Oregon-R genomic sequences. Sequencing was initiated from primers in the vector or the cDNA sequence.

For the *in situ* hybridizations and Northern blots, we used clone pSS*oaf* for the probe. This contains a *Sal*I-*Spe*I fragment (nt 15 to 3089) from cDNA clone NN inserted into the respective sites of pBluescript II KS.

RNA isolation and analysis: Drosophila embryos (Oregon R) were collected and aged at 25° to obtain staged embryos of 0–4, 4–8, 8–12 and 12–24 hr of development. For first, second and early third instar larvae, 0–12 hr embryos were added to a tub containing standard food and aged for 30, 54 and 78 hr, respectively. Wandering third instar larvae and pupae aged 24–48 hr or 96–120 hr after pupariation were collected from fly food bottles. Adult flies were anesthetized under CO_2 and separated by sex. Samples were frozen at −80° until needed.

To prepare RNA, the samples were lysed in 7 M urea, 2% SDS, 0.35 M NaCl, 10 mM Tris and 1 mM EDTA, pH 8, in a Dounce homogenizer and extracted with phenol/chloroform (50:50) three times and chloroform once. The nucleic acid was ethanol precipitated and centrifuged. To further purify these RNAs, the pellets were dissolved in water and then brought to 3 M sodium acetate, pH 5, before precipitation overnight at −20°. The pelleted RNAs were dissolved in 0.3 M sodium acetate, pH 5, and stored as ethanol precipitates at −20°.

For Northern analysis, RNAs were glyoxalated and electrophoresed on 1.2% agarose gels containing 0.5 μg/ml ethidium bromide in 10 mM sodium phosphate buffer (pH 7.0). After electrophoresis, the gels were photographed with UV light, equilibrated for 30 min in 10 mM NaOH and blotted with the same solution for 6 hr onto Zetaprobe GT nylon membranes. Blots were air dried, prehybridized for 2–3 hr in the hybridization solution of CHURCH and GILBERT (1984) as modified by PARIS *et al.* (1993) and hybridized for 18 hr in fresh solution containing random-primed ^{32}P-labeled *oaf* region probes. After hybridization, blots were washed once for 15 min at 65° in 4% SDS, 40 mM $NaPO_4$, 1 mM EDTA and three times for 15 min at 65° in 1% SDS, 40 mM $NaPO_4$, 1 mM EDTA. Blots were exposed to XAR-5 film at −70° with Lightning Plus intensifying screens.

Primer extension: Ten picomols of a 30 nucleotide primer near the 5′ end of *oaf* (*oaf* 103R, 5′-GGTGTGGGTGCTCCTCCTTTAAGATCATTG-3′) were end labeled with γ-^{32}P-ATP^{32}P-ATP (30 μCi, 3000 Ci/mmol) and T4 polynucleotide kinase (8–10 U) for 10 min at 37° in a 10-μl reaction containing 1× Forward Exchange Buffer (50 mM Tris, pH 7.5, 10 mM $MgCl_2$, 0.5 mM DTT and 0.1 mM spermidine). After the inactivation of the kinase for 2 min at 90°, unincorporated isotope was removed by Sephadex G-25 chromatography. Labeled primer was hybridized for 12 hr at room temperature to 100 μg of total RNA in 30 μl of hybridization solution (80% formamide, 0.4 M NaCl, 40 mM PIPES and 1 mM EDTA, pH 6.4). After hybridization, the mixture was ethanol precipitated, resuspended, and treated with AMV reverse tran-

scriptase (1 U) for 30 min at 42° in a 20 μl reaction containing 1× Primer Extension Buffer [50 mM Tris (pH 8.3 at 42°), 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, 2.8 mM $Na_4P_2O_7$, 1 mM each dNTP and 0.5 mM spermidine]. After the addition of EDTA to 25 mM, the samples were treated with 5 μg RNase A for 30 min at 37°, phenol/chloroform extracted, ethanol precipitated and resuspended in loading dyes. The extension products were resolved on 6% acrylamide, 7 M urea and 1× TBE denaturing polyacrylamide gels adjacent to the sequencing reactions of a genomic DNA clone spanning the start of *oaf* transcription. This DNA was prepared using the TA-Quence Cycle Sequencing Kit (U.S. Biochemicals) and the same labeled *oaf*103R oligonucleotide, such that the primer extension product band would lie adjacent to the band from the sequencing reaction that represents the 5′ end of the *oaf* transcript.

Whole-mount *in situ* hybridization: Antisense RNA probes were labeled with digoxigenin-11-UTP according to the manufacturer's protocol (Boehringer Mannheim). These probes were then processed by the method of JIANG *et al.* (1991) except that the probe was hydrolyzed for only 20 min. Dissociated ovaries and embryos were fixed, hybridized and treated as described previously (TAUTZ and PFEIFLE 1989; JIANG *et al.* 1991). Imaginal disks and male gonads were collected and fixed as described (MASUCCI *et al.* 1990) except we omitted glutaraldehyde from the final fixation step and our PBT contained 0.1% Tween-20 in place of the Triton. Subsequent manipulations with these tissues were performed as above. Tissues were mounted in 70% glycerol/30% PBS and photographed using Nomarski optics.

Histochemical staining for β-galactosidase: Embryos were stained for β-galactosidase activity according to the method of BELLEN *et al.* (1989) except the embryos were devitellinized before mounting. After staining in X-gal solution, the embryos were washed with PBS in a 1.5 ml tube. Devitellinization was accomplished by adding 0.5 ml heptane to 0.5 ml of PBS/embryos, removing the PBS, adding 0.5 ml methanol and agitating the tube to fracture the vitelline membranes. After removing both the methanol and the heptane, the embryos were washed twice in methanol, twice in PBS and equilibrated in 70% glycerol/30% PBS.

Larval gonads and brains and adult testes and ovaries were stained using the same conditions as described previously for imaginal disks (BLACKMAN *et al.* 1991).

E-32 insertion site analysis: To determine the insertion site of the E-32 transposon, we employed plasmid rescue to recover the left half of the transposon (containing the *lacZ* gene) and the adjoining genomic sequences. Using the method of PIRROTTA (1986), we digested genomic DNA from *E-32;ry* flies with *Xba*I and *Nhe*I and self-circularized the DNA by ligation. These enzymes produce overhanging ends that are complementary. The DNA was electroporated into JM109 cells and the bacteria were plated on LB + kanamycin plates. Three recovered clones extended from an *Xba*I site in the middle of the PZ transposon (JACOBS *et al.* 1989) to an *Nhe*I site in the *dpp* disk region at map coordinate 112.3. Clone pPR2 was sequenced using a primer from within the *P* element's left end reading outward into the genomic DNA.

The insertion junction at the other end of the transposon was PCR amplified using a primer in the *P* element's right end and an *oaf* primer. For these experiments, genomic DNA was prepared from a single fly according to the method of GLOOR *et al.* (1993) except we omitted the proteinase K step. PCR was performed with 2 μl DNA in a 100 μl reaction containing PCR buffer III (PONCE and MICOL 1992), 200 μM dNTP, 0.4 μM of each amplification primer and 2 U Taq DNA polymerase (Boehringer Mannheim). Samples were subjected to a thermocycling scheme of 2 min at 95°; 40 cycles of 1 min at 95°, 1 min at 55°, and 1.5 min at 75°; and 10 min at 75°. Amplified products were run on gels, purified using Magic PCR Preps (Promega), cloned into ddT-tailed plasmid vectors (HOLTON and GRAHAM 1990) and sequenced.

RESULTS

Isolation and characterization of *oaf* cDNAs: Using probes that spanned the 113 to 133.5 region, we screened 0–4, 4–8, 8–12 and 12–24 hr embryonic cDNA libraries. Among the positives from the probes covering the 113 to 125 region, we purified and characterized 13 clones (Table 1). All correspond to the same gene which is present in a single copy in the genome (D. BERGSTROM, unpublished results). Another set of cDNAs, isolated using a more proximal probe (map units 129–133.5), derive from a transcribed middle repetitive element and were not analyzed further.

Based on restriction digest and sequence analysis, we have found that the 13 single-copy clones fall into four classes based on their 3′ ends (Table 1 and Figure 2). The differing 3′ extents of Classes 1–3 are defined by alternate sites of polyadenylation (see below). The remaining class, comprised of five cDNAs, appears to have resulted from internal priming at an adenine-rich stretch within the transcript (see MATERIALS AND METHODS) but these clones still yield valuable 5′ end information. We have named the gene encoding these cDNAs *out at first* (*oaf*) based on the mutant phenotype (death during the first larval instar, see below) of animals bearing lethal alleles of this gene.

Sequence analysis: We chose clone NN for further analysis as we expected that it would be full length and representative of the longest class of *oaf* transcript. We determined the 3171 bp sequence from both strands of the cDNA (Figure 3). Open reading frame (ORF) analysis of the sequence identifies three ORFs of greater than 400 nt (ORFs 1–3; Figure 4). Searches of the NCBI databases with the sequence of the entire cDNA or the polypetides predicted by the three ORFs have uncovered no significant matches. Thus, *oaf* appears to represent a novel protein.

ORF 1 extends from the 5′ most AUG of the sequence (nt 76 to 78) to a UGA codon at 1072 to 1074. Conceptual translation of this ORF predicts a protein of 332 amino acids (aa), a molecular weight of 37 kD and a pI of 9.1. An unusual feature of ORF 1 is that it contains a high percentage of cysteines (15 residues, 5% of the ORF), of which 10 are located within its last 78 aa. To determine whether this ORF might be of functional importance, we have compared its sequence to a partial sequence of the *oaf* gene from the distantly related species *Drosophila virilis* (R. SHELBY, unpublished results). Over a span of 252 amino acids, there

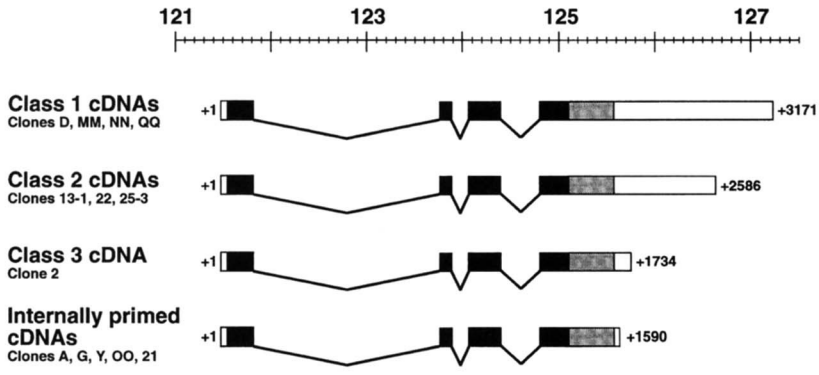

FIGURE 2.—Classes of recovered *oaf* cDNAs. The three classes of *oaf* cDNAs, distinguished by alternate sites of polyadenylation, are diagrammed beneath the molecular map (coordinates in kilobases). A fourth class, whose synthesis was initiated at an internal adenine-rich region (see MATERIALS AND METHODS), is also shown. mRNA exons are shown as rectangles. Their positions were determined by high-resolution restriction mapping and sequencing of the appropriate genomic and cDNA clones. For each cDNA, ORF 1 is filled with black, ORF 2 with gray and the remaining sequences with white. Transcription goes from left to right.

is 94.5% amino acid identity between the two species. This confirms the reading frame presented in Figure 3 and strongly suggests that ORF 1 encodes a functional protein.

ORF 2 follows immediately after the UGA at the end of ORF 1 and continues for 462 nt in the same reading frame. Because it was surprising that two substantial ORFs should be separated by a single stop codon, we sequenced this region from four more cDNAs (A, D, G and OO) and also from the *melanogaster* and *virilis* genomes. In each case, the UGA was present.

Suppression of UGA codons within protein coding regions has been found in nearly every organism studied (HATFIELD and DIAMOND 1993), including *Drosophila* (XUE and COOLEY 1993). Conceptual translation of the ORF 1 + 2 sequence predicts a 487 aa (54 kD) protein and a pI of 9.6 (assuming the incorporation of a selenocysteine at the UGA codon; see DISCUSSION). The polypeptide derived from ORF 2, however, would have some unusual characteristics. Overall, the ORF is highly enriched for serine (18%), threonine (12%) and glutamine (10%), and most of these are found in three domains comprised of only a few types of amino acids (Figure 3). The first domain (aa 369–396) contains 85% alanine, glycine and threonine, whereas the second (aa 420–447) has 82% serine, histidine, and glutamine. The third region, aa 456–480, contains only serine, threonine and proline. Thus, over half (81 of 154 aa) of ORF 2 is comprised of these simple sequence domains. Unlike ORF 1, ORF 2 is terminated by three consecutive stop codons (TGA TAA TGA).

Hydropathy analysis of the ORF 1 and ORF 2 sequences (data not shown) does not predict a prototypical signal sequence or a transmembrane domain (JAHNIG 1990). In fact, the only significantly hydrophobic region found in the two ORFs is the 16 aa stretch from aa 62 to 78. There are two moderately hydrophilic domains: aa 150–170, present in ORF 1, and aa 420–480, within the clusters of simple amino acid repeats in ORF 2.

ORF 3, from nt 1205 to 1711, presumably could be used only if alternative splicing would connect it in frame to ORF1. Because we have no evidence for such a splicing event (J. CYGAN, unpublished results), this ORF is unlikely to contribute to *oaf* protein synthesis.

A striking feature of the clone NN sequence is that its 3′ half, punctuated by numerous stop codons, represents an unusually long 3′ untranslated region (Figure 4). In this clone, polyadenylation occurs 1.6 kb beyond the end of ORF 2 and 2.1 kb beyond the UGA of ORF 1. A near-consensus polyadenylation signal (UAUAAA) is 20 nt upstream of the end of the clone. As noted above, Class 2 and 3 cDNAs use alternative sites of polyadenylation (nt 2586 and nt 1734, respectively). These sites have the prototypical polyadenylation signal (AAUAAA) ~20 nt upstream of their cleavage sites.

Intron/exon organization: From a comparison of the clone NN and *virilis* genomic DNA sequences, we could predict that the ORF 1 sequence contains three introns. Each of these intron/exon boundaries, noted on Figures 2 and 3, was verified in the *melanogaster* sequence by partial sequencing of the genomic clone p119B. High-resolution restriction mapping of the *melanogaster* genomic and cDNA clones failed to uncover any additional introns in the gene.

RNA blot analysis: We used a derivative of clone NN (pSS*oaf*) to probe a blot of developmentally staged total RNAs (Figure 5A). Three distinct groups of transcripts, distinguished by their sizes and expression profiles, are resolved. Group 1 is composed of a 3.3 kb transcript whose expression declines during embryogenesis and larval development reaching a minimum during the third instar. Transcripts levels rise again in pupae and adults. The size of this RNA is in close agreement with the sequenced Class 1 cDNA clone and presumably the two are synonymous. Group 2 is composed of a broad range of transcripts varying in size from 2.4 to 3.0 kb. These RNAs are abundant in 0–4 hr embryos and adult females, suggesting that they are maternally contributed to the embryo. In 4–8 hr embryos, Group 2 transcripts are rarer (most significantly in the 2.4–2.7 kb range) and by 8–12 hr, the transcripts are undetectable. This group includes the Class 2 cDNAs (which were only recovered from the 0–

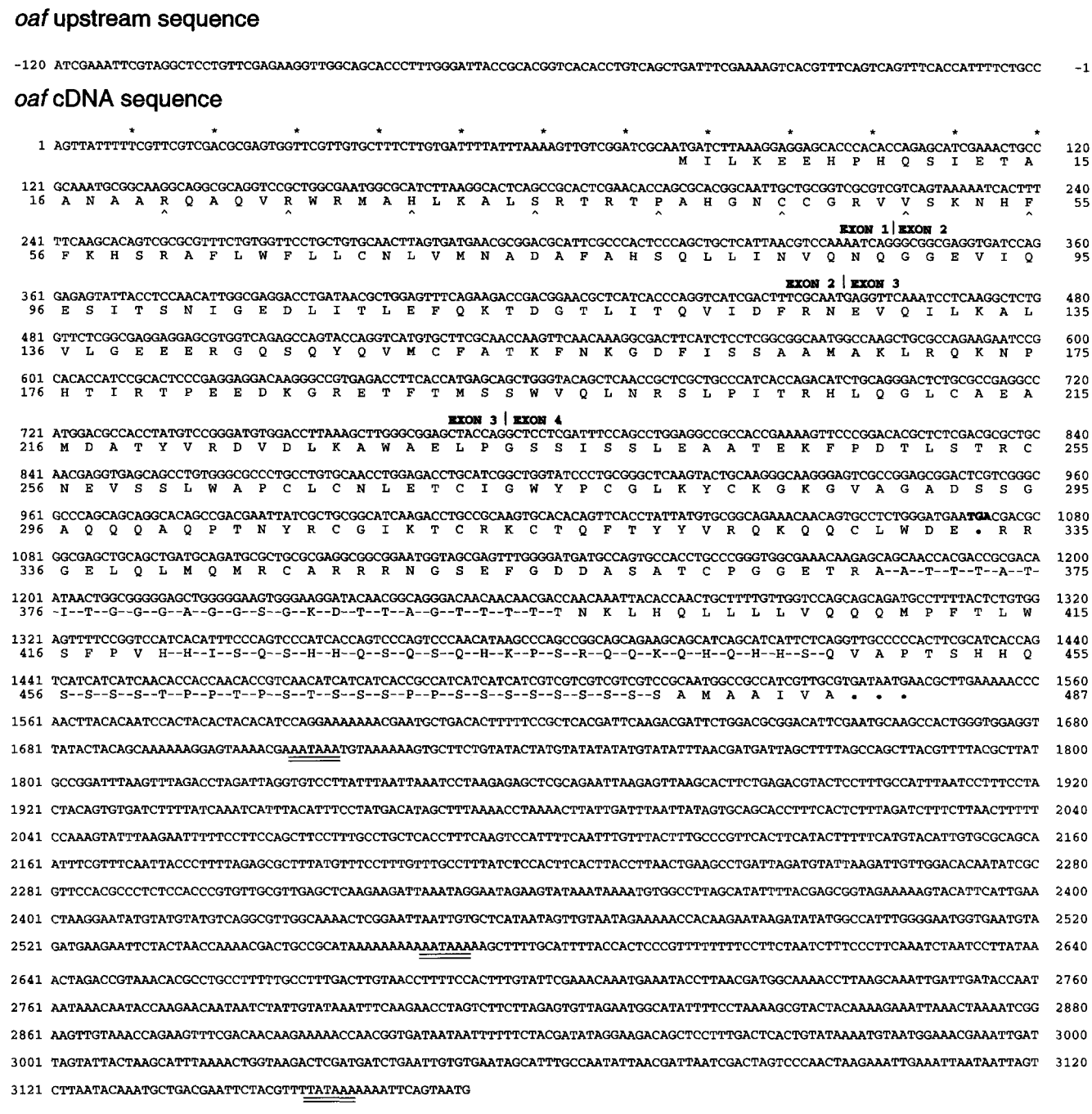

FIGURE 3.—Nucleotide sequence and conceptual translation products derived from the *oaf* locus. The sequence of Class 1 cDNA clone NN (GenBank accession number L31349) is shown. The predicted amino acid sequences of ORF 1 and ORF 2 are indicated below the cDNA sequence in single letter code. Three ORF 2 regions, all highly enriched for a few amino acids, are noted by hyphens connecting their amino acid designations. The TGA codon separating the two ORFs at nt 1072 to 1074 is in bold. This stop codon and the three at the end of ORF 2 are indicated by a dot (●) in the protein sequence. The polyadenylation signals used by the three different classes of cDNAs are underlined. Exon/intron boundaries are noted by vertical lines. The genomic DNA sequence from the 5′ flanking region of *oaf* (GenBank accession number L36089) is from plasmid p119B.

4 hr library) but the broad sizing of the RNAs suggests that there may be additional types of transcripts also. Group 3 RNAs are represented by two bands on the autoradiogram. The smaller of the two migrates just ahead of the two major rRNA bands whereas the larger runs between them. When polyA$^+$ RNA from 0–4 hr embryos is probed, this group runs as a single broad band at 1.8–2.0 kb (D. BERGSTROM, unpublished results). The Group 3 RNAs are abundant in 0–8 hr embryos but become rare in older embryos and larvae.

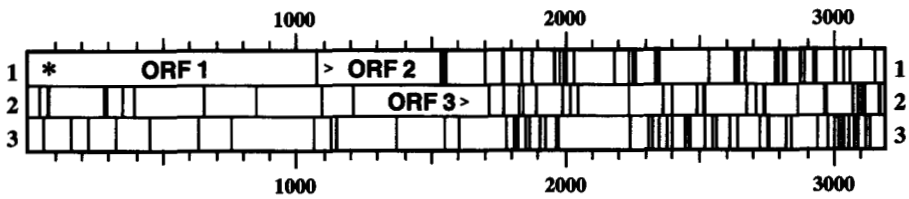

FIGURE 4.—Open reading frame (ORF) analysis. The three phase ORF map of *oaf* clone NN is shown. ORF 1, ORF 2 and ORF 3 are labeled (see text). The 5′ most AUG of the sequence is designated *. The 5′ most in-frame AUGs of ORF 2 and ORF 3 are designated >.

Transcripts levels increase in late third instar larvae, pupae, and adults. This transcript group presumably includes the Class 3 cDNAs.

Hybridization of a similar blot with a probe containing only the 5′ most 386 bp of clone NN gives the identical banding pattern (Figure 5B). In conjunction with the sequence analysis of our *oaf* cDNAs, these results suggest that all *oaf* transcripts begin at the same site. Thus, the size variation of the different transcript groups arises from the use of alternate sites of polyadenylation. The reason for the heterogeneity of sizes within Groups 2 and 3 is still unresolved.

***oaf* transcription starts from a TATA-less promoter:** To map the transcriptional start site of *oaf*, we performed primer extension experiments with total RNA from 12–24 hr embryos and late third instar larvae (Figure 6). Using an oligonucleotide primer homologous to nucleotides 74–103 of clone NN, we see a predominant extension product (ending at an adenosine) that corresponds exactly with the 5′ end of the cDNA clone, confirming that NN is full length. We observe much less extension product with the larval RNA than with the embryonic sample, consistent with the levels of *oaf* transcripts at these stages.

Genomic sequence immediately upstream of the start site of *oaf* transcription contains no TATA box consensus sequence at or near −25 indicating that *oaf* is transcribed from a TATA-less promoter (Figure 3). In such promoters, initiator (*Inr*) elements spanning the transcriptional start site play an important role in the initiation of transcription (SMALE and BALTIMORE 1989). A near match (CCAGT) of the Drosophila initiator consensus sequence (TCAGT; CHERBAS and CHERBAS 1993) is present at nt −2 to 3.

***oaf* RNA localization:** In precellularized embryos (Stage 3 and younger), we see an uniform distribution of *oaf* RNA (Figure 7A). Because the youngest of these embryos have not begun to accumulate zygotic transcripts, this RNA must be of maternal origin. During cellularization, gastrulation and germband extension (Stages 4–10), accumulated *oaf* transcripts remain uniformly distributed throughout the cellular portion of the embryo but not in the central yolk portion. At maximum germband extension (Stage 11), localized accumulations of higher levels of *oaf* RNA appear in most or all segments (Figure 7D). In the gnathal segments, the RNA is in clusters of two to four cells located adjacent to the anterodorsal margin of the maxillary protuberance and along the dorsal margin of the labial protuberance. The high level of ubiquitous maternal products makes it difficult to assess if localized expression occurs in the labial segment. Transcription in each of the three thoracic segments occurs in clusters of cells located more ventrally than the gnathal clusters and positioned near the posterior margin of the segment. Transcription in the abdomen is found in single cells located near the posterior margin of each segment. In some favorable preparations, we have also observed a single *oaf* expressing cell in each thoracic and abdominal segment at the upper edge of the germband bordering the amnioserosa. Because the ubiquitous RNA has remained at the same level as before, these localized accumulations must result from newly made zygotic transcripts.

These localized patterns disappear during germband retraction. In addition, the high level of ubiquitous *oaf* RNA declines to a lower amount that remains constant throughout embryonic development (Figure 7, E and F). This result is consistent with the data of Figure 5, which shows much less *oaf* RNA at 4–8 hr than 0–4 hr, particularly in the maternally contributed Group 2 RNAs. Thus, this reduction probably represents the degradation of most maternal *oaf* RNA.

Localized expression of *oaf* reappears at Stage 15 (Figure 7, E and F) in a segmentally repeated pattern in clusters of three to four cells along the midline of the nerve cord and bilaterally in smaller groups of cells near the periphery of the nerve cord. *oaf* expression also appears in the brain and gonad.

In wandering third instar larvae, we detect abundant *oaf* transcription in the male gonads (Figure 7G). Here, *oaf* expression is highest in the spermatocytes in the terminal end of the gonad but very low or absent in the proliferative gonial cells at the apical end. Although we have not attempted *in situ* hybridization to female larval gonads, an enhancer trap line with its transposon inserted in the *oaf* gene (described below) shows abundant expression in this tissue.

In the adult, we have looked at *oaf* RNA accumulation in female ovaries and found the gene to be expressed within cells budding from the germarium and in the nurse cells at all stages of development (Figure 7H). We also see *oaf* RNA accumulating in the developing oocytes, confirming the maternal contribution of *oaf*.

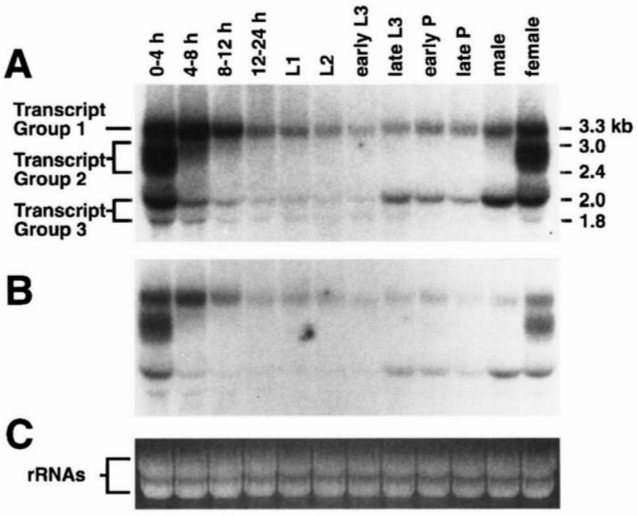

FIGURE 5.—Developmental Northern blots. (A) Staged total RNAs (7.5 μg/lane) were blotted and probed with nucleotides 15 to 3089 of clone NN. The first four lanes are embryonic samples, in hours (h) after egg laying. The remaining lanes, in order, are from first, second, early third and wandering third instar larvae, early and late pupae and male and female adults. Three distinct groups of transcripts are noted. (B) A second blot, prepared at the same time as the blot in (A), was probed with a purified fragment containing nucleotides 15 to 386 of clone NN. (C) Photograph of ethidium bromide stained gel used in (B) before blotting. Ribosomal RNAs (rRNA) serve as a control for uniform loading.

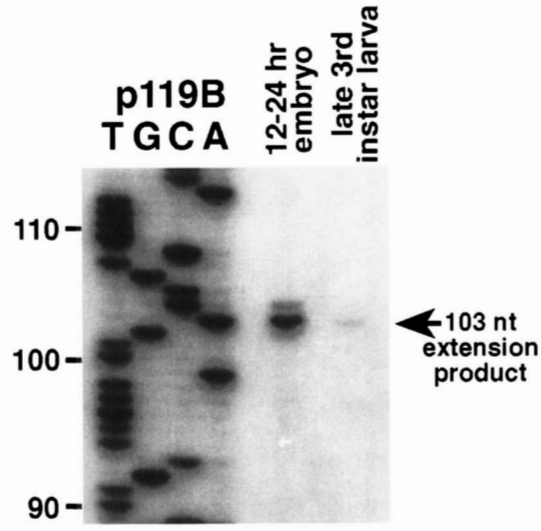

FIGURE 6.—Mapping the *oaf* transcriptional start site. Primer extension analysis using an *oaf* primer yields a 103-nt extension product from both embryonic and larval RNAs and confirms that the clone NN sequence (Figure 3) begins at the true transcriptional start site. Plasmid p119B, a genomic DNA clone spanning the *oaf* promoter, was sequenced with the same primer and run next to the extension products.

In males, the enhancer trap line expression suggests that *oaf* RNA may be present in the testes as well.

Because of our laboratory's interest in imaginal disk development, we have tried to detect the expression of *oaf* in the disks of late third instar larvae. Despite repeated attempts, we have never seen localized or nonlocalized expression of *oaf* in these tissues (D. BERGSTROM, C. MERLI and R. BLACKMAN, unpublished results).

The E-32 enhancer trap insertion in *oaf*: BROOK *et al.* (1993) have previously characterized the E-32 enhancer trap line whose insert is near *dpp*. We undertook a molecular characterization of this line and cloned the genomic DNA flanking the ends of the transposon. Sequence analysis shows that the E-32 transposon is inserted within the 5′ untranslated region of *oaf* following nt 40 of the cDNA sequence (Figure 8). The transposon is oriented such that the 5′ end of the *lacZ* gene is nearer *dpp*. The 8 bp *oaf* sequence from nt 33 to 40 is present at both ends of the transposon, a result of the typical target site duplication associated with *P*-element insertion (ENGELS 1989).

All *oaf* sequences are present in the E-32 line and the protein coding regions remain intact. Flies bearing the *E-32* chromosome are viable and fertile as homozygotes or when heterozygous for a deficiency of the entire *oaf* gene. Transcription of *oaf* sequences in E-32 is likely to initiate from a cryptic promoter within the 3′ end of the transposon (see below). This RNA would contain the same protein coding information as normal *oaf* transcripts.

E-32 staining patterns: Patterns of β-galactosidase activity (β-gal) in the E-32 line are shown in Figure 9. In contrast to the *in situ* hybridization results with *oaf* probes, we have not observed β-gal in E-32 embryos prior to or during cellularization of the blastoderm (Stages 1–5) (Figure 9A). One plausible explanation for the lack of β-gal at these early stages is that the maternally contributed *lacZ* mRNA may be degraded or nonfunctional in the embryo because of inappropriate targeting, processing or packaging. This result also suggests that zygotic transcription of *oaf* does not normally begin this early or, if it does, that it is controlled by an enhancer to which the enhancer trap is unresponsive.

We first observe β-gal in E-32 embryos during gastrulation (Stage 6). At this point, β-gal is observed as uniform staining distributed throughout the embryo (Figure 9B). This pattern is maintained throughout gastrulation and germ band extension (Stages 6–10). At Stage 11, β-gal is enhanced in small clusters of cells in the maxillary and labial protuberances, reproducing the pattern of localized expression of the *oaf* RNA in these gnathal tissues. During Stages 12–13, localized β-gal becomes evident in the thoracic and abdominal pattern usually seen for *oaf* RNA during Stage 11 (Figure 9C). The reason for this delay is unknown. A thin stripe of cells bordering the amnioserosa also expresses β-gal.

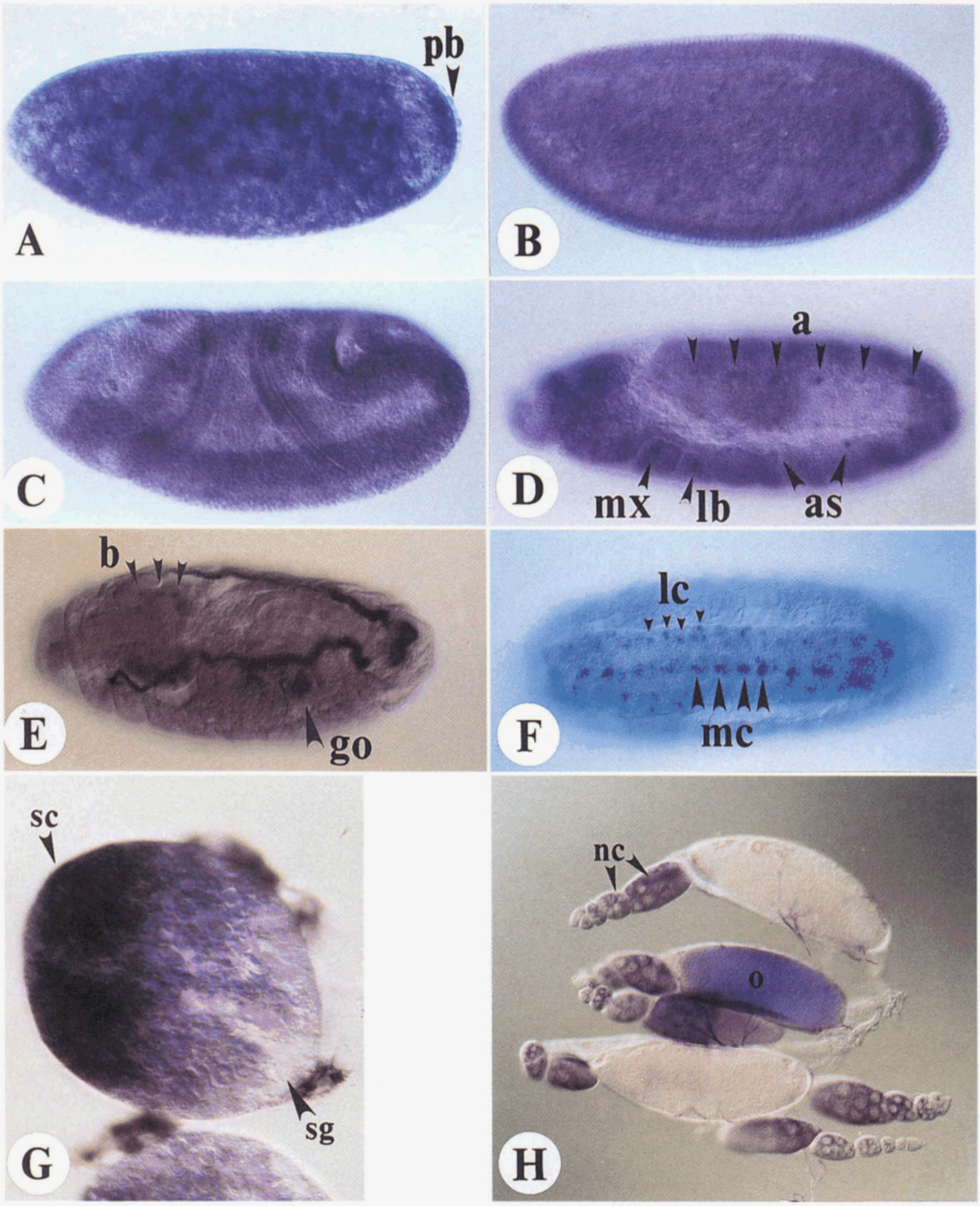

FIGURE 7.—RNA expression patterns of *oaf*. (A) Stage 3 embryo with maternally contributed *oaf* RNA throughout the embryo and the polar buds (pb). (B) Stage 5 embryo showing *oaf* RNA present throughout the blastoderm but basally localized within blastodermal cells. (C) Stage 6 embryo. (D) Stage 10 embryo with localized accumulations of *oaf* RNA in the maxillary (mx) and labial (lb) protuberances, in single cells bordering the amnioserosa (as), and in single cells in the abdominal segments (a). (E) Dorsolateral view of a Stage 15 embryo showing *oaf* RNA in clusters of cells in the brain (b) and in the embryonic gonad (go). The staining seen in the dorsal tracheoles is artifactual. (F) Ventral view of a Stage 15 embryo showing *oaf* RNA in midline and more lateral cells of the CNS. A-F, anterior is left. A-D, dorsal is up. (G) Male gonad of third instar larva showing *oaf* RNA in primary spermatocytes (sc) but not in spermatogonia (sg). (H) Adult female ovarioles showing accumulation in nurse cells (nc) and developing oocytes (o). RNA is not detected in the most mature oocytes because the egg membranes block probe penetration.

At Stage 15 and later, uniform β-gal staining is present throughout the embryo but its level is lower than before. At these stages, higher levels of *oaf* RNA normally appear in a segmentally repeated pattern in the CNS. However, E-32 does not reproduce this pattern (Figure 9D), perhaps because the CNS enhancer lies downstream of the transposon and is separated from the *lacZ* promoter by the intervening *rosy* gene.

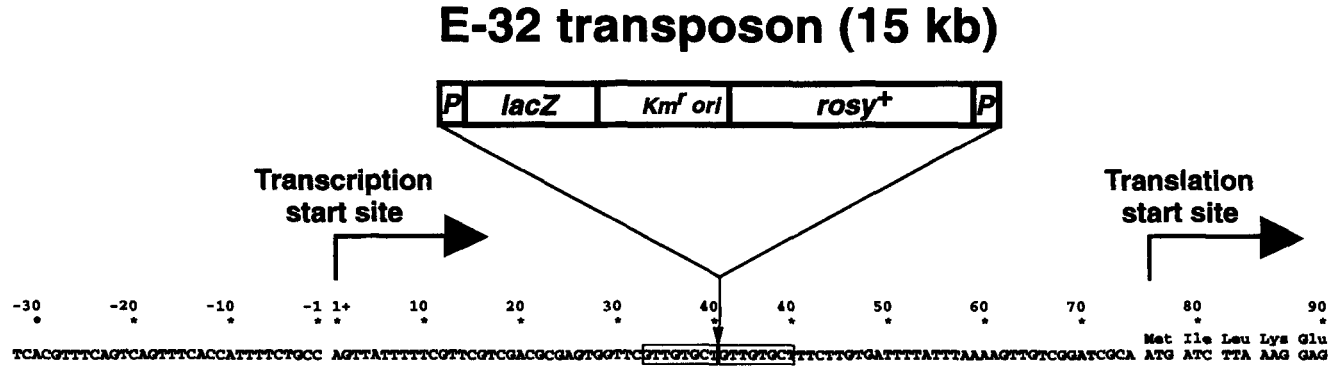

FIGURE 8.—The insertion site of the E-32 enhancer trap transposon. The 15 kb transposon is inserted in the 5′ UTR of the *oaf* transcription unit after nucleotide 40. The locations of the *P* element ends (*P*), the *E. coli lacZ* gene, the kanamycin resistance gene (Km^r), bacterial origin of replication (*ori*), and the Drosophila *rosy* gene (*rosy*$^+$) within the transposon are noted. The duplicated 8-bp target sequence (33–40) is boxed. The transcriptional and translational start sites of *oaf* are indicated by bent arrows. The sequence numbering is from Figure 3.

β-gal staining in male gonads of third instar larvae is in the same graded distribution as *oaf* RNA (Figure 9E). Female larval gonads also stain suggesting that *oaf* is transcribed there as well (C. MERLI, unpublished results). Patterns of β-gal in the larval brain and nerve cord are complex (Figure 9F). Higher levels of expression are seen in the central portion of each brain lobe, along the lateral edges and midline of the nerve cord, and transversely across the nerve cord halfway down its length. Areas of weaker expression include the lateral, rounded surfaces of each brain lobe, the area between the brain lobes and the extreme posterior of the nerve cord. We also see β-gal in individual nuclei within the intersegmental nerves extending from the nerve cord.

As reported by BROOK *et al.* (1993), E-32 also produces β-gal in the imaginal disks. This is in striking contrast to the lack of authentic *oaf* transcripts in these tissues. The analysis of these patterns has important implications for the regulation of the *dpp* and *oaf* genes and will be the subject of another communication.

We have also observed β-gal in adult male testes and female ovaries in patterns reflecting *oaf* RNA expression. In the testis, β-gal staining reflects the distribution of *oaf* RNA seen in the larval gonad. No β-gal is observed in the proliferative gonial cells at the apical end of the testis but expression is seen in the primary and secondary spermatocytes, spermatids and spermatogonia (Figure 9G). In ovarioles, staining is seen within the germarium and in nurse cells of all stages (Figure 9H). A small amount of β-gal is also seen in the maturing oocyte nucleus but more would be expected if all of the active enzyme in the nurse cells had been contributed to the oocyte.

Isolation and characterization of *oaf* lethal alleles: By remobilizing the E-32 transposon, we have generated new mutations in the *oaf* gene resulting from the imprecise excision of the transposon. From 70 independent remobilization events, we obtained seven lines that failed to homozygose (*oaf*$^{6.17}$, *oaf*$^{6.28}$, *oaf*$^{6.36}$, *oaf*$^{8.18}$, *oaf*$^{8.19}$, *oaf*$^{8.33}$ and *oaf*$^{8.37}$). Six of these alleles (all but *oaf*$^{8.19}$) failed to complement *dpp*^{d14}, a deficiency removing the three nearest lethal complementation groups [*l(2)ND1*, *l(2)ND2*, and *l(2)ND3*] proximal of *dpp*, showing that these chromosomes carried newly induced lethal mutations in the appropriate region. We then tested the six lines against *dpp*^{d19}, a deletion beginning in *dpp* and removing *l(2)ND1* only, and *l(2)ND1*^{H39} (an EMS-induced allele). All six failed to complement these tester alleles. From this, we conclude that the *oaf* gene is equivalent to the *l(2)ND1* complementation group previously identified by SPENCER *et al.* (1982). This locus has also been designated *l(2)22Fb* (LINDSLEY and ZIMM 1992) and *transcript near dpp (tnd)* (FlyBase 1994).

Southern blot and sequence analysis of the six lethals shows that each has lost some or all of the rightmost end of the transposon (Figure 10). Only some of these deletions actually extend into the *oaf* coding sequences. To explain all these results, we hypothesize that the 3′ end of the transposon contains a cryptic promoter needed for *oaf* transcription in the *E-32* chromosome. The size of the *oaf* larval RNA from the *E-32* line, 3.4 kb (D. BERGSTROM, unpublished results), also suggests that the promoter must be close to end of the transposon. By way of comparison, we have examined five homozygous viable alleles (*oaf*$^{6.3}$, *oaf*$^{6.31}$, *oaf*$^{6.35}$, *oaf*$^{8.9}$ and *oaf*$^{8.32}$). Each of these contain internal deletions of the transposon but retain its 3′ end (Figure 10). The deletion in the *oaf*$^{6.35}$ line extends beyond the 5′ end of the transposon and removes the entire *oaf* promoter and 1500 bp of upstream flanking sequence. Thus, the endogenous *oaf* promoter is not needed for the viability of the E-32 line.

Phenotypic analysis of *oaf* mutants: To determine

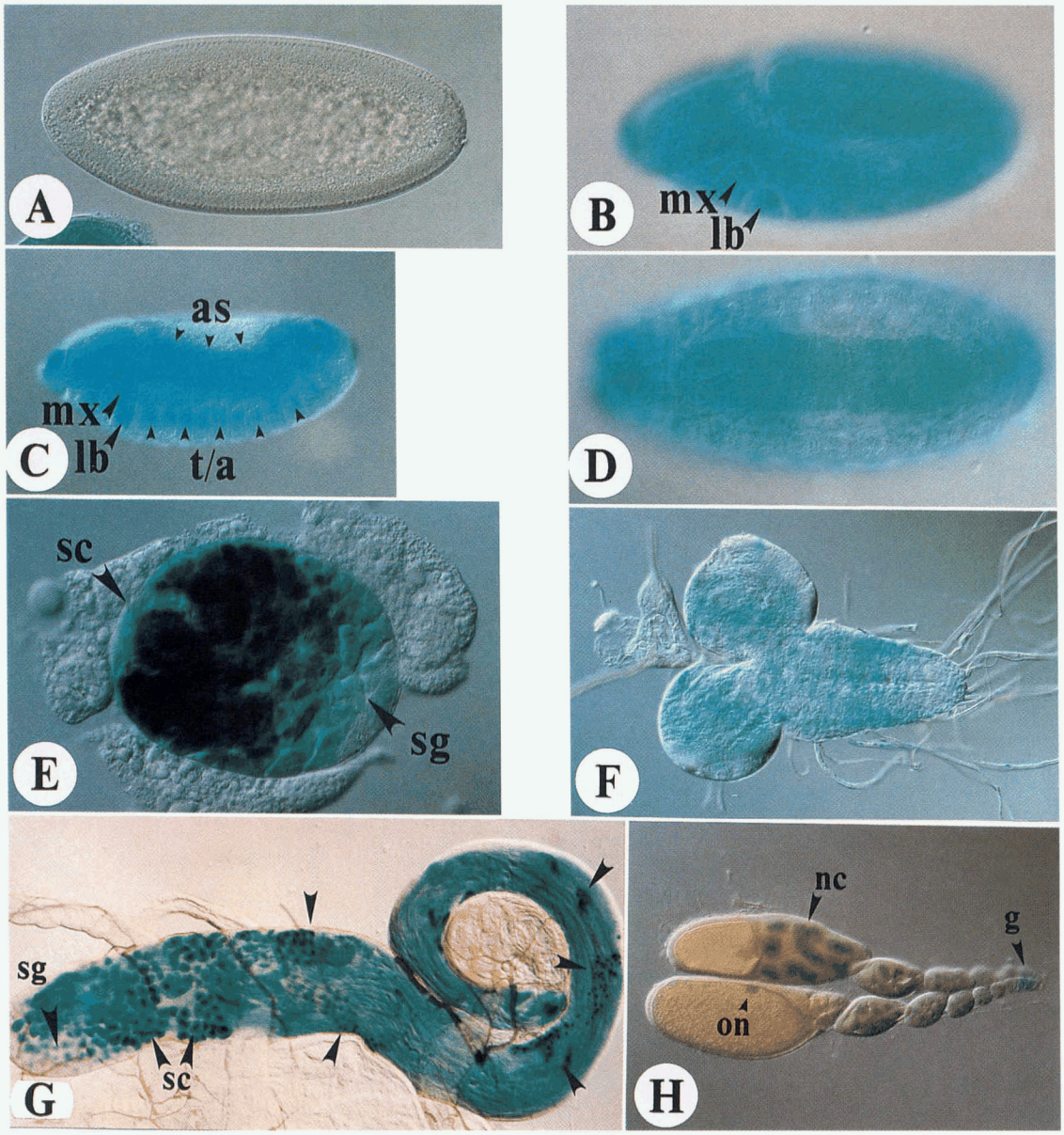

FIGURE 9.—Patterns of β-gal expression in *E-32* embryos, larvae and adults. (A) Stage 5 embryo showing no activity. (B) Stage 11 embryo with ubiquitous β-gal and localized regions of more intense staining in the maxillary (mx) and labial (lb) protuberances. (C) Stage 13 embryo with localized regions of staining in the maxillary (mx) and labial (lb) protuberances, in a thin stripe of cells bordering the amnioserosa (as), and in the thoracic and abdominal segments (t/a). (D) Ventral view of a Stage 15 embryo showing general expression throughout the embryo but no localized staining in the CNS. A-C, anterior is left, dorsal is up. D, anterior is left, ventral is up. (E) Male gonad of third instar larva showing strong β-gal staining in the primary spermatocytes (sc) and little or no activity in the spermatogonia (sg). (F) β-gal activity in the brain and ventral nerve cord of a third instar larva. (G) Adult testis showing β-gal in primary spermatocytes (sc) and in cells in the more advanced stages of spermatogenesis (arrows). Little or no β-gal is seen in the spermatogonia. (H) Adult ovarioles showing β-gal activity in the germarium (g), nurse cells (nc) and oocyte nucleus (on). The nuclear localization of the β-gal, obvious in the ovarioles, is caused by a signal present in the P element sequences fused to the β-galactosidase protein.

the effective lethal phase for the *oaf* mutations, each was made heterozygous for a wild-type chromosome and crossed *inter se*. With each allele, ~25% of the progeny (presumably oaf^- homozygotes) died late in embryogenesis or early in the first larval instar. Many of the dead larvae remained partially enclosed in their egg shells. Because all six alleles, including those lacking part of the *oaf* protein coding region, gave the same lethal phenotype, we feel that this must be the zygotic loss of function phenotype for the gene. Given the time of lethality for the mutants, we have chosen to name this gene *out at first*.

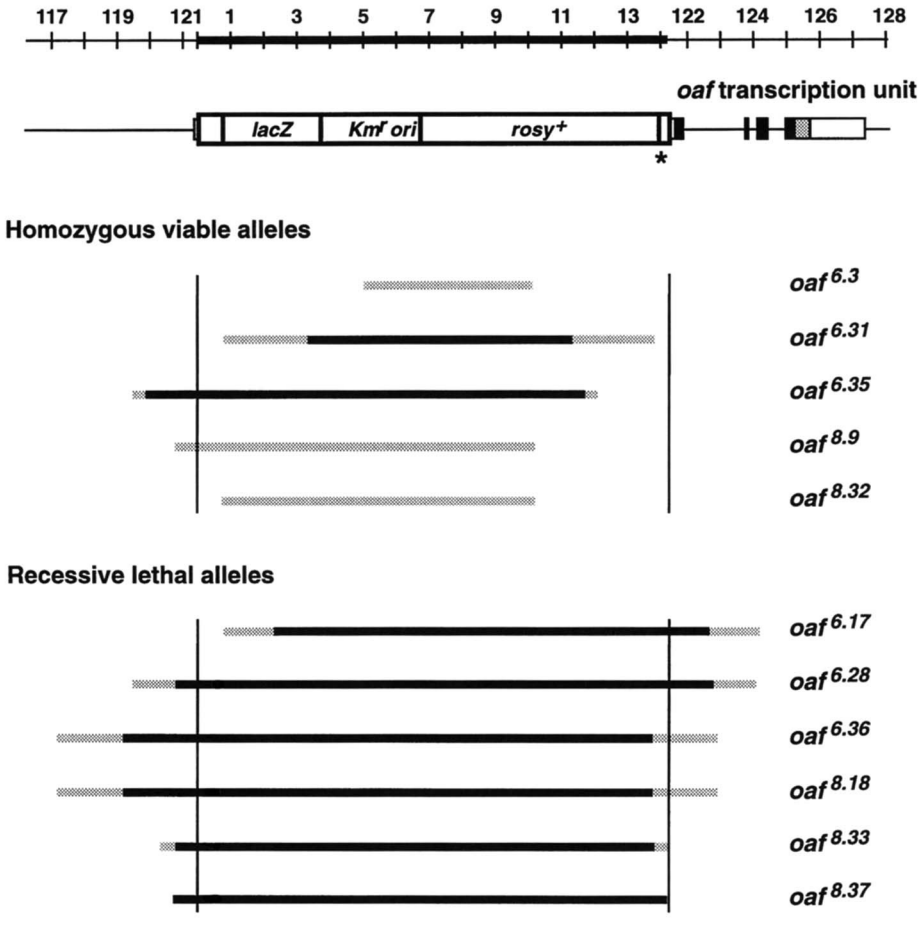

FIGURE 10.—Molecular analysis of *oaf* alleles produced by imprecise excision of the E-32 transposon. Map coordinates (in kilobases) are shown at the top; the transposon region (map units 1 to 14.2) is distinguished from the *oaf* sequences by the increased thickness of the line. Below this is a schematic representation showing important features of the E-32 transposon (large rectangles) and the exons of *oaf* (smaller rectangles). ORF 1 is depicted in black, ORF 2 in gray and the remaining sequences of a Class 1 transcript are in white. The presumed location of a cryptic promoter allowing *oaf* transcription in viable alleles is noted *. Deletions present in five homozygous viable and six recessive lethal *oaf* alleles are shown below the map. The minimum extent of each deletion is noted by a black line. The actual end points fall within the gray regions, which are the smallest intervals our Southern analysis could resolve. The breakpoints of the $oaf^{8.37}$ deletion were determined by sequencing the DNA spanning the deletion. The leftward and rightward extents of the E-32 transposon are demarcated by thin vertical lines.

As judged by cuticle preparations, the dead larvae exhibit no external patterning defects (D. BERGSTROM, unpublished results). Because *oaf* is expressed in the CNS, we investigated whether mutant animals exhibited nervous system defects. To this end, we stained mutant embryos with the monoclonal antibody BP102, which labels most or all CNS axons (SEEGER *et al.* 1993), and monoclonal antibodies directed against *engrailed* (*en*) and *prospero* (*pros*), two patterning genes expressed in portions of the CNS (VAESSIN *et al.* 1991; CUI and DOE 1992; MATSUZAKI *et al.* 1992). Using this small set of molecular probes, we have observed no reproducible CNS defects (D. BERGSTROM, unpublished results). LITTLETON and BELLEN (1994) have also looked for mutant phenotypes in embryos bearing *oaf* alleles. They studied the peripheral nervous system (PNS) of oaf^{-} embryos using the monoclonal antibody 22C10. Embryos heterozygous for a deficiency and oaf^{T28}, an EMS-induced lethal allele, exhibit sporadic growth cone guidance defects, particularly in tracts descending from the dorsal to lateral PNS clusters. Some lateral and dorsal clusters have reduced numbers of neurons and some segments have dorsally displaced lateral chordotonal organs. When they analyze mutants bearing our $oaf^{6.17}$ and $oaf^{6.28}$ alleles, they observe similar PNS defects but with lower penetrance (5% of the embryos showing phenotypic defects compared with 25% for oaf^{T28}/Df and 0% for wild type; H. BELLEN, personal communication). Taken together, these data suggest that *oaf* is necessary for proper neuronal development and hatching. However, this does not preclude the possibility that *oaf* is needed earlier in development, where its function is supplied or rescued by maternally contributed products, or later in development, where somatic clones will be needed to overcome the embryonic/larval lethality that prevents this analysis.

DISCUSSION

This paper presents the first molecular characterization of the Drosophila *out at first* gene. Transcription of *oaf* begins at 121.5 on our molecular map while the last known *dpp* enhancer is at 116.5 (Figure 1). Previous analysis has shown that there are no other transcribed sequences in the 116.5–119 region (ST. JOHNSTON *et al.* 1990) and our library screens failed to detect cDNAs derived from the 5-kb interval separating the two genes. Thus, we conclude that *oaf* is the first gene proximal of *dpp*.

***oaf* encodes multiple transcripts:** *oaf* produces three size classes of transcripts from the same TATA-less promoter. All three contain the same protein coding sequences but differ by the extent of their 3′ untranslated regions (UTRs). In the longest transcript, the 3′ UTR accounts for >50% of the mRNA. Because each of the transcript classes has a different developmental expression pattern, the differing 3′ UTRs may have functional significance, perhaps in the regulation of *oaf* mRNA translation (see below) and/or stage-specific stability.

The variation in UTR lengths occurs through the use of alternate polyadenylation sites. cDNA Classes 2 and 3, both likely to be contributed maternally during oogenesis, use a consensus AAUAAA polyadenylation signal while the predominant zygotic transcript (Class 1) contains the variant UAUAAA sequence. Use of a nonconsensus signal has been shown to reduce the efficiency with which transcripts are cleaved and polyadenylated (SHEETS *et al.* 1990). Thus, *oaf* transcripts levels may be regulated at the level of RNA cleavage and polyadenylation.

The heterogeneity of RNA sizes within Transcript Groups 2 and 3 suggests that additional processing of some RNAs must occur. At present, we cannot account for this level of variation.

***oaf* RNAs may encode multiple protein products:** Conceptual translation of ORF 1 predicts a basic protein of 37 kD with a cysteine-rich C-terminus. Interestingly, all of the *oaf* cDNAs contain a second ORF separated from the first by a single UGA stop codon. No significant matches could be found in the NCBI databases with either of these ORFs.

This unusual two-ORF structure raises the possibility that multiple protein products could arise from *oaf*. If the single UGA codon at the end of ORF 1 is suppressed, a read-through protein of 54 kD would be produced. Read-through translation of an UGA in Drosophila has previously been observed for the *kelch* gene (XUE and COOLEY 1993). The ORF 1 + 2 *kelch* protein has been detected as a minor product in western blots of ovarian proteins where the ORF 1 protein predominates (XUE and COOLEY 1993). However, the ORF 1 and ORF 1 + 2 *kelch* proteins are expressed equally in larvae and pupae (LYNN COOLEY, personal communication), showing that suppression can be developmentally regulated.

In bacteria and animals, UGA termination suppression is often mediated by the cotranslational insertion of selenocysteine at the stop codon (BERRY and LARSEN 1993). Selenocysteinyl tRNAs have been identified in representatives of all five kingdoms of life, including Drosophila (LEE *et al.* 1990). In mammals, selenocysteine incorporation requires specific selenocysteine insertion sequence (SECIS) elements found within the 3′ UTR of the transcript (BERRY *et al.* 1991, 1993; SHEN *et al.* 1993). If such sequences are important for *oaf* UGA suppression, it is likely that the three alternately polyadenylated *oaf* transcripts would have different capabilities to produce the ORF 1 + 2 protein. The smallest *oaf* transcript has <200 nt of 3′ UTR, whereas the longest has 1600 nt, a size more amenable to accommodate the structures proposed for SECIS function (BERRY *et al.* 1993). If produced, the ORF 1 + 2 protein would likely have very different properties than the ORF 1 polypeptide due to its several domains of relatively simple amino acid composition. We are in the process of making antibodies to the *oaf* polypeptides to address some of these issues.

Maternal and zygotic *oaf* transcripts have different structures and distributions: Our analysis of *oaf* transcripts in the embryo, larva and adult indicates that *oaf* transcription has both maternal and zygotic components. The maternally contributed RNAs, present ubiquitously in the maturing oocyte and the early embryo, are composed mostly or wholly of the Group 2 and 3 transcripts. These RNAs persist until most are degraded at the time of germ-band retraction, after which a lower level of ubiquitous *oaf* RNA, probably of maternal and zygotic origin, is maintained. Localized zygotic transcription of *oaf* is first observed at Stage 11 in small clusters of cells in the gnathal, thoracic and abdominal segments and in single cells bordering the amnioserosa. By Stage 15, only CNS and gonad expression is seen above the ubiquitous *oaf* levels. The 3.3 kb RNA is the predominant zygotic transcript during these stages and it continues to be expressed throughout larval development. In the oldest larvae, *oaf* is expressed in male (and probably female) gonads but is absent from the imaginal disks.

The role of *oaf* in development: Flies homozygous for recessive lethal alleles of *oaf* die late in embryogenesis or early during the first larval instar with no gross morphological defects. These observations have been confirmed by the analysis of LITTLETON and BELLEN (1994) who have independently observed the same effective lethal phase for EMS-induced alleles of *oaf* and found that these animals had PNS patterning defects, albeit with low penetrance. They have observed similar phenotypes in two of our transposon-induced mutations which remove portions of ORF 1 (H. BELLEN, personal communication), and we feel that these phenotypes are indicative of the null phenotype. It is likely that the incomplete penetrance observed for every mutation examined to date results from the presence of residual maternal *oaf* activity (supplied by the heterozygous mother) that partially rescues the zygotic requirement for *oaf*. However, homozygous *oaf* mutants that escape PNS defects are still destined to die by the first instar. This suggests that additional roles for *oaf* remain to be discerned, a likely situation given the additional

patterns of *oaf* expression in the CNS and the larval and adult gonads of both sexes. In addition, the ubiquitous accumulation of maternally contributed *oaf* RNA during early embryogenesis suggests that *oaf* may be needed at these times also. Germline clones will be needed to remove this class of mRNA from embryos to investigate these earlier roles.

The independent transcriptional regulation of *oaf* and *dpp*: Enhancers that control the imaginal disk-specific transcription of *dpp* lie 3′ of the *dpp* transcription unit in the 30 kb *disk* region (MASUCCI *et al.* 1990; ST. JOHNSTON *et al.* 1990; BLACKMAN *et al.* 1991; R. BLACKMAN and T. GILLEVET, unpublished results). The most distant of these enhancers are from 35 kb to over 55 kb away from the alternate *dpp* promoters with which they interact (Figure 1). The same *dpp* enhancers however, are within 5–10 kb of the *oaf* promoter. The lack of detectable *oaf* transcription within imaginal disks suggests that the *oaf* promoter does not respond to the *dpp* enhancers despite their proximity.

Two models are commonly invoked to explain how adjacent genes can maintain such regulatory autonomy. The first, the domain boundary model, proposes that both ends of the gene are tethered to a physical structure (*e.g.*, a nucleoprotein complex or the nuclear scaffold) and this serves to insulate the domain from external regulatory influences (EISSENBERG and ELGIN 1991; LAEMMLI *et al.* 1992). In Drosophila, excellent candidates for such boundary elements have been identified (GYURKOVICS *et al.* 1990; KELLUM and SCHEDL 1991, 1992; ROSEMAN *et al.* 1993). When one of these sequences is placed between an enhancer and its promoter, the boundary element blocks the interaction of the two regulatory elements. Such a boundary sequence placed between *dpp* and *oaf* would keep these genes independent of the other. A second model, the promoter specificity model, states that some property inherent to the promoter itself allows it to respond to some enhancers but not others (CHERBAS and CHERBAS 1993). There is good evidence that this type of mechanism is involved in keeping the Drosophila *gooseberry* and *gooseberry-neuro* genes from cross-regulating one another (LI and NOLL 1994).

Although we cannot answer which, if either, of these models is the mechanism establishing the autonomy of *dpp* and *oaf*, we have one observation that may be relevant to this issue. During a previous analysis of the *disk* region regulatory sequences, we found evidence for an enhancer that directed expression in embryos (BLACKMAN *et al.* 1991). From a more extensive analysis, we now know that this element lies within the 114.5–115 interval, with other *dpp* disk-specific enhancers to its right (Figure 1; R. BLACKMAN and T. GILLEVET, unpublished results). This enhancer activates reporter gene expression in a pattern that looks remarkably similar to the localized expression of the *oaf* gene in Stage 11 embryos (BLACKMAN *et al.* 1991). Both are stained in the mandibular and maxillary protuberances, in small clusters laterally in the thoracic and abdominal segments, as well as in single cells at the ectoderm/amnioserosa border. If, in fact, this enhancer does affect *oaf* transcription during embryogenesis but other (closer) *disk* region enhancers ignore the *oaf* promoter in the disks, it would argue against the presence of a domain boundary situated between the two genes unless its presence is developmentally regulated.

We are actively pursuing the mechanism involved in maintaining the autonomy of these genes. The E-32 transposon affords us the opportunity to test the promoter specificity model directly. GLOOR *et al.* (1991) have shown that it is possible to introduce new sequences at the site occupied by a *P* element. Because the transposon is removed during this process, the alteration is introduced into an otherwise normal chromosome. For our experiments, we can exchange the *oaf* promoter for one known to interact with the *dpp* enhancers and then test if *oaf* transcription is influenced by the *dpp* regulatory elements. This should provide a definitive test for the promoter specificity model.

We thank BILL BROOK and MIKE RUSSELL for the E-32 enhancer trap line and sharing unpublished data. We are grateful to HUGO BELLEN for his efforts in characterizing our mutant lines, CHRIS DOE for nervous system antibodies, HUGH ROBERTSON and BILL ENGELS for *P*-element primers and the Genetic Engineering Facility at the University of Illinois for their DNA sequencing assistance. This work was supported by grants from the National Science Foundation (DCB 90–18618 and MCB 93–17701) and the Research Board of the University of Illinois.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

BELLEN, H. J., C. J. O'KANE, C. WILSON, U. GROSSNIKLAUS, R. K. PEARSON *et al.*, 1989 P-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. Genes Dev. **3:** 1288–1300.

BENSON, M., and V. PIRROTTA, 1988 The Drosophila *zeste* protein binds cooperatively to sites in many gene regulatory regions: implications for transvection and gene regulation. EMBO J. **7:** 3907–3915.

BERRY, M. J., and P. R. LARSEN, 1993 Recognition of UGA as a selenocysteine codon in eukaryotes: A review of recent progress. Biochem. Soc. Trans. **21:** 827–831.

BERRY, M. J., L. BANU, Y. CHEN, S. J. MANDEL, J. D. KIEFFER *et al.*, 1991 Recognition of UGA as a selenocysteine codon in Type I deiodinase requires sequences in the 3′ untranslated region. Nature **353:** 273–276.

BERRY, M. J., L. BANU, J. W. HARNEY and P. R. LARSEN, 1993 Functional characterization of the eukayotic SECIS elements which direct selenocysteine insertion at UGA codons. EMBO J. **12:** 3315–3322.

BLACKMAN, R. K., M. SANICOLA, L. A. RAFTERY, T. GILLEVET and W. M. GELBART, 1991 An extensive 3′ cis-regulatory region directs the imaginal disk expression of *decapentaplegic*, a member of the TGF-β family of *Drosophila*. Development **111:** 657–665.

BLAIR, S., 1992 *engrailed* expression in the anterior lineage compart-

ment of the developing wing blade of *Drosophila*. Development **115:** 21–33.

BROOK, W. J., L. M. OSTAFICHUK, J. PIORECKY, M. D. WILKINSON, D. J. HODGETTS *et al.*, 1993 Gene expression during imaginal disc regeneration detected using enhancer-sensitive P-elements. Development **117:** 1287–1297.

BROWN, N. H., and F. C. KAFATOS, 1988 Functional cDNA libraries from *Drosophila* embryos. J. Mol. Biol. **203:** 425–437.

CAPOVILLA, M., M. BRANDT and J. BOTAS, 1994 Direct regulation of *decapentaplegic* by *Ultrabithorax* and its role in *Drosophila* midgut morphogenesis. Cell **76:** 461–475.

CHERBAS, L., and P. CHERBAS, 1993 The arthropod initiator: the capsite consensus plays an important role in transcription. Insect Biochem. Mol. Biol. **23:** 81–90.

CHURCH, G. M., and W. GILBERT, 1984 Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991–1995.

COHEN, S. M., 1993 Imaginal disk development, pp. 747–841 in *Development of Drosophila*, edited by A. MARTINEZ-ARIAS and M. BATE. Cold Spring Harbor Press, Cold Spring Harbor, NY.

CUI, X., and C. Q. DOE, 1992 *ming* is espressed in neuroblast sublineages and regulates expression in the *Drosophila* central nervous system. Development **116:** 943–952.

EISSENBERG, J. C., and S. C. R. ELGIN, 1991 Boundary functions in the control of gene expression. Trends Genet. **7:** 335–340.

ENGELS, W. R., 1989 P elements in *Drosophila melanogaster*, pp. 437–484 in *Mobile DNA*, edited by D. E. BERG and M. M. HOWE. American Society for Microbiolgy, Washington, D.C.

FERGUSON, E. L., and K. V. ANDERSON, 1992 *decapentaplegic* acts as a morphogen to organize dorsal/ventral pattern in the *Drosophila* embryo. Cell **71:** 451–461.

FLYBASE, 1994 Genetic loci. Dros. Info. Serv. **73:** 461.

GELBART, W. M., and C. T. WU, 1982 Interactions of *zeste* mutations with loci exhibiting transvection effects in *Drosophila melanogaster*. Genetics **102:** 179–189.

GLOOR, G. B., N. A. NASSIF, D. M. JOHNSON-SCHLITZ, C. P. PRESTON and W. R. ENGELS, 1991 Targeted gene replacement in *Drosophila* via P-element induced gap repair. Science **253:** 1110–1117.

GLOOR, G., C. R. PRESTON, D. M. JOHNSON-SCHLITZ, N. A. NASSIF, R. W. PHILLIS *et al.*, 1993 Type I repressors of P element mobility. Genetics **135:** 81–95.

GYURKOVICS, H., J. GAUSZ, J. KUMMER and F. KARCH, 1990 A new homeotic mutation in the *Drosophila* bithorax complex removes a boundary separating two domains of regulation. EMBO J. **9:** 2579–2585.

HATFIELD, D., and A. DIAMOND, 1993 UGA: a split personality in the universal genetic code. Trends Genet. **9:** 69–70.

HOLTON, T. A., and M. W. GRAHAM, 1990 A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res. **19:** 1156.

HUANG, J.-D., D. H. SCHWYTER, J. M. SHIROKAWA and A. J. COUREY, 1993 The interplay between multiple enhancer and silencer elments defines the pattern of *decapentaplegic* expression. Genes Dev. **7:** 694–704.

HURSH, D. A., R. W. PADGETT and W. M. GELBART, 1993 Cross regulation of *decapentaplegic* and *Ultrabithorax* transcription in the embryonic visceral mesoderm of *Drosophila*. Development **117:** 1211–1222.

IMMERGLÜCK, K., P. A. LAWERENCE and M. BIENZ, 1990 Induction across germ layers in *Drosophila* mediated by a genetic cascade. Cell **62:** 261–268.

IRISH, V. F., and W. M. GELBART, 1987 The *decapentaplegic* gene is required for dorsal-ventral patterning of the *Drosophila* embryo. Genes Dev. **1:** 868–879.

JACKSON, P. D. and F. M. HOFFMANN, 1994 Embryonic expression patterns of the *Drosophila decapentaplegic* gene: separate regulatory elements control blastoderm expression and lateral ectodermal expression. Dev. Dynamics **199:** 28–44.

JACOBS, J. R., Y. HIROMI, N. H. PATEL and C. S. GOODMAN, 1989 Lineage, migration, and morphogenesis of longitudinal glia in the *Drosophila* CNS as revealed by a molecular lineage marker. Neuron **2:** 1625–1631.

JAHNIG, F., 1990 Structure predictions of membrane proteins are not that bad. Trends Biochem. Sci. **15:** 93–95.

JIANG, J., D. KOSMAN, Y. T. IP and M. LEVINE, 1991 The *dorsal* morphogen gradient regulates the mesoderm determinant *twist* in early *Drosophila* embryos. Genes Dev. **5:** 1881–1891.

KELLUM, R., and P. SCHEDL, 1991 A position-effect assay for boundaries of higher order chromosomal domains. Cell **64:** 941–950.

KELLUM, R., and P. SCHEDL, 1992 A group of scs elements function as domain boundaries in an enhancer-blocking assay. Mol. Cell Biol. **12:** 2424–2431.

KINGSLEY, D. M., 1994 The TGF-β superfamily: new members, new receptors, and new genetic tests of function in different organisms. Genes Dev. **8:** 133–146.

LAEMMLI, U. K., E. KÄS, L. POLJAK and Y. ADACHI, 1992 Scaffold-associated regions: *cis*-acting determinants of chromatin structural loops and functional domains. Curr. Opin. Gen. Dev. **2:** 275–285.

LEE, B. J., M. RAJAGOPALAN, Y. S. KIM, K. H. YOU, K. B. JACOBSON *et al.*, 1990 Selenocysteine tRNA [Ser]Sec gene is ubiquitous within the animal kingdom. Mol. Cell. Biol. **10:** 1940–1949.

LI, X., and M. NOLL, 1994 Compatibility between enhancers and promoters determines the transcriptional specificity of *gooseberry* and *gooseberry neuro* in the *Drosophila* embryo. EMBO J. **13:** 400–406.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego, CA.

LITTLETON, T., and H. BELLEN, 1994 Genetic and phenotypic analysis of 13 essential genes in the cytologic interval 22F1–2; 23B1–2 reveals three genes required for neural development in *Drosophila*. Genetics **138:** 111–123.

MASUCCI, J. D., and M. F. HOFFMANN, 1993 Identification of two regions from the *Drosophila decapentaplegic* gene required for embryonic midgut development and larval viability. Dev. Biol. **159:** 276–287.

MASUCCI, J. D., R. J. MILTENBERGER and F. M. HOFFMANN, 1990 Pattern-specific expression of the *Drosophila decapentaplegic* gene in imaginal disks is regulated by 3' *cis*-regulatory elements. Genes Dev. **4:** 2011–2023.

MATSUZAKI, F., K. KOIZUMI, C. HAMA, T. YOSHIKA and Y.-I. NABESHIMA, 1992 Cloning of the *Drosophila prospero* gene and its expression in ganglion mother cells. Biochem. Biophys. Res. Comm. **182:** 1326–1332.

PADGETT, R. W., R. D. ST. JOHNSTON and W. M. GELBART, 1987 A transcript from a *Drosophila* pattern gene predicts a protein homologous to the Transforming Growth Factor-β family. Nature **325:** 81–84.

PANGANIBAN, G. E. F., R. REUTER, M. P. SCOTT and F. M. HOFFMANN, 1990 A *Drosophila* growth factor homolog, *decapentaplegic*, regulates homeotic gene expression within and across germ layers during midgut morphogenesis. Development **110:** 1041–1050.

PARIS, M., H. XIAO and P. FLICK, 1993 Sensitive northern and Southern analysis using chemiluminescence. U.S.B. Editorial Comments **20:** 29–35.

PIRROTTA, V., 1986 Cloning *Drosophila* genes, pp. 83–110 in *Drosophila: A Practical Approach*, edited by D. B. ROBERTS. IRL Press, Oxford.

PONCE, M. R., and J. L. MICOL, 1992 PCR amplification of long DNA fragments. Nucleic Acids Res. **20:** 623.

POSAKONY, L. M., L. A. RAFTERY and W. M. GELBART, 1991 Wing formation in *Drosophila melanogaster* requires *decapentaplegic* gene function along the anterior-posterior compartment boundary. Mech. Dev. **33:** 69–82.

RAFTERY, L. A., M. SANICOLA, R. K. BLACKMAN and W. M. GELBART, 1991 The relationship of *decapentaplegic* and *engrailed* expression in *Drosophila* imaginal disks: do these genes mark the anterior-posterior compartment boundary? Development **113:** 27–33.

REUTER, R., G. E. F. PANGANIBAN, F. M. HOFFMANN and M. P. SCOTT, 1990 Homeotic genes regulate spatial expression of putative growth factors in the visceral mesoderm of *Drosophila* embryos. Development **110:** 1031–1040.

ROBERTSON, H. M., C. R. PRESTON, R. W. PHILLIS, D. M. JOHNSON-SCLITZ, W. K. BENZ *et al.*, 1988 A stable source of *P* element transposase in *Drosophila melanogaster*. Genetics **118:** 461–470.

ROSEMAN, R. R., V. PIRROTTA and P. K. GEYER, 1993 The *su(Hw)*

protein insulates expression of the *Drosophila melanogaster white* gene from chromosomal position-effects. EMBO J. **12:** 435–442.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SEEGER, M., G. TEAR, D. FERRES-MARCO and C. GOODMAN, 1993 Mutations affecting growth cone guidance in *Drosophila*: genes necessary for guidance toward or away from the midline. Neuron **10:** 409–426.

SHEETS, M. D., S. C. OGG and M. P. WICKENS, 1990 Point mutations in AAUAAA and the poly(A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation *in vitro*. Nucleic Acids Res. **18:** 5799–5805.

SHEN, Q., F.-F. CHU and P. E. NEWBURGER, 1993 Sequences in the 3′-untranslated region of the human cellular glutathione peroxidase gene are necessary and sufficient for selenocysteine incorporation at the UGA codon. J. Biol. Chem. **268:** 11463–11469.

SMALE, S. T., and D. BALTIMORE, 1989 The "Initiator" as a transcription control element. Cell **57:** 103–113.

SPENCER, F. A., F. M. HOFFMANN and W. M. GELBART, 1982 *Decapentaplegic*: a gene complex affecting morphogenesis in *Drosophila melanogaster*. Cell **28:** 451–461.

ST. JOHNSTON, R. D., F. M. HOFFMANN, R. K. BLACKMAN, D. SEGAL, R. GRIMAILA *et al.*, 1990 Molecular organization of the *decapentaplegic* gene in *Drosophila melanogaster*. Genes Dev. **4:** 1114–1127.

TAUTZ, D., and C. PFEIFLE, 1989 A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. Chromosoma **98:** 81–85.

THUMMEL, C., 1993 Compilation of *Drosophila* cDNA and genomic libraries. Dros. Info. Serv. **72:** 180–183.

VAESSIN, H., E. GRELL, E. WOLFF, E. BIER, L. Y. JAN *et al.*, 1991 *prospero* is expressed in neuronal precursors and encodes a nuclear protein that is involved in the control of axonal outgrowth in *Drosophila*. Cell **67:** 941–953.

WHARTON, K. A., R. P. RAY and W. M. GELBART, 1993 An activity gradient of decapentaplegic is necessary for the specification of dorsal pattern elements in the Drosophila embryo. Development **117:** 807–822.

WU, C.-T., and M. L. GOLDBERG, 1989 The *Drosophila zeste* gene and transvection. Trends Genet. **5:** 189–194.

XUE, F., and L. COOLEY, 1993 *kelch* encodes a component of intercellular bridges in *Drosophila* egg chambers. Cell **72:** 681–693.

Communicating editor: P. CHERBAS