Cell, Vol. 61, 85–99, April 6, 1990, Copyright © 1990 by Cell Press

The Drosophila 74EF Early Puff Contains *E74*, a Complex Ecdysone-Inducible Gene That Encodes Two *ets*-Related Proteins

Kenneth C. Burtis,*‡ Carl S. Thummel,*† C. Weldon Jones,*§ Felix D. Karim,† and David S. Hogness*

* Department of Biochemistry
Stanford University School of Medicine
Stanford, California 94305-5307
† Howard Hughes Medical Institute
Department of Human Genetics
University of Utah Medical Center
Salt Lake City, Utah 84132

Summary

We have isolated an ecdysone-inducible gene, *E74*, from the early puff at position 74EF in the Drosophila polytene chromosomes. We show that *E74* consists of three nested transcription units that derive from unique promoters but share a single polyadenylation site. The 60 kb *E74A* unit is directly induced by ecdysone and leads to the synthesis of a 6.0 kb mRNA that contains an unusually long 5′ leader (1891 nucleotides) with 17 short ORFs. Within the fifth of the seven *E74A* introns are two *E74B* promoters that direct the synthesis of 4.8 and 5.1 kb mRNAs. The nested arrangement of these transcription units leads to the formation of two *E74* proteins, each with a unique N-terminal domain joined to a common C-terminal domain. The unique N-terminal domains contain regions rich in acidic amino acids while the C-terminal domain is rich in basic amino acids and is very similar to proteins encoded by the *ets* proto-oncogene superfamily.

Introduction

The larval-to-adult metamorphosis of Drosophila melanogaster is initiated by a pulse of the steroid hormone 20-OH ecdysone (hereafter referred to as ecdysone) that occurs at the end of the third larval instar. This pulse triggers a coordinated change in the developmental pathways of both the imaginal tissues that yield the adult structures and the strictly larval tissues, most of which eventually histolyse.

The first clues to the genetic regulation of these coordinated changes in tissue-specific developmental pathways were derived from the effects of ecdysone on the pattern of transcription puffs in the polytene chromosomes of the larval salivary glands. Exposure of late third instar glands to ecdysone results in the rapid induction of a small set of "early" puffs, followed by the induction of a much larger set of "late" puffs and the regression of the early puffs. Detailed studies of the conditions required for these changes in puffing activity led Ashburner and his associates to propose the hierarchical model for the genetic regulation of puffing shown in Figure 1 (Ashburner et al., 1974).

We are interested in the possibility that this model is applicable to each target tissue and forms the basis for the temporal coordination of their developmental pathways by ecdysone. Thus, in each of these tissues, an ecdysone–receptor complex would be formed that induces the transcription of a small set of regulatory genes (early genes) whose proteins control two fundamental processes: the turning-on of a large set of effector genes (late genes), whose products effect a change in those morphological and functional properties of the tissue that define its developmental state; and the turning-off of early gene expression so as to determine the duration of that expression. Such a hierarchical model for the coordination of tissue development may also apply to other developmental stages similarly characterized by a pulse of ecdysone. Such additional pulses have been observed at mid-embryogenesis and during the first and second larval instars, as well as during prepupal and pupal development (reviewed by Richards, 1981).

The postulate that ecdysone acts via a complex with a receptor protein was based on studies of vertebrate steroid hormones, as it preceded evidence for an ecdysone receptor in D. melanogaster (Ashburner et al., 1974). Evidence for such a receptor initially consisted in the detection of proteins that exhibit ecdysone-specific binding ($K_D \approx 10^{-7}$ M) in extracts or partially purified preparations of tissue culture cells, embryos, and third instar imaginal discs and salivary glands (Maroy et al., 1978; Yund et al., 1978; Sage et al., 1982; Schaltmann and Pongs, 1982; Cherbas et al., 1984). Subsequently, Riddihough and Pelham (1987) detected a protein in extracts of cultured cells that exhibited specific binding to a 23 bp ecdysone-responsive element that they had defined in the promoter of the ecdysone-inducible Drosophila heat shock gene *hsp27*. The question as to whether this protein also exhibited ecdysone-specific binding, however, remained open. This question has recently been answered by the isolation of a D. melanogaster gene, *EcR*, that encodes a member of the steroid receptor superfamily (M. Koelle, W. A. Segraves, W. Talbot, M. Bender, and D. S. H., unpublished data; see Evans, 1988, for a review of this superfamily). The *EcR* protein exhibits specific binding with both ecdysone and the above 23 bp element; furthermore, the ecdysone activation of promoters carrying this element in receptor-deficient cultured cells requires the expression of the *EcR* gene in a cotransfection assay (M. Koelle, P. Cherbas, and D. S. H., unpublished data).

Given this evidence for an ecdysone receptor with properties akin to those postulated in the Ashburner model (Figure 1), the early regulatory genes constitute a critical focus for tests of this model and its extension proposed here. It is that focus which has prompted the experiments described here and in a subsequent study (Thummel et

‡ Present address: Department of Genetics, University of California, Davis, California, 95616.
§ Present address: Department of Biology, Bethel College, St. Paul, Minnesota, 55112.

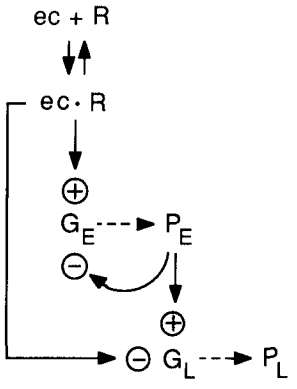

Figure 1. The Ashburner Model for Changes in the Activity of Early and Late Puffs Triggered by Ecdysone

The Ashburner model is from Ashburner et al. (1974). Ecdysone (ec) combines with its receptor (R) to form an ecdysone-receptor complex (ec·R) that induces the transcription of all early genes (G_E) and represses transcription of the late genes (G_L). One or more of the proteins (P_E) encoded by the G_E induce transcription of the G_L, overcoming their initial repression by ec·R. At least one of the P_E also represses all G_E to overcome their induction by ec·R, causing the early puffs to regress. The model allows, but does not demand, the existence of lower ranks in this regulatory hierarchy. In this case, G_L would represent only a subset of the genes responsible for the late puffs, and one or more of their proteins (P_L) would induce transcription of other late puff genes.

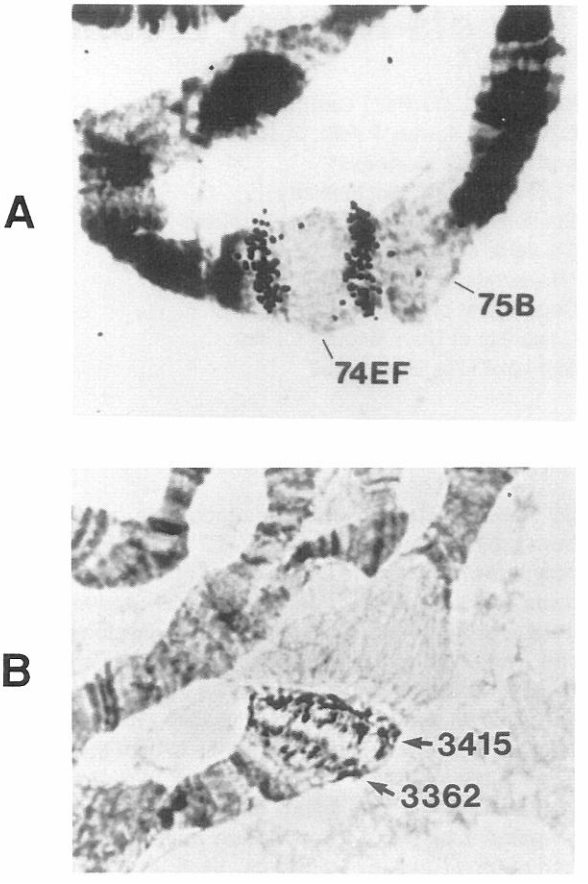

Figure 2. In Situ Hybridization of Genomic DNA Probes from the 74EF Region to Late Third Instar Salivary Gland Chromosomes

(A) Tritium-labeled probes were prepared from restriction fragments located at the ends of the 74EF walk, separated by 260 kb in the genome. These were hybridized to third instar salivary gland polytene chromosomes as described (Bonner and Pardue, 1976). The large puffs at 74EF and 75B are labeled.

(B) Biotin-labeled probes prepared from the EcoRI restriction fragment 3362 and phage 3415 DNA (see Figure 3) were hybridized to polytene chromosomes and detected as described (Langer-Safer et al., 1982).

al., 1990). Our entrée to the study of these genes is presently restricted to those defined by the early puffs in the larval salivary gland polytene chromosomes. The first part of our experimental plan is therefore to obtain a molecular definition of these genes that will provide the basis for testing the models.

In this paper, we provide such a molecular definition of an ecdysone-inducible gene called *E74*, which occupies the locus of a large, well-characterized early puff at 74EF. This gene consists of three nested transcription units that have a common polyadenylation site but different 5′ ends: a 60 kb *E74A* unit, and two units, *E74B1* and *E74B2*, with 5′ ends located 300 bp apart within the fifth and largest of the seven *E74A* introns (Figure 3). As a result, three of the eight *E74A* exons are included in the two *E74B* mRNAs in-frame with an AUG-initiated open reading frame (ORF) common to the 5′ exons of the *E74B* mRNAs. Consequently, the *E74* gene encodes two proteins with different N-terminal regions but common C-terminal regions. This common C-terminal region exhibits homology to the proteins encoded by the *ets*-1 and *ets*-2 proto-oncogenes.

Results

A Chromosomal Walk through the 74EF Puff Region

Cloning of genomic DNA from the 74EF region was made possible by the isolation of a cDNA clone, adm135A8, that hybridized in situ to 74E as well as to seven other chromosomal sites (Wolfner, 1980). This clone derives from the *Jonah* multigene family, which is expressed in the larval and adult guts (Carlson and Hogness, 1985). Because of sequence variation among the members of this family, a cDNA clone, pkdm26H2, failed to hybridize to 74E but did hybridize to the other seven positions detected by adm135A8. These two probes were therefore used in a differential screen of a lambda genomic library to isolate DNA from the 74E region. Two plaques were identified that hybridized to adm135A8, but not to pkdm26H2. Unique restriction fragments from one of these phage, lacking homology to adm135A8, were used to initiate chromosomal walks in both directions (Bender et al., 1983). Each walk was extended until it was possible to determine its direction within the polytene chromosome by in situ hybridization. Only that walk which proceeded from 74E toward 74F was continued, until genomic DNA spanning the entire puff region was included. Figure 2A shows an in situ hybridization using two tritium-labeled probes prepared from restriction fragments that are 260 kb apart and represent the ends of this walk.

The search for the gene or genes responsible for forma-

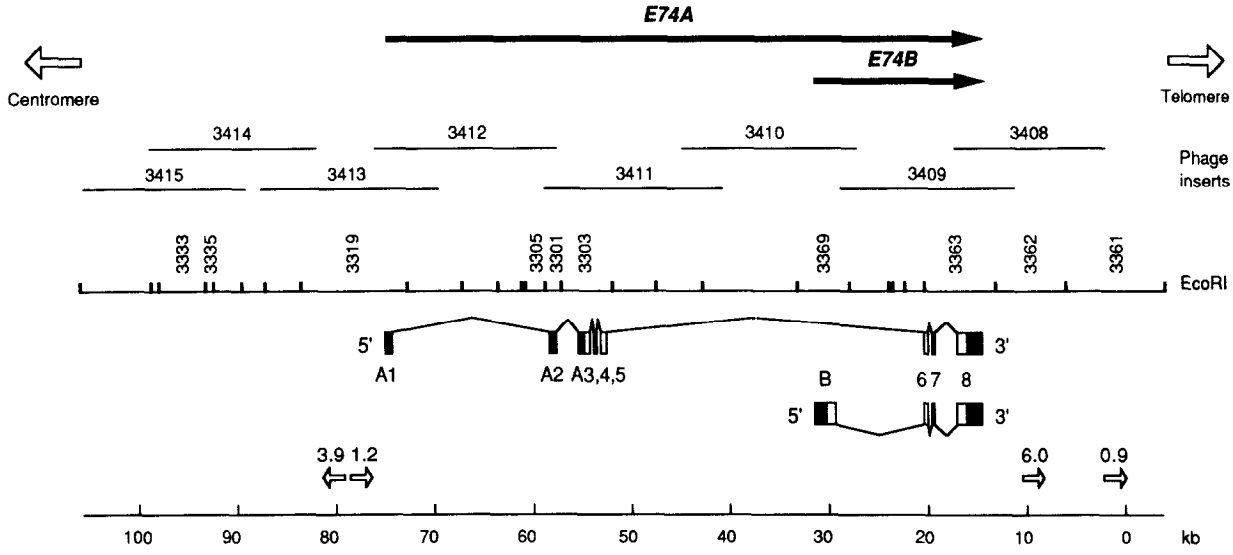

Figure 3. Structural Map of the *E74* Region

A map of 100 kb of genomic DNA is shown, encompassing the E74A *and E74B* transcription units (bold arrows). The direction of the arrows represents the orientation of *E74* transcription in relation to the centromere and telomere of the left arm of the third chromosome. The locations and names of genomic phage inserts spanning the region are shown above the EcoRI restriction map, as are the names of EcoRI fragments mentioned in the text. The boxes below the map show the approximate size and location of *E74* exons. Black boxes represent the 5' leader and 3' trailer regions, and open boxes represent the protein coding regions. The exons are named according to the convention described in the text. The different 5' ends of the *E74B* transcripts are not shown. Open arrows at the bottom indicate the location of EcoRI fragments that hybridize to other transcripts in the region and indicate the polarity of those transcripts. The sizes of the transcripts (in kb) are shown above the arrows. Map coordinates are shown on the bottom line in kb; the origin is the location of the *Jonah* gene used to initiate the walk.

tion of the 74EF puff was focused on a 105 kb region. A map of this portion of the walk is shown in Figure 3. This region spans the majority of the puff, as demonstrated in Figure 2B, which shows the in situ hybridization of biotin-labeled probes prepared from restriction fragment 3362 and the DNA of phage 3415 (see Figure 3 for the positions of these probes).

The *E74A* Transcription Unit

Low Resolution Mapping of an Ecdysone-Inducible Transcription Unit

Low resolution mapping of ecdysone-inducible exons within the 105 kb region was carried out by differential hybridization of ^{32}P-labeled cDNA probes to restriction fragments from phage clones 3408–3415, whose DNAs cover this region (Figure 3). The cDNA probes were prepared by randomly primed reverse transcription of poly(A)$^+$ RNA isolated from third instar salivary glands that had been incubated in vitro with or without ecdysone (Experimental Procedures). Hybridization of these probes to duplicate Southern blots of EcoRI and PstI digests of the cloned phage DNAs revealed three separated genomic regions that hybridized preferentially to the induced probe (data not shown)—regions delineated in left-to-right order in Figure 3 by four EcoRI fragments: 3319, the adjacent 3301 and 3303 fragments, and 3363.

Evidence that the ecdysone-inducible exons within these three regions are part of the same ecdysone-inducible transcription unit derives from Northern blot analyses of induced and noninduced RNAs hybridized with single-stranded probes from each of the four EcoRI fragments. Figure 4A shows the results of such an analysis in which both poly(A)$^+$ and poly(A)$^-$ RNAs from induced and noninduced third instar salivary glands were hybridized with a single-stranded probe from 3301 that detects only rightward transcription (i.e., from left to right in Figure 3). A single major poly(A)$^+$ RNA of 6.0 kb was observed in induced glands at an abundance at least 100-fold greater than that in noninduced glands, as determined by densitometry of the autoradiogram. Single-stranded probes from the 3319, 3303, and 3363 fragments similarly detect a 6.0 kb ecdysone-inducible RNA of the same polarity (data not shown). This provides strong evidence that the ecdysone-inducible exons in the four fragments are present in a single transcription unit that yields a 6.0 kb mRNA and is 61 ± 9 kb long, as judged from the positions of the 5' and 3' terminal fragments 3319 and 3363, respectively (Figure 3). We call this presumptive unit the *E74A* unit and the 6.0 kb poly(A)$^+$ RNA the *E74A* mRNA.

Figure 4B shows that ecdysone also induces the *E74A* unit in the presence of the protein synthesis inhibitor cycloheximide. It mimics the ecdysone induction of the 74EF and other early puffs in this respect (Ashburner et al., 1974), indicating both processes represent primary responses to this steroid hormone. The correspondence between these processes has been investigated in more detail as has the mechanism by which ecdysone induces *E74A* transcription (Thummel et al., 1990).

Identification and High Resolution Mapping of the Eight E74A Exons

Figure 3 shows the positions of the eight *E74A* exons on

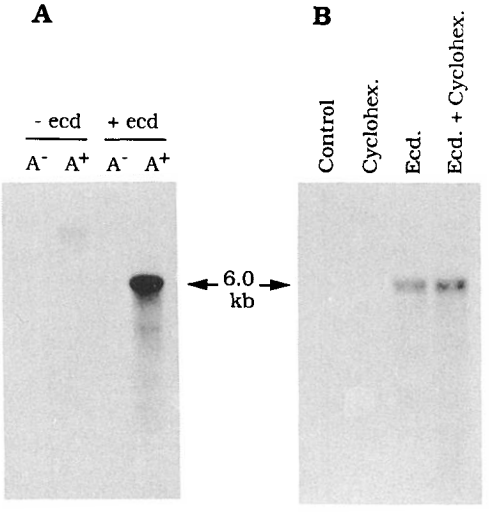

Figure 4. Effects of Ecdysone and Cycloheximide on *E74A* Transcription

(A) Northern blot analysis of poly(A)$^-$ and poly(A)$^+$ RNA isolated from third instar salivary glands incubated in the absence (−ecd) or presence (+ecd) of ecdysone.

(B) Northern blot analysis of total RNA isolated from third instar larval tissues incubated in Robb's saline alone (control) or in the presence of cycloheximide and/or ecdysone as indicated. In both cases, radioactive single-stranded DNA from EcoRI fragment 3301 was used as a probe. The high molecular weight band in the second lane of (A) was not reproducible and may have been DNA contamination.

the molecular map of the 105 kb puff region. The first five exons are not shared by the *E74B* transcription units described below and are therefore labeled A1–A5, whereas the last three exons are labeled 6–8 because they are shared. These exons were identified and molecularly mapped, and their nucleotide sequences were determined by a combination of S1 nuclease protection analysis, nucleotide sequence analysis of cloned cDNAs and genomic DNAs, and primer extension analysis.

Figure 5A shows the results of one of several S1 protection experiments that determined the lengths and positions of the *E74A* exons in each of the three regions shown in the preceding section to contain ecdysone-inducible exons. This experiment provided the lengths of the four exons in the central region defined by the adjacent EcoRI fragments 3301 and 3303 (Figure 3). As will become apparent, these exons correspond, respectively, to A2 and A3–A5. Lane 1 shows that a single 714 nucleotide fragment (A2) was protected from S1 nuclease digestion after a uniformly labeled, single-stranded probe from 3301 was annealed with poly(A)$^+$ RNA isolated from second day pupae. Similarly, lane 3 shows that an equivalent probe from 3303 yielded protected fragments that exhibit lengths of 611 (A3), 282 (A4), and 508 (A5) nucleotides. In addition, the position of A3 in relation to the XbaI site within 3303 can be deduced from the lengths of the two protected fragments shown in lane 2. These derive from the XbaI fragment 3321 shown in the schematic at the bottom of Figure 5A and correspond to the 714 nucleotide A2 exon and a 591 nucleotide XbaI-truncated derivative of the 611 nucleotide A3 exon.

Similar S1 protection analyses, in which end-labeled single-stranded probes were used (Weaver and Weissman, 1979), yielded the positions of A2, A3, and A4 (Figure 5A) as well as the length and position of A1, the only ecdysone-inducible exon detected in the upstream region defined by the EcoRI fragment 3319 (Figure 3). S1 protection and primer extension analyses provided the evidence that the 5′ end of exon A1 is the transcriptional initiation site of the *E74A* mRNA (Figure 5B). Although second day pupae provided the source of the poly(A)$^+$ RNA used in these experiments, the same result was obtained by primer extension with RNA templates derived from 18 hr embryos, ecdysone-induced third instar tissues, and fourth day pupae, as well as ecdysone-induced K_c tissue culture cells (data not shown). These results are consistent with the temporal and tissue distributions of *E74A* transcripts (Thummel et al., 1990) and indicate that the *E74A* transcription initiation site is invariant during development.

S1 protection analysis was also used for the approximate mapping of exons 6, 7, and 8, which are shared by the *E74A* and *E74B* units and were the only exons detected by S1 analysis within the 3363 EcoRI fragment that defines the 3′ terminal region of ecdysone-inducible exons (data not shown). The precise mapping of exons 6 (370 bp), 7 (251 bp), and 8 (2598 bp) was accomplished by comparison of the nucleotide sequence of a cDNA clone called c17 with the genomic DNA sequence of appropriate parts of 3363 and its subclones (Figure 6). The c17 cDNA is the longest of five clones isolated from an early pupal cDNA library that were identified as *E74A* cDNAs by their hybridization to probes from both 3363 and 3303. It includes the 3′ terminal 399 bp of A3 and extends downstream to a polyadenylated 3′ end that marks the 3′ terminus of exon 8, and hence that of *E74A* (Figure 6). Comparison of the c17 sequence with the complete sequence of 3303 therefore provided a map of the 3′ terminus of A3 and the boundaries of A4 and A5, which confirmed those obtained by S1 mapping.

The only ambiguity observed for S1 mapping was that for exon 6. Probes containing this exon detected two S1-protected fragments of 312 and 370 nucleotides (data not shown). Additional S1 analyses demonstrated that these two protected fragments shared a common 3′ terminus but had 5′ termini that were 58 nucleotides apart. Although we cannot eliminate the possibility that these two fragments result from different *E74A* mRNAs formed by alternative splicing, confirmatory evidence from cDNA clones exists only for the 370 nucleotide exon. All of the five *E74A* cDNA clones noted above derive from RNAs employing the upstream acceptor site. Furthermore, the sequence of the 5′ boundary of the 312 nucleotide fragment does not exhibit a good fit to a consensus Drosophila splice acceptor sequence, whereas the 5′ boundary of the 370 nucleotide fragment does. We therefore suppose that the 312 nucleotide fragment is an artifact resulting from an S1-sensitive site in the 370 nucleotide exon 6.

The sequence of the cDNA clone demonstrated that exons A3–A8 are linked and therefore derive from a single transcription unit. The linkage of A2 to A3 was demon-

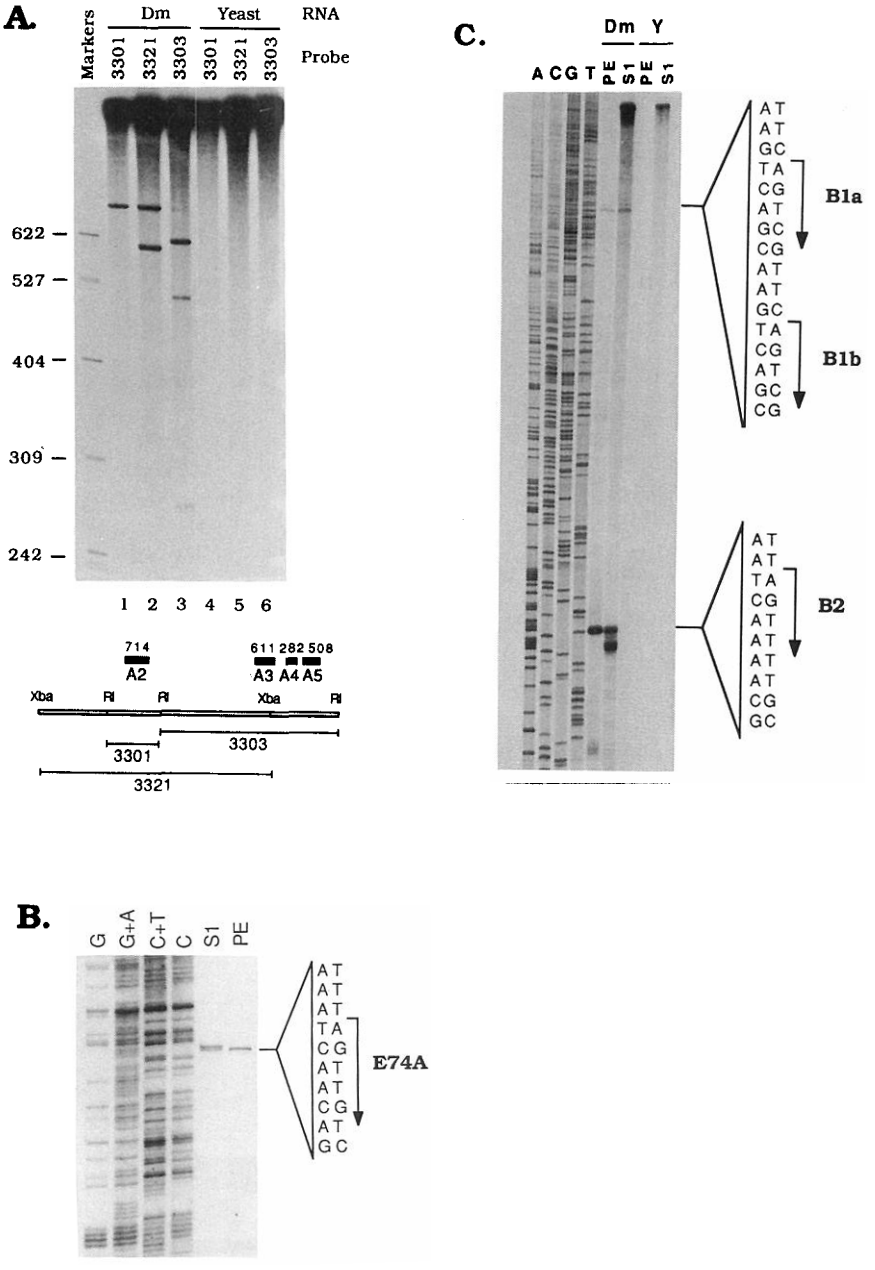

Figure 5. S1 Nuclease Protection and Primer Extension Analyses of the *E74A* and *E74B* Units

(A) S1 protection of exons A2–A5 of *E74A*. Uniformly labeled, single-stranded probes were prepared in vitro from plasmids carrying EcoRI fragments 3301 and 3303 or XbaI fragment 3321 as described in Experimental Procedures. These probes were annealed to either poly(A)$^+$ RNA isolated from second day pupae (Dm; lanes 1–3) or yeast RNA (lanes 4–6) and digested with S1 nuclease, and the products were separated by electrophoresis in urea-acrylamide gels, as described in Experimental Procedures. Size markers (in bp) are shown on the left. The schematic below the autoradiogram shows the arrangement of the probes and the exons they detect.

(B) Determination of the 5′ end of the *E74A* unit. The single-stranded DNA probes used for primer extension and S1 analyses had common 5′-labeled ends at a Sau3A site within exon A1 (273[a], Figure 6) and extend upstream to, respectively, a HhaI site within A1 (+184[a]) and an AluI site upstream of A1 (−131[a]). After annealing to poly(A)$^+$ RNA isolated from second day pupae, the S1 probe was digested with S1 nuclease and the PE probe was used as a template for reverse transcription (Experimental Procedures). The resulting 5′-labeled fragments were resolved by electrophoresis in sequencing gels to yield lanes S1 and PE. The sequence ladders to the right derive from the S1 probe (Maxam and Gilbert, 1977), and part of the resulting sequence is shown indicating the 5′ end of *E74A*.

(C) Determination of the 5′ ends of the *E74B* units. An oligonucleotide from 498[b] to 526[b] (Figure 6) was annealed to 3369 single-stranded DNA and extended with radioactive nucleotides to position 449[b] (SacII) for a primer extension probe or −243[b] (SphI) for an S1 probe. Poly(A)$^+$ RNA from second day pupae (Dm) was used for primer extension or S1 analysis, along with yeast RNA (Y) controls. The same oligonucleotide was used to prime DNA sequencing (Sanger et al., 1977). The sequences of the 5′ ends are depicted on the right. The set of bands below the full-length *E74B2* S1-protected fragment is an S1 artifact caused by occasional melting of the run of four AT bp.

A. E74A specific

```
AAGCTTGCCATTGAATTACGATATTTTCAATTTAAGTAAGTTTTTGAGTATAAAAGTGAGTATAATGTACTGCTTACTTTTTTGAGAAAGCAGAATACTT  -387a
GCTGAATGAGTTTCCGCCGCATTGTAATTGAAATGGATTTAGCGTTAATCGAACGCAAAGCAATGTTAATATTTTTTGGCCTTGAAATTAAATAAGGT  -287a
GTTGTTGGAAATAAAGTACGCAAAAACAAAAACTATTTCCCACAATTTCGACGTACAGAGGATGCACTGTAGCTCCCTACTCTCTTTGGCTCTCCCTGCA  -187a
ATTGCACTTGCATTCCGCTCTCTTAGCCTCGCTTGCGCTCTCTACGCTAACGAGCTTTGAAGTTCAGAGCGTGAACTTGAGCGTTTCTCTGCCGTTGTCG  -87a
TCGCGAGGTGGAGGTGCGCTGGTGTGCGTGAGCGGGTCTCAGTGCAGCGCAGTGCCTGTGTGCAAAAATACGCTTGTTCGCCGTTTAGTTGTCTTTTGAC  14a
TGCTGTAACGGACAGTCGCAAATTTTGCCCCAACTCTGATTCGAGCGATTCGAATGCGAAGAGGAAAACCTTCGCTTTTCAAACACACATAACAAAAACC  114a
AAACAACAAAACAAAATAGCAGCAACAAAAATCAATTGTTCAAACGTAACTGTAACTGCAGCGTTCGCGCCAACCGACAGAACGCCCGCCCGCCCACTTT  214a
TCCCGCCCACCGTCGTCGTCGCCGCCGTCGCGTCGCTAATTAAACGCTTTGCTACGATCAAAAGGGCAATCCCAAAATATCAAACAAATAAAATTGGCGT  314a
TACAAAGAGGTTCTGAAGCAGCAAAAATCTCCTGCGAAGAAATCTCGCTGTTCAAGTGGCCTTAGCGTGATGATTAAAATGTCAATTAATTAATCCAAAT  414a
ATCTGGTGGCATCTGACAGCCTTTGGTCGGCTTTTCCCGCCCTCCCCTGACCACGCCCCCGCCCCCATAAAGAAAAATGCCTGAAAGTTTATACATATTT  514a
ACACAAGAATTTTAGAAATTAAGTGATAAATCGCAAAGTAACTGATAGAAACTTCGAGGCAATAGGGTATATGATTTTTTTCTGCCGTTGCCGGAACATT  614a
ATGGATATATCGCCAATTGGATTACTTTTGCTTGAGATGAGGCCGCACAGCAAGCCGACATAGGGCGACATTGGCTGATTGACAATTCGGCTGATTGTGC  714a
AATCGGAAATTGGGAGCCATTTCCACGGAATCCTCTTGTCTTCGGGGCACCGCCCACCGACGCCCACCGCCGCCCGCCCCCAAATGCCCCAGCGCACCCT  814a
TCAAACACCATAAATGCGTGCCAAAAGCAGCAAACTGGGAACTAATGCAAATTTGGTCAATGCCATCGATATCGATGCTAATGGCAATAAAACTTTATAC  914a
ATCTATCGCATTCGCTGTAAGATAAAATCGGCCAAGTCGGCAAACTGTGAGACTGCTTCCCCGAACTCCATCATTTACCTGTGACAAGTGCGGCGGCCAC  1014a
AAAAGACCAAAATAATTAACTAAACTATATAAAATGCGCCCGAGTGCGCGGCCAAGCAAATACAACAACATCGAGTGGCACATCGGTCAATGTGCAACAA  1114a
GTGAAGCAACCCAGAAATGCAGAAAAACACTTTTTGATAACTTTACACTGCCTTACGGTTAACATAAATCGTTAATATTAACTTACAAATTACGCGTAGT  1214a
TCAATATACAAACTATACAACAAGCAAAAAACAATCAAAATTTATATATTTACAACAACAACGAAACCTTTTACAAATAAGTGAAAAACAAGAAAAAAA  1314a
ACATTCAGCAACGTCGAGGATCGTCAGCGAACAACGACAACAAAAAGCGAAAGACTCCAAAACCAAGCCAAAATAATATTAAAATTAAAAATATATATAC  1414a
AAGAGGATTTTAATATCCCTACACTGTTAAGTGCGTCGAAGAACTAACAACTTTTACACATCCACAAGGAACCTAGTATCTTCACTTTTCGTGTGATCAA  1514a
AATCCAGGAGGCGCAACAAAACTCTATACAAATTTTGAAGAACCAAACCAAAACCGTCTTAAATAAAGTGCCTAAAACATCTTTTTCGCGCGATTCTATG  1614a
AGAAAATTGTAAAGCGAACGGAGATCGAAATTTTATACCAAGTGTGATAGGTATACAAATCGAGGTATTGCAAGTTCGTATCGAGCAGTGTTTCCGCAGC  1714a
GAAACTAGAAAACTAGTGGACACTAGATACCCTTTCTTGGTTTGAGACGCGAGGAATACCTTGGACTTGGGTGGCATCAAGATTGAGCAGAACTTTCGCG  1814a
ACTGGCAGGGGTATTCAGGAAACGGCTACACGCTGGCAGTTAACAGTGTTTGCAGTGCAGGCTTTGCCCTATCAGCG                          1891a

ATGCCCTTTATCGACGATGCACTGCTCTGGTGTCCAGATAACGATGGTCGTCTTGTTGGAGGTCTAGACTTGGGAACATGCATAGCCGATGATTCTACG  1990a
METProPheIleAspAspAlaLeuLeuTrpCysProAspAsnAspGlyArgLeuValGlyGlyLeuAspLeuGlyThrCysIleAlaAspAspSerThr  33a

GCTAACGGAACGGAAAACTTGAACCCCTCGATCCAATCAGCTGGCAATCCGAACAACCCGCAGCAGAGCGTCGGTGGCGAGATCCTGGGATCCGTGGAG  2089a
AlaAsnGlyThrGluAsnLeuAsnProSerIleGlnSerAlaGlyAsnProAsnAsnProGlnGlnSerValGlyGlyGluIleLeuGlySerValGlu  66a

TCCGCCGGCAACGAACTGAATGGAGCGGCGGCCCGCAACGTGAACGTGGTGGTGGAACCGCTCTGCGGCGGCGACTCCTCCGACGAGCTGTTCCGCAGC  2188a
SerAlaGlyAsnGluLeuAsnGlyAlaAlaAlaArgAsnValAsnValValValGluProLeuCysGlyGlyAspSerSerAspGluLeuPheArgSer  99a

TTCTCCGAGAGCAACTTCGAGATCGAGAGCCTGCTGTCCGATTTGGCCACCGTCGAGGTGAAAGTCGAAAATGAGGAGAACAACAACAATGTCATCACC  2287a
PheSerGluSerAsnPheGluIleGluSerLeuLeuSerAspLeuAlaThrValGluValLysValGluAsnGluGluAsnAsnAsnAsnValIleThr  132a

GATGATGACTTCGCCAGCGTCGCTGCCGCCGTGGTGGCCAACGATGATCTGCTAGCCAAAGAGAATGCCCAGCTGAGTGCCCAGGGTCTGGTCGATTCG  2386a
AspAspAspPheAlaSerValAlaAlaAlaValValAlaAsnAspAspLeuLeuAlaLysGluAsnAlaGlnLeuSerAlaGlnGlyLeuValAspSer  165a

GTGGCCGCCAGTCTGGCCGATTCCGGCGATGCCGGTGGTCAGCAGGCGCTCTTGGCCTTCGGCTCTTCCAGCTCGGCAGCCTCGGCGATTGCAGCAGCA  2485a
ValAlaAlaSerLeuAlaAspSerGlyAspAlaGlyGlyGlnGlnAlaLeuLeuAlaPheGlySerSerSerSerAlaAlaSerAlaIleAlaAlaAla  198a

GCTGCCGCCCTCTGTGGCGATTTGATCAACAACAACAACAACAACAACAGCAACAGCAACAACAACAGCAATGGCAATGGCAATCACGGCGGCGGGGGAGGA  2584a
AlaAlaAlaLeuCysGlyAspLeuIleAsnAsnAsnAsnAsnAsnSerAsnSerAsnAsnAsnSerAsnGlyAsnGlyAsnHisGlyGlyGlyGlyGly  231a

GGAGCTAGTTCCGGCGGCGGAGTAGCCGGCGACTGTGCCACCAAGCTGGAGTACGCCCTCATGGGCGGACAGCCGCTGGCGGAGGAGCCGCGCTTCGTG  2683a
GlyAlaSerSerGlyGlyGlyValAlaGlyAspCysAlaThrLysLeuGluTyrAlaLeuMETGlyGlyGlnProLeuAlaGluGluProArgPheVal  264a

ACCAGCGCCGCCGCCAATCCCCTTCTCGTCGAGAAGCTAATGTCCAAGTGCCTGAACATTGAGAAGCGAATGGACAAGCTCAGCG                  2768a
ThrSerAlaAlaAlaAsnProLeuLeuValGluLysLeuMETSerLysCysLeuAsnIleGluLysArgMETAspLysLeuSerAsp                  293a
```

B. E74B specific

```
CTCGAGAATTGCATGCAAATGCGCATGTGGCACACAAATCCCCCAAACATGCACACACACACACACACATACCGGTGCAACAACGAAGTGGTGGGGGTGA  -159b
AGAGGAAGACGCAGCAGAGAGAGCTCCAAACATTCGCGTGCTCTCTCGCTCTCCCCCTGTCTTTGCCCCGCGATTACGTTTGTTTGCCATTTTTGCGCTC  -59b
TCTCGCAACGTTTTTCGCAACACCAGCGGCAGCGGCAGCAACGTCGTGCCGTCGATTCAGTCGTTCAGTCGGCGTAAGCGAGTCGAGCGAGCGAGCCCCA  42b
TTCAGTCCGGCGATTTGGCTCTTTTGCACCTGTGCAAGTGTGTGTTTCTCCCGCGGTGCGTGCGTCCGTGTAAGTGTGTGTGCGTGAGCGGTGGAACTGG  142b
TGCACTGGTATTGGGGCATCGGTGAGTGGGGCGCGCGCGCTTTTGCAGCCGCAGCATCGCAGTCGGCGTCTCTGCTGGCGCTGACGTCGAAGCCGACGTT  242b
TGTTGTTCGGTGTTTGCGCGTAACAAACACTCGCAGCGGTTTTTGTGCTGACTTTAGATTTTAGTTTTGCTGATACCGTAAGAGATAAATGACGTGCCGC  342b
GGCGAGCGACGAATTAGCGGCCCAAAACAGAGAAAAACTCAACAGCCAAAACACACACAGAATACCAGAGAAACAGAAGCAAGCAACAAGCAAATGTTTG  442b
ACCCGCGGGAAAACAAAATGTAATAGCTAATGTAGCAAAATGTGTTTAAGTTTAAGCCAGAAAGTTAGCATATAAAACGTCCACAATTAGTGCAAAAAAG  542b
TTCAAAACCTGATCAATTTTCGAAATCAAACAAATACCTCAAGCGTAAAGAAGTAACAATAAACCAACAAGCAAACAAACAAACAAGCAACGAACCGAGT  642b
GCAAAAACTTAGTGCAATTTCTTTAATAAGAAAATTTGAAGCCCCCAAAAAAACAGACATAATAATCATGTCTAGTGTGTAGAACGTTGCGTATACGCAA  742b
TATTGCCAAACCCATTAACGCAGACATCACGCATACGCCGCGTCGTGCAAA                                                    793b

ATGATTATGGTGCAACATTTGGTGGCCGCCTCGGCGCACAACTTCGCCAGCCAGGCGGCGGCATCGCTTGTCAATGTGTCCAGCTCCAGCTCTAGTTCC  892b
METIleMETValGlnHisLeuValAlaAlaSerAlaHisAsnPheAlaSerGlnAlaAlaAlaSerLeuValAsnValSerSerSerSerSerSerSer  33b

AGTTCGAGTTCCTCGTCCTCGCTGTCGCTGTCTTCATCCTCATCCAGCTCCTCGCTGTCCTCGGCCACGCCCACGCCGGTGGCCTCGCCCGTGACGCCC  991b
SerSerSerSerSerSerSerLeuSerLeuSerSerSerSerSerSerSerSerLeuSerSerAlaThrProThrProValAlaSerProValThrPro  66b

ACATCTCCACCGCCGGCAGCCGCCGCACCCGCAGAAGCCTCGCCGCCAGCAGGAGCTGAGTTGCAGGAGGATGGGCAGCAGGCTAAGACTCAGGAGGAT  1090b
ThrSerProProProAlaAlaAlaAlaProAlaGluAlaSerProProAlaGlyAlaGluLeuGlnGluAspGlyGlnGlnAlaLysThrGlnGluAsp  99b

CCTACTATGAAGGACCAAGACATGCTGGAAAAGACGAGGCAAGAGGTGAAGGATCCAGTTAATGTGGAGGAACCAGGTGCCATAGTTGATACCGAGTCA  1189b
ProThrMETLysAspGlnAspMETLeuGluLysThrArgGlnGluValLysAspProValAsnValGluGluProGlyAlaIleValAspThrGluSer  132b

GTGATGGCGCGGCAAAGTCCAAGTCCAGTTGCGTCCACCAAAGTCCCCGAGTCACTAGAAGAGATTTCCAACAAATCGCCGCCCGTTCAAGAGGATGAA  1288b
ValMETAlaArgGlnSerProSerProValAlaSerThrLysValProGluSerLeuGluGluIleSerAsnLysSerProProValGlnGluAspGlu  165b

GAGGAGTCTGAATCCGTAGCCTCCGACTGTCGCGAGTTCAAAGTGCTCTATAATCACCTAAGGCAGCAGCAACACCACCACTCTCCCTCCTCGCCAGAC  1387b
GluGluSerGluSerValAlaSerAspCysArgGluPheLysValLeuTyrAsnHisLeuArgGlnGlnGlnHisHisSerProSerSerProAsp  198b
```

Figure 6. Nucleotide Sequences of the *E74A* and *E74B* cDNAs and 5′-Flanking Genomic DNA, and Predicted Amino Acid Sequences of the *E74A* and *E74B* Proteins

As no complete cDNAs were obtained for either the *E74A* or *E74B* mRNAs, the sequence information presented is a composite of genomic and cDNA sequence. Nucleotides −486[a]–1579[a] were derived from genomic sequence, and nucleotides 1580[a]–6017[a] were derived from the *E74A* cDNA c17 (this cDNA also included the 490 bp intron between exons 6 and 7, which is not included in the figure). The *E74B*-specific DNA sequence was determined from genomic DNA. The predicted transcription start sites at positions 1[a], 1[b], 9[b], and 305[b] are indicated in boldface. Underlined sequences include potential translation initiation codons located in the 5′ leader sequences of *E74A* and *E74B*; the αI, αII, βI, and βII sequence homologies (Figure 9); the TATA box sequences upstream of the *E74A* and *E74B2* transcription start sites; the nucleotides (654[a], 1368[a], 1979[a], 2261[a], 2769[a], 3138[a], and 3393[a]) immediately following each intron–exon boundary; and the putative polyadenylation signal. Nucleotides and amino acids are

```
AAGACGCGCAGCACCCTGGACGATGTGTCCAAGATCCTCTGGGAGCGAAAGCAGCAGCTGCAGCGCAGCTCAGTGATTACCGCAGCACCCACTCTGCAG  1486[b]
LysThrArgSerThrLeuAspAspValSerLysIleLeuTrpGluArgLysGlnGlnLeuGlnArgSerSerValIleThrAlaAlaProThrLeuGln  231[b]

CCGCAGCAGCACCAGCAGCCCATGAGCGACATCGAGGACGAGGAGACGCTGGAGGATGTCGATGACGCGGATGCGGACGTGGAGGCAGACGCCGAGGAC  1585[b]
ProGlnGlnHisGlnGlnProMETSerAspIleGluAspGluGluThrLeuGluAspValAspAspAlaAspAlaAspValGluAlaAspAlaGluAsp  264[b]

GAGGAGCTACTGGAGCAGTACCAAAACGGCTACGACTCGCCACTGGATCTGTCCCTGGGTGGAGCCACCAGTGCAGCCGCATCAGCTGCCGCTGCCGCC  1684[b]
GluGluLeuLeuGluGlnTyrGlnAsnGlyTyrAspSerProLeuAspLeuSerLeuGlyGlyAlaThrSerAlaAlaAlaSerAlaAlaAlaAlaAla  297[b]

TCGGCGGTCAGTCGCCGACGGGGACGCACCTACTCCGGCACGGAATCCGACGACTCCGCGCAGTGCGAACGGGCGAGGATGCGACTCAAGCCGGAACGC  1783[b]
SerAlaValSerArgArgArgGlyArgThrTyrSerGlyThrGluSerAspAspSerAlaGlnCysGluArgAlaArgMETArgLeuLysProGluArg  330[b]

AAGGCGGAGAGATCGGCGGCCTACAAGAAGAGCTTGATGAAGCGATATT                                                      1832[b]
LysAlaGluArgSerAlaAlaTyrLysLysSerLeuMETLysArgTyrTyr                                                    347[b]
```

C. Common

```
ACACCGAAATTCCTATTGTCAAGCAATCGACAAGTCCGGCACCGCAGCAGCAACTGCAGCAGCAGCACCACCTGCAACAGCAGCAACAGCAGCAGCCG  2866[a],1930[b]
  ThrGluIleProIleValLysGlnSerThrSerProAlaProGlnGlnGlnLeuGlnGlnGlnHisHisLeuGlnGlnGlnGlnGlnGlnGlnPro  325[a],379[b]

CACAACGGCTCCACCTTCGCCGGCGCCACCGCTCTGCTCCACATAAAGACGGAGCAAAATACGCTGCTCACTCCGCTCCAGCTGCAGCAGCAGCAACAA  2965[a],2029[b]
HisAsnGlySerThrPheAlaGlyAlaThrAlaLeuLeuHisIleLysThrGluGlnAsnThrLeuLeuThrProLeuGlnLeuGlnGlnGlnGlnGln  358[a],412[b]

CAGCAGCAGGGTCTACACGGTGCGGCAGGCAACGGAGGCAGCAGCAACGGCAACAATGCCCACCAGCAGCAACAACCGCTGGCCATCCCACAACGCCCG  3064[a],2128[b]
GlnGlnGlnGlyLeuHisGlyAlaAlaGlyAsnGlyGlySerSerAsnGlyAsnAsnAlaHisGlnGlnGlnGlnProLeuAlaIleProGlnArgPro  391[a],445[b]

CTGCTCCACAATCTGCTTAGCGGCGGTGCCATCCACAATCCGCACCACAGAAACTACACGACGGCCACAACAGGTTCATTTCCGCCCAGTCCCGCGGAC  3163[a],2227[b]
LeuLeuHisAsnLeuLeuSerGlyGlyAlaIleHisAsnProHisHisArgAsnTyrThrThrAlaThrThrGlySerPheProProSerProAlaAsp  424[a],478[b]

AGCGGCGTCTCCGACGTGGACAGCTCCAGTTCCGGCGGCCAGCCCTGCGCCGACGAGCTGAAGGCGCGGCTCGGGATGCCACCAGCCACCTCGGCGTCG  3262[a],2326[b]
SerGlyValSerAspValAspSerSerSerSerGlyGlyGlnProCysAlaAspGluLeuLysAlaArgLeuGlyMETProProAlaThrSerAlaSer  457[a],511[b]

GCAGCGGCCGCGGCAGCAGCGGCGGCGGCGGCAGCGGCACACCTGCACACGGGCACCTTTCTGCATCCGAATTTATACCAGAACAATGCGGCAAATTCG  3361[a],2425[b]
AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaHisLeuHisThrGlyThrPheLeuHisProAsnLeuTyrGlnAsnAsnAlaAlaAsnSer  490[a],544[b]

CTGCGGAACATATGGAATCGCAGTGTTGGGGTGCCCGACAACTACTACGGAAGCAGTGGCGCCGGCAGTGGGGGCACCCAGCCAGGTGGTCCGGGAAAC  3460[a],2524[b]
LeuArgAsnIleTrpAsnArgSerValGlyValProAspAsnTyrTyrGlySerSerGlyAlaGlySerGlyGlyThrGlnProGlyGlyProGlyAsn  523[a],577[b]

CCCCAAACGCCTGGCTACCTGACAACGTCCTACTTCAACGCACCGACAGCGGCAACGGCAGCGGCGTCCCAGCGAGGAACCACAATCAATGGCTATCAT  3559[a],2623[b]
ProGlnThrProGlyTyrLeuThrThrSerTyrPheAsnAlaProThrAlaAlaThrAlaAlaAlaSerGlnArgGlyThrThrIleAsnGlyTyrHis  556[a],610[b]

TCCCTGCACCAGCAGCAGCAGCAGCAACAGCAGTCGCAGCAATCGCAGCAGCAACAGCAACTGGCCCACCAGCAACTGTCGCACCAGCAGCAGCAGGCG  3658[a],2722[b]
SerLeuHisGlnGlnGlnGlnGlnGlnGlnSerGlnGlnSerGlnGlnGlnGlnGlnLeuAlaHisGlnGlnLeuSerHisGlnGlnGlnGlnAla  589[a],643[b]

TTGCATCAGCAACTGTCGCACCAACAGCAGCAGCAGCAACAGCAACAGCAGCAACACCCGCACTCCCAACTGAACGGTCCGCATCCGCACTCACATCCG  3757[a],2821[b]
LeuHisGlnGlnLeuSerHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisProHisSerGlnLeuAsnGlyProHisProHisSerHisPro  622[a],676[b]

CACTCTCATCCGCATTCGCATCCGCACGCGGGCCAGCACACGCATTCCACAATCGCTGCAGCGGCGGCAGCGGCAGCAGCATCTGTCGTCAGCAGCAGC  3856[a],2920[b]
HisSerHisProHisSerHisProHisAlaGlyGlnHisThrHisSerThrIleAlaAlaAlaAlaAlaAlaAlaAlaAlaSerValValSerSerSer  655[a],709[b]

AGCAGCGCCGTTGCAGCGGCAGCCATGCTGAGCGCCAGTGCAGCGGCAGCAGCAACGGCGGCAGCTGCGGCGGGCGGATCCCAGAGTGTTATCCAACCG  3955[a],3019[b]
SerSerAlaValAlaAlaAlaAlaMETLeuSerAlaSerAlaAlaAlaAlaAlaThrAlaAlaAlaAlaAlaGlyGlySerGlnSerValIleGlnPro  688[a],742[b]

GCGACGTCCAGCGTCAGCTACGATCTCTCCTACATGCTGGAGCTGGGCGGATTCCAGCAGCGGAAGGCGAAGAAGCCGCGCAAACCGAAGCTGGAGATG  4054[a],3118[b]
AlaThrSerSerValSerTyrAspLeuSerTyrMETLeuGluLeuGlyGlyPheGlnGlnArgLysAlaLysLysProArgLysProLysLeuGluMET  721[a],775[b]

GGCGTAAAGCGGCGCAGCCGGGAGGGATCCACCACCTATCTGTGGGAGTTCCTCCTCAAACTGCTCCAGGATCGCGAATACTGTCCGCGTTTCATCAAG  4153[a],3217[b]
GlyValLysArgArgSerArgGluGlySerThrThrTyrLeuTrpGluPheLeuLeuLysLeuLeuGlnAspArgGluTyrCysProArgPheIleLys  754[a],808[b]

TGGACGAACCGGGAGAAGGGCGTCTTCAAGCTGGTCGACTCGAAGGCTGTGTCCCGCCTGTGGGGCATGCACAAGAACAAGCCGGACATGAACTACGAG  4252[a],3316[b]
TrpThrAsnArgGluLysGlyValPheLysLeuValAspSerLysAlaValSerArgLeuTrpGlyMETHisLysAsnLysProAspMETAsnTyrGlu  787[a],841[b]

ACGATGGGCAGGGCGCTGAGGTACTACTACCAGCGCGGCATACTGGCCAAGGTGGATGGCCAGCGGCTGGTCTACCAGTTCGTGGATGTGCCCAAGGAC  4351[a],3415[b]
ThrMETGlyArgAlaLeuArgTyrTyrTyrGlnArgGlyIleLeuAlaLysValAspGlyGlnArgLeuValTyrGlnPheValAspValProLysAsp  820[a],874[b]

ATCATCGAGATTGACTGCAACGGTGTGTGA                                                                         4381[a],3445[b]
IleIleGluIleAspCysAsnGlyVal .                                                                          829[a],883[b]

AGTGATGAGGACGAGGAGGACGACGAGGCCTGTAAATATAGTTCGATCCCAGATCGCAGACGGATCGAGCTGGAATGGCAATAGGTTTTGGGTGGGGTAT  4481[a],3545[b]
AAATCAACCGGCTAGCGGCAATAGTGGCTGAAGGCCACGCGAGATAGGGATTTCTTTGTACATGCGTTGAGTTTCTTGGTTAGCCAGCGATAAGCGTAGG  4581[a],3645[b]
TTACACACATAGCTGATAGTCCCAATACGCGTACATATACTTATATATATATATACATATATATATAAATATATTTCACACCCTGTGTCCTGTACCCTA  4681[a],3745[b]
TACATACACCCAGCTAAATCGACGGTAAATAATGATCTATTCATGGGCGTTAGTGTTGTACTTGTAACTGTAGATTTAACTGTAAACTGTAACTGTAATG  4781[a],3845[b]
AATATTGTTATTATTGGTTATTATTATTATTATTACTATTACTATTATTAATATTATAATTATTATTGCTACTGCTATGGCGGAAGGAAGAGGACGA  4881[a],3945[b]
AGGACGACCGAATTGATTCATATTGATATTCGATAGATGTGTGTAATATGTAAATGAAGTACGAACCCCTATAGCTGCAAGAATTTCTCTATCTCTATTG  4981[a],4045[b]
ATCTAATGCTAAATCTAAATGTATATGAGTATGTATATGCCTAGCCTAATGACCTTTCAACTGTCCCCTAGCCCAACCCGCTGCCCGAGGACAAGCTCAG  5081[a],4145[b]
TGCAGATGCAGATATGGGTCTATGTATGAACATATGTATACGGAGATCTGAACGATCTGCACGGGTGTCTGTGCTTGGGCTAATTGTTTTTGTTGCTTAT  5181[a],4245[b]
TTTATAAACGAAGAGAGATGGAAATTTATTATTTCATAGTTTTAGCTATTTGTACGACTACGTTTAGGATATTTATTTAAGTGTTTCTCCAGCTAATCCC  5281[a],4345[b]
CCGAAACTTTCCCCAAAACAACAAAAACACACACATCCTAACTGGAGCTAAGAACCCAGCAAACTGGTGGTCTAAACTTTCTTGTAATTATTATTTGATA  5381[a],4445[b]
AGTTTTGTATATTTACAACATCAACAGCAGAAAAGCTCACAGGATAAACCCGAAGAAGGCACTTCTTGAGCCCTCGTTTAAGCAGTCTTAGAACACCAAA  5481[a],4545[b]
CCAAACTCAAGTCTCTCGATCTGAAGTTTATTTATAAAGCATAAAGTTAAACTACCGTATTCTTGTTAACGTCACACTGCCACAAATCGAAACAAGCCTT  5581[a],4645[b]
ACATAATGGAACAGATCTGGACATTTAATCGAACCAAACCGAGATACTGACCCATAGTTTCAACAATACTGCCAACCGCAATCCTCAATGCAATATACCG  5681[a],4745[b]
ACAACTTCTACAACAATTTCCAAACACTGCCCAAAACGTGCAATCTCTTGAAAGGATACCTTGAATAGGAAAGAGAATAAAACCAACTCCTAGTTCTTTA  5781[a],4845[b]
ATCTAATTCTTTAATAGTTGCGTTCTCTAAGTACCTTTATCTATTATTATTATTATTATGTGTACGTGAAGATGAAATCTACAATATATCATTTGTTTAT  5881[a],4945[b]
GATGATTGGTGATTGATGATGCAAATAAACAAAGTGAGTCTTGAGTTCAACTCTGTTCATTGTAAACATTTTTTATGATTATTTTTTGTTTTTTAAATAA  5981[a],5045[b]
AATTATTATTTCAAAAAAAAAAAAAAAAAAAAAAA                                                                    6017[a],5081[b]
```

---

numbered in the right margin, with coordinates for the *E74A* and *E74B* products indicated by superscripts. The intron donor and acceptor sequences located at the 5′ and 3′ ends of each intron are as follows (intron names refer to the 5′ flanking exon, intron size in bp is indicated between the donor and acceptor sequences): A1, GTAAGTGA-15000-TTCTACTTGCAG; A2, GTGAGAAT-3208-TTCTTTTTTCAG; A3, GTAAGTAC-357-TTTTCTTCAAAG; A4, GTACATCC-214-ACTGTTTCCTAG; A5, GTAAGTCG-32300-CCTTCATTTCAG; B, GTAAGTGG-9000-CCTTCATTTCAG; 6, GTGATTAT-490-TCCCGT-TCGCAG; 7, GTGAGTAT-1900-TTGTTCTTGCAG. The sizes of introns A2, A3, A4, and 6 were determined by DNA sequencing; the others were determined by restriction mapping.

strated by comparing the electrophoretic mobilities observed for the S1 nuclease–protected fragments of the 3321 probe (which overlaps both exons, Figure 5A) in denaturing and nondenaturing gels (Berk and Sharp, 1978). In nondenaturing gels, the DNA fragments representing the A2 exon (714 nucleotides) and the XbaI-truncated A3 exon (591 nucleotides), which we suppose are adjacent in the RNA, should be held together by that RNA in the form of an RNA–DNA duplex with a break in the DNA strand, and should migrate equivalently to a DNA duplex of 1.305 kb. We observed that a duplex of 1.3 kb was indeed formed under nondenaturing conditions (data not shown), whereas under the denaturing conditions used in Figure 5A, the separate A2 and truncated A3 fragments were observed (lane 2).

Linkage of A1 and A2 was established by the use of an extended primer. A single-stranded DNA fragment derived from the 5′ proximal sequences of A2 was annealed to poly(A)$^+$ RNA isolated from ecdysone-induced third instar tissues and used as a primer for reverse transcription of the RNA template in the presence of ^{32}P-labeled deoxynucleotide triphosphates. This labeled cDNA was then used to probe a Southern blot of genomic HindIII restriction fragments from a 50 kb region upstream of the primer. The only fragment to hybridize was a 1.5 kb HindIII fragment (3302), whose central portion is occupied by the 0.653 kb A1 exon. The sum of the mapping data therefore demonstrates that the *E74A* transcription unit contains the indicated eight exons that are spliced to yield the ecdysone-inducible 6.0 kb *E74A* mRNA. This conclusion is supported by the subsequent isolation of a cloned cDNA consisting of all but the 23 5′ proximal nucleotides of exon A1, all of exons A2–A4, and part of exon A5. The structure of this cDNA, as determined by extensive restriction enzyme analysis, is entirely consistent with that of the deduced *E74A* mRNA. The sum of the lengths of the seven *E74A* introns is 53.5 kb (see legend to Figure 6) which, when added to the combined lengths of its eight exons, yields a total length for the *E74A* unit of 59.5 kb.

The E74A DNA Sequence

Figures 6A and 6C show the nucleotide sequence for a complete cDNA of the *E74A* mRNA and for the 486 bp immediately upstream of its transcription initiation site. The exon boundaries are indicated in the figure, and the sequences flanking these splice sites are presented in the figure legend. All of these splice donor and acceptor sequences are consistent with a consensus sequence derived from the sequences flanking more than 200 Drosophila introns (K. B., unpublished data).

A single long AUG-initiated ORF starts in A3 at +1892[a], includes 829 codons that have been translated in Figure 6, and is terminated by a UGA stop codon at +4379[a] in exon 8, which also contains the 1615 nucleotide 3′ untranslated region. We call this long ORF the *E74A* coding region and reserve consideration of the amino acid sequence it encodes to the Discussion, where it is compared with that encoded by the *E74B* coding region. Similarly, we reserve to the Discussion a consideration and comparison of the 5′ flanking sequences for the *E74A* and *E74B* transcription units. We consider here only the exceptionally long (1891 nucleotides) 5′ leader sequence of the *E74A* mRNA and the sequences surrounding the *E74A* and *E74B* transcription initiation sites.

The most striking feature of the *E74A* 5′ leader sequence, apart from its length, is the presence of 17 AUG-initiated ORFs, each of which is closed within this leader sequence. The locations of the AUG triplets initiating these ORFs are indicated in Figure 6A by underlining. All of the ORFs are short, the longest containing 56 potential codons. These characteristics raise the question of whether any of the 17 ORFs are translated or, more generally, whether the 5′ leader sequences play specific roles in the regulation of *E74A* translation, as is the case for the *GCN4* gene in yeast (Mueller and Hinnebusch, 1986).

While these questions have yet to be answered by direct experimentation, the sequence data do not encourage a positive answer. For example, certain of the 17 ORFs, which are not included within other members of this set, are extremely short; thus, the first, second, fourth, eighth, and seventeenth consist of 12, 7, 12, 9, and 4 codons, respectively. Similarly, the codon usage frequencies in the 17 ORFs, with the possible exception of the ninth, do not exhibit the bias of a typical D. melanogaster gene. In contrast, the *E47A* coding region does exhibit such a bias. (The Framescan program of Staden and McLachlan, 1982, and a standard set of Drosophila codon usages compiled by K. C. B. were used for these computations.)

We also compared the sequences immediately upstream of the AUG codons for the first eighteen ORFs with an optimal consensus sequence for eukaryotes: $CC^{A}_{G}CC$-AUG. This consensus, originally derived from sequences flanking AUG initiation codons in a very large number of genes from a variety of eukaryotes (Kozak, 1984), is consistent with results of extensive experiments in vertebrate systems designed to define the importance of particular flanking nucleotides (Kozak, 1986, 1987). The result of this comparison is that the flanking sequence upstream of the AUG initiation codon beginning the *E74A* coding region exhibits a better fit, with three matching positions out of five, than do any of the other seventeen. Cavener (1987) obtained a somewhat different consensus sequence ($^{C}_{A}AA^{A}_{C}$–AUG) from a much smaller number of genes, all from Drosophila. This consensus overlaps that of Kozak at all positions other than −2 (relative to the AUG) and lacks nucleotide specificity at −5. It is not clear whether these differences indicate a functional difference between Drosophila and other eukaryotes, because no experiments testing the functional importance of particular flanking nucleotides have been carried out in Drosophila systems, as they have in vertebrates (Kozak, 1986). It is also curious that the flanking sequence for the *E74A* initiation codon gives a poor fit to the Cavener sequence, with only one match, while all but one of the flanking sequences for the 17 leader sequence AUGs exhibit an equal or better fit.

Table 1 shows that the sequence surrounding the *E74A* transcription initiation site is quite similar to those for the three *E74B* transcription initiation sites determined in the next section. Furthermore, each of these *E74* start sites is in reasonable agreement with consensus sequences de-

Table 1. DNA Sequences Surrounding the *E74* Transcription Start Sites

E74A	TTTAGTTGT
E74B1a	ATTAGTCGT
E74B1b	TTCAGTCGG
E74B2	TTTAGTTTT
Consensus 1	ATCA$^{G}_{T}$T$^{C}_{T}$
Consensus 2	AGT$^{G}_{T}$N$^{A}_{T}$

The A nucleotide at position +4 is the start site for each *E74* transcript. Consensus 1 is from Hultmark et al. (1986); Consensus 2 is from Pirrotta et al. (1987).

rived from other Drosophila genes (Hultmark et al., 1986; Pirrotta et al., 1987).

The *E74B* Transcription Units

The *E74B* mRNAs were first observed as a single 4.9 kb band on Northern blots of total RNA from ecdysone-induced late third instar tissues. This RNA was detected with labeled probes from the 3369 and 3363 EcoRI fragments but not with probes from any of the other EcoRI fragments spanning the region from 3319 to 3361 (Figure 3). The direction of transcription of the 4.9 kb RNA (i.e., the *E74B* unit[s]), as determined with single-stranded probes, is the same as that for *E74A*. These observations suggest that the 5′ exon(s) of *E74B* is within the 3369 fragment and thus within the long A5 intron of *E74A*, while the 3′-proximal exons of *E74B* are in the 3363 fragment and could be shared with *E74A* (i.e., exons 6–8, Figure 3). The following analyses confirm this suggested organization for *E74B* and define the respective exons at the nucleotide level.

A cDNA library prepared from late third instar RNA was screened with a 3369 probe to yield two cDNA clones (C-1 and E-1), both of which also hybridize with a 3363 probe. Sequences within these clones specific to 3369 hybridized uniquely with the 4.9 kb RNA, while those specific to 3363 hybridized with both the 4.9 kb RNA and the 6.0 kb *E74A* mRNA. Determination of the 3′ terminal sequence of E-1 showed that it contained a poly(A) tail at the same addition site as that defining the 3′ terminus of exon 8. Comparison of a detailed restriction map for the 3.3 kb of E-1 cDNA lying upstream of this polyadenylation site with that deduced from the sequence for the linked exons 6–8 found in the *E74A* cDNA demonstrated that the 3′-terminal region of E-1 shares these three exons with *E74A*. Similar analyses showed that the C-1 clone has its 3′ terminus in the middle of exon 7.

Determination of the sequence immediately upstream of the 5′ end of exon 6 in these cDNAs showed that both cDNAs contain the same 5′ exon and span the same splice junction. Sequence analyses of the appropriate genomic DNA in 3369 showed that this exon (exon B in Figure 3) lies within the 3369 fragment and is located 9.0 kb upstream of exon 6, as this intervening DNA comprises intron B. Further sequence analysis indicated that the 5′ ends of C-1 and E-1 are located at $+701^{b}$ and $+1362^{b}$ in the exon B sequence given in Figure 6B.

Primer extension and S1 protection analyses (Figure 5C) identified three transcription initiation sites for the *E74B* units, located at $+1^{b}$ (B1a), $+9^{b}$ (B1b), and $+305^{b}$ (B2) in the sequence shown in Figure 6B. The fact that the B1a and B1b sites are only 8 bp apart and are at the same position in a tandem repeat of the sequence TTCAGTCG makes it unlikely that they are differently controlled. We therefore consider them as alternative initiation sites for the same transcription unit, *E74B1*. Because the B2 initiation site is 300 bp downstream of the B1 sites, we assume that it could be differentially controlled and therefore presume that it defines a separate transcription unit, *E74B2*. Although second day pupae provided the source of the poly(A)$^{+}$ RNA used in these S1 and primer extension experiments, the same result was obtained by S1 analysis using RNA templates derived from late embryos, late third instar larvae, fourth day pupae, as well as ecdysone-induced K_c tissue culture cells (data not shown). These results indicate that, like *E74A*, the *E74B* transcription initiation sites are invariant during development. As noted above, the sequences surrounding all three *E74B* sites fit the consensus sequence for such sites in Drosophila genes (Table 1). The *E74B1a*, *E74B1b*, and *E74B2* mRNAs are predicted to contain 5051, 5059, and 4755 transcribed nucleotides, respectively, plus a poly(A) tail. We should therefore expect the 4.9 kb RNA band to consist of two size classes of mRNAs with approximate lengths of 5.1 kb and 4.8 kb, respectively. Consistent with this expectation, the 4.9 kb band has recently been resolved into two bands of the expected mobilities (F. D. K. and C. S. T., unpublished data).

Each of the *E74B* mRNAs contains the same 883 codon ORF initiated by the ATG at $+794^{b}$ (Figure 6B) and continues in-frame to include the same 536 codons from exons 6–8 that comprise the 3′ portion of the *E74A* coding region. The *E74* gene therefore encodes two related proteins, as discussed in more detail in the Discussion. The 5′ leader upstream of this 883 codon *E74B* coding region contains six short ATG-initiated ORFs (ATGs are underlined in Figure 6B), which we think are unlikely candidates for translation for the same reasons given above for the 17 ORFs that precede the *E74A* coding region.

Transcripts Adjacent to the *E74* Gene

Figure 3 shows the orientation and approximate position of DNA sequences corresponding to four transcripts that lie outside of the *E74A* transcription unit. These transcripts were detected by Northern blot analyses of total RNA from ecdysone-induced and noninduced third instar tissues using appropriate probes derived from the 3319, 3362, and 3361 EcoRI restriction fragments. A pair of oppositely oriented transcripts of 1.2 kb and 3.9 kb derive from genomic sequences upstream of the 5′ end of *E74A*, with the 1.2 kb transcript mapping within 1 kb of *E74A* and transcribed in the same direction as *E74*. Neither of these RNAs increases in abundance in the presence of ecdysone, despite their proximity to the ecdysone-inducible *E74A* promoter.

A transcript with the same orientation as the 6.0 kb *E74A* mRNA, and virtually the same length, was detected by the 3362 probe downstream of the 3′ end of *E74A* (Figure 3). This RNA is distinguished from the *E74A* mRNA not only

by its insensitivity to ecdysone in third instar tissues, but also by the constancy of its abundance in whole animals throughout development. In Northern analyses, it is also generally accompanied by a minor 5.3 kb RNA band of the same orientation (Burtis, 1985). Neither of these RNAs have been further investigated. The 0.9 kb RNA detected by the 3361 probe derives from the *Jonah* gene referred to earlier in the Results (Carlson and Hogness, 1985).

Möritz et al. (1984) have reported the isolation of genomic DNA from the 74EF region by microdissection and identified two EcoRI restriction fragments as having homology to a 2.7 kb ecdysone-inducible RNA from salivary glands. Comparison of their restriction map with ours indicated that these two fragments corresponded to 3333 and the small adjacent 3335 fragment. When, however, we used 3333 as a probe for Northern analyses of total RNA isolated from whole animals at all stages of development, we observed a 2.4 kb RNA that was present at all developmental stages and showed no correlation with ecdysone pulses, unlike the transcriptional response of the *E74A* mRNA (Thummel et al., 1990). We did not, therefore, further define this transcription unit.

Discussion

The ecdysone-inducible early genes of D. melanogaster were initially defined as those responsible for the early puffs induced by ecdysone in the polytene chromosomes of late third instar salivary glands (Ashburner et al., 1974). We isolated and determined the structure of a gene that satisfies this definition. This *E74* gene occupies the locus of the 74EF early puff (Figure 2) and exhibits a complex structure consisting of a 60 kb *E74A* transcription unit inside of which are nested two smaller units, *E74B1* and *E74B2*, that employ the same polyadenylation site as *E74A* but are initiated at different sites, 300 bp apart, within the fifth of the seven *E74A* introns (Figures 3 and 6). The eight *E74A* exons comprise a 6.0 kb mRNA induced directly by ecdysone in late third instar salivary glands (Figure 4), the kinetics of this primary response mimicking that of the 74EF early puff (Thummel et al., 1990). Both *E74B* mRNAs (4.8 and 5.1 kb) consist of four exons, of which the 3′ proximal three are the same as those in the *E74A* mRNA (Figures 3 and 6). The question of the ecdysone inducibility of the *E74B* mRNAs is addressed in the accompanying paper (Thummel et al., 1990).

As a consequence of the nested organization of these transcription units, the *E74* gene encodes two related proteins (*E74A* and *E74B*) that contain the same carboxy-terminal region but differ in their amino-terminal regions (Figures 6 and 7). A similar organization has been observed for another ecdysone-inducible early gene, *E75*, which occupies 50 kb of genomic DNA at the locus of the 75B early puff shown in Figure 2 adjacent to the 74EF puff (Segraves, 1988; Segraves and Hogness, 1990). The two related proteins encoded by *E75* exhibit sequence similarities with the DNA binding and hormone binding domains of the steroid receptor superfamily and therefore appear to satisfy the prediction of the Ashburner model that early genes encode nuclear regulatory proteins (Figure 1). This may also be the case for the proteins encoded by ecdysone-responsive transcripts that derive from the locus of the 2B5 early puff (Chao and Guild, 1986; G. Guild, personal communication). In the following section, we analyze the deduced amino acid sequences of the *E74A* and *E74B* proteins for evidence that they may also satisfy this prediction of the Ashburner model.

Are the *E74* Proteins Nuclear Regulatory Proteins?

The deduced amino acid sequences for the *E74A* and *E74B* proteins consist of 829 (87.1 kd) and 883 (94.7 kd) amino acids, respectively (Figure 6). The carboxy-terminal 536 amino acids in each protein are encoded by the common exons 6, 7, and 8, whereas their amino-terminal sequences are encoded, respectively, by exons A3, A4, and A5 (293 amino acids) and by exons B1 or B2 (347 amino acids).

Figure 7 shows that the common region, which occupies almost two-thirds of each protein, is divisible into two parts according to the distribution of homopolymeric sequences and charged amino acids. The preponderance

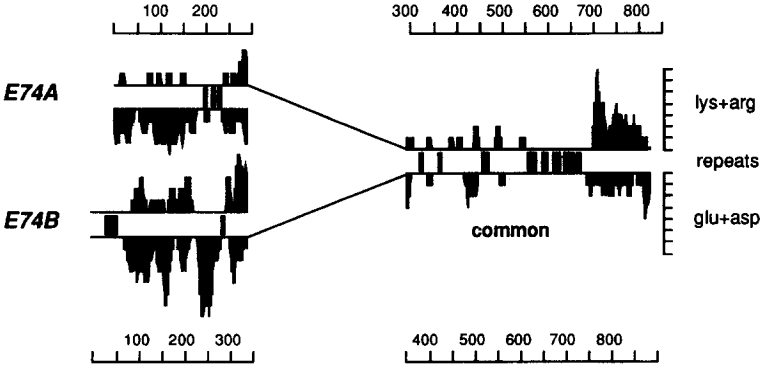

Figure 7. Distribution of Charged and Repeated Amino Acids in the *E74A* and *E74B* Polypeptides

The distribution of basic (Lys and Arg), acidic (Glu and Asp), and repeated amino acids (defined as a single amino acid reiterated five or more times) is displayed graphically. The ordinates above and below the horizontal axes represent the number of occurrences of basic or acidic residues, respectively, in sequential windows of ten amino acids, with values calculated around each successive amino acid residue in the polypeptides. The values in the extreme amino-terminal portion of the common region reflect the sequence of the *E74A*-specific polypeptide. Each division in the scale at the right side of the common region represents one residue out of a window of ten. The locations of repeated residues are indicated by the filled boxes located between the horizontal lines, with the length of each block corresponding approximately to the length of the repeats. The stippled box shows the location of the Ser-His-Pro-His repeats discussed in the text. The lengths of the *E74A*-specific, *E74B*-specific, and common regions are 293, 347, and 536 residues, respectively, and are drawn to the same scale. Amino acid coordinates for both the *E74A* and *E74B* polypeptides are indicated on the horizontal scales at the top and bottom of the figure (*E74A* coordinates at top, *E74B* coordinates at bottom).

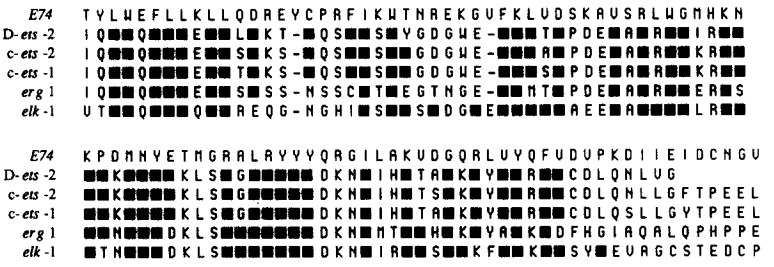

Figure 8. The *E74* Proteins Are Similar to Members of the *ets* Proto-Oncogene Superfamily

The 97 amino acids at the carboxyl terminus of the *E74* proteins are depicted on the top line. Below this are sequences from the Drosophila D-*ets*-2 (Pribyl et al., 1988), human c-*ets*-2 (Watson et al., 1985), human c-*ets*-1 (Reddy and Rao, 1988), *erg* 1 (Reddy et al., 1987), and *elk*-1 (Rao et al., 1989) proteins arranged for maximum alignment with the *E74* sequence. The closed boxes represent amino acid identities with *E74* and the dashes indicate gaps to facilitate alignment.

of basic (Lys and Arg) over acidic (Glu and Asp) amino acids among the 121 carboxy-terminal residues defines a basic domain with a computed pI of 10.6. In contrast, the remaining 415 amino acids of the common region exhibit a low density of charged amino acids without significant acidic or basic bias and are marked by a high frequency of homopolymeric sequences that consist of one Ser and several Glu repeats encoded by *opa* sequences (Wharton et al., 1985) and a few Ala repeats. In addition, this homopolymeric region contains a PRD repeat motif (Frigerio et al., 1986) consisting of His residues that alternate with either Pro or Ser residues (615[a]–630[a] in Figure 6C).

Figure 7 also shows that the *E74A*- and *E74B*-specific regions are each marked by acidic domains near the amino termini of the respective proteins (residues 1[a]–186[a] in *E74A*, pI = 3.1; residues 55[b]–280[b] in *E74B*, pI = 3.9). In *E74B*, this acidic domain is bounded at its amino terminus by a homopolymeric Ser repeat and at its carboxyl terminus by an Ala repeat, which separates the acidic domain from a short region with a basic amino acid bias (residues 302[b]–347[b], pI = 11.0). The *E74A* acidic domain is similarly bordered at its carboxyl terminus by homopolymeric sequences that consist of single repeats of Ala, Asn (*opa* encoded), and Gly (encoded by the *pen* sequence; Haynes et al., 1987) followed by a short region of charged amino acids that exhibits no significant bias.

Thus, in a first approximation, each *E74* protein is characterized by an amino-terminal acidic domain and a carboxy-terminal basic domain separated by a long homopolymeric region. The basic domain is of particular interest because it contains an 84 amino acid sequence that is 50% identical with an 82 residue sequence located near the carboxyl terminus of the protein encoded by the human c-*ets*-2 proto-oncogene (Watson et al., 1985). This region is also highly conserved in the c-*ets*-1 proto-oncogene-encoded protein (Reddy and Rao, 1988). Figure 8 shows these sequence similarities, as well as those between the *E74* sequence and the Drosophila homolog of c-*ets*-2 (D-*ets*-2, 50% identical) and two human homologs (*erg*-1, 44% identical; *elk*-1, 54% identical).

The c-*ets*-2 gene forms part of the transforming gene of the E26 avian erythroblastosis virus (Watson et al., 1985). This gene is conserved in vertebrates (Watson et al., 1986) as well as Drosophila, the c-*ets*-2 and D-*ets*-2 proteins exhibiting a 94% identity over a 94 amino acid region near their carboxy termini (Figure 8; Pribyl et al., 1988).

Although the function of the c-*ets*-2 protein has not been precisely defined, it exhibits certain similarities to nuclear proto-oncogene proteins such as the *myc* and *fos* proteins. Thus, it is localized in the nucleus (Fujiwara et al., 1988a), is phosphorylated and protein kinase C responsive, and exhibits a short half-life (Fujiwara et al., 1988b). Furthermore, the murine homolog of the c-*ets*-2 gene, like *myc* and *fos*, is expressed primarily in proliferating tissue, and the kinetics of that expression indicates that it may play a role in the G0 to G1 transition of the cell cycle (Papas et al., 1987).

These properties of c-*ets*-2 and its murine homolog, in conjunction with the sequence similarity between the c-*ets*-2 and *E74* proteins, suggest that the *E74* proteins are nuclear regulatory proteins, as predicted by the Ashburner model. This suggestion is supported by the nuclear localization of the *E74A* protein observed by in situ labeling with *E74A*-specific antibodies (L. Boyd and C. S. T., unpublished data). Nuclear localization of proteins labeled by antibodies to epitopes common to both *E74* proteins has also been observed (W. Talbot and D. S. H., unpublished data). In addition, recent experiments have demonstrated that the *E74A* protein exhibits sequence-specific DNA binding activity (L. Urness and C. S. T., unpublished data).

The carboxy-terminal basic domain that includes the *ets* sequence homology is a reasonable candidate for the *E74* DNA binding domain. This domain also contains a sequence (Lys-Lys-Pro-Arg-Lys-Pro-Lys-Leu; residues 712[a]–719[a] in Figure 6C) that is a candidate for a nuclear localization signal on two counts: it contains the consensus four amino acid sequence (Lys-Arg/Lys-X-Arg/Lys) found in nuclear localization signals (Chelsky et al., 1989) and it bears some resemblance to the sequence (Pro-Lys-Lys-Lys-Arg-Lys-Val) required for nuclear localization of the SV40 T antigen (Kalderon et al., 1984).

At a similar level of speculation, studies of transcriptional activator proteins in yeast have indicated that acidic regions function as activation domains (reviewed by Ptashne, 1988), suggesting the possibility that the *E74* amino-terminal acidic domains might function in a similar manner. The *E74B* acidic domain is further distinguished by exhibiting a high frequency of proline residues, particularly at its amino terminus, where 10 out of 29 residues (34%) are prolines; in this respect it resembles the transcriptional activation domain of CTF/NF-1 (Mermod et al., 1989).

The homopolymeric repeats clustered in the region

separating the terminal acidic and basic domains are frequently, although by no means uniquely, found in nuclear regulatory proteins involved in the determination of developmental pathways in Drosophila (Burtis, 1985; Scott and Carroll, 1987). In spite of this correlation, the functional significance of these repeats has not yet been defined. Indeed, they may not have specific functions. Rather, the proteins in which they appear may be unusual in that they contain combinations of discrete functional domains linked by "spacers" that have little other than joining functions, which allow the functional domains to act upon their substrates in a coordinated manner. The insertion of repeats—continuously generated by mechanisms such as slippage during DNA replication (Tautz et al., 1986)—outside of the functional domains might then have little effect, allowing their preferential accumulation in such spacers.

According to this view, the central region connecting the acidic and basic domains might, to first approximation, function as such a spacer. If this is the case, one would expect the central region, and particularly its homopolymeric repeats, to exhibit a considerably lower level of evolutionary conservation than the terminal acidic and basic domains. Sequence analyses of the *E74* genes in different Drosophila species are consistent with this expectation (C. W. J., unpublished data).

It should be emphasized that this view does not exclude the possibility that certain repeated amino acid sequences, such as the PRD repeat, have specific functions. This repeat appears to be highly correlated with Drosophila nuclear regulatory proteins. Originally defined by Frigerio et al. (1986) as a sequence of alternating His and Pro residues in the *paired* gene, they also found it in *bicoid*, one of 11 additional "PRD genes" that they isolated by low stringency screening. Indeed, the clone they identified as PRD gene 8 is probably from *E74* according to its location in polytene chromosomes, its EcoRI restriction map, and the presence of an *opa* repeat in the same EcoRI fragment containing the PRD repeat. In this *E74* repeat, the His residues alternate with either a Pro or Ser residue, which occur with equal frequency (Figures 6 and 7). We examined a Drosophila data base for amino acid sequences similar to the *E74* PRD motif and identified three additional genes with similar motifs: *Deformed*, which contains the sequences HSHSHSHTHS and HPHSHPHAHPH (Regulski et al., 1987), *fushi tarazu*, with sequences SHSHSH and HPHS (Laughon and Scott, 1984), and *doublesex*, with SHPHSH (Burtis and Baker, 1989). This motif does not occur commonly in known protein sequences; a search of the Protein Identification Resource data bank for the sequence His-X-His-X-His, where X could be either serine or proline, found only four non-Drosophila examples. As with the homopolymeric sequences, it remains, however, an open question whether the PRD repeats exhibit specific functions.

The *E74* Promoter Sequences and Ecdysone-Responsive Elements

A computer search of genomic DNA sequences encompassing the *E74A* A1 exon (extending to −486, Figure 6A), as well as genomic sequences containing exons A2–A5 and associated introns (i.e., the EcoRI fragments 3301 and 3303, Figure 3), failed to reveal significant sequence similarities with that of the 23 bp ecdysone-responsive element defined by Riddihough and Pelham (1987) and referred to in the Introduction. This result is consistent with the absence of a functional ecdysone-responsive element near the *E74A* initiation site, as determined by P element germline transformations; indeed, no responsive elements were detected by this test within 7.3 kb of 5′ flanking DNA and genomic DNA encompassing exons A1–A3, excluding most of the 15 kb A1 intron (C. S. T., unpublished data). Nor could a potential ecdysone-responsive sequence be found within the B exons and the flanking 5′ sequences (Figure 6B) by the computer search noted above. Functional tests for such an element in the vicinity of the *E74B* transcription initiation sites have not yet been carried out. However, sites that bind the ecdysone receptor encoded by the Drosophila *EcR* gene (see Introduction) have recently been detected in a region 7.7–13.8 kb downstream of the *E74A* initiation site (W. Talbot and D. S. H., unpublished data). The sum of the existing data therefore suggests that the ecdysone-responsive element(s) for the *E74A* promoter lies downstream of the start site, within the long A1 intron.

The *E74A* and *E74B* promoter sequences shown in Figures 6A and 6B are not, however, without interest, even with respect to the TATA boxes. A sequence similar to the canonical TATA box ($TATA^{A}_{T}A^{A}_{T}$; Goldberg, 1981; Breathnach and Chambon, 1981) is located between -17^{a} and -24^{a} in the *E74A* promoter. Unexpectedly, an examination of the sequences in the *E74A* promoter required for in vitro transcription indicated that this TATA box–like sequence did not exhibit the normal TATA box functions—with respect either to the accuracy of transcription initiation or to transcriptional activity. Rather, a TATA box–like sequence located downstream of the transcription initiation site at $+34^{a}$ to $+40^{a}$ may be required for such functions (Thummel, 1989). No TATA box-like sequence was observed within 100 bp of the *E74B1* transcription initiation site, although such a sequence was found at an appropriate position (36 bp) upstream of the *E74B2* initiation site ($+263^{b}$ to $+269^{b}$, Figure 6B).

Four additional 5′ flanking sequences, two upstream of *E74A* (αI, αII) and two upstream of *E74B2* (βI, βII), are also of interest because of their sequence similarity to one another and to a sequence located 159 bp upstream of another ecdysone-inducible transcription unit, *E75A*, that is responsible for the 75B early puff (Segraves, 1988; Segraves and Hogness, 1990). These sequences are aligned in Figure 9 to show their regions of identity, and the locations of the *E74* sequences are indicated by underlining in Figure 6.

Although our attention was first attracted to these sequences by the 13/14 identity between the 3′-proximal regions of αII and the *E75A* sequence (Burtis, 1985), subsequent DNAase I protection studies showed that the αI and αII sequences bind a purified protein encoded by the Drosophila *zeste* gene (Thummel, 1989). These sequences, as well as the βII sequence, which exhibits a

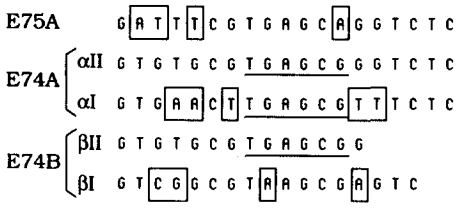

Figure 9. Sequence Similarities within the 5′ Flanking Regions of *E74A*, *E74B*, and *E75A*

Five 5′ flanking sequences are aligned to show their regions of identity: one from 159 bp upstream of the *E75A* start site (Segraves, 1988; Segraves and Hogness, 1990); two upstream of *E74A*, αI (−98[a] to −116[a]) and αII (−47[a] to −65[a]); and two upstream of *E74B2*, βI (+10[b] to +26[b]), and βII (+120[b] to +133[b]). These sequences are also underlined in Figure 6. The boxes mark nucleotides that do not match the other aligned sequences. The consensus *zeste* binding sequences (Benson and Pirrotta, 1988) are underlined.

striking 14/14 identity with the 5′-proximal portion of αII, contain a hexanucleotide (TGAGCG) that conforms to a consensus binding sequence for the *zeste* protein deduced by Benson and Pirrotta (1988). Whether the *zeste* protein is relevant to the regulation of *E74A* in vivo, however, is not known.

Experimental Procedures

General Methods

Standard recombinant DNA procedures (i.e., plasmid preparation, bacterial transformation, restriction enzyme digestions, nick translation, agarose gel electrophoresis, blot hybridization, etc.) were done essentially as described by Maniatis et al. (1982). Lambda growth, purification, and DNA extraction were performed according to Davis et al. (1980) and Huynh et al. (1985). Two lambda libraries of Drosophila genomic DNA were used: a library of Oregon R DNA (Meyerowitz, Kemp, and D. S. H., unpublished data) in lambda Sep6 (Davis et al., 1980) and a library of Canton S DNA in Charon 4 (Maniatis et al., 1978). Several cDNA libraries were screened, including an early pupal cDNA library in λgt10 (Poole et al., 1985; from which five 3′ proximal *E74* clones were isolated, including c17), a cDNA library in λgt10 constructed using RNA isolated from larval organs treated with ecdysone and cycloheximide (C. S. T., unpublished data; Huynh et al., 1985; from which the *E74A* cDNA clone carrying exons A1–A5 was isolated), and a late larval cDNA library in lambda ZAP (Stratagene; C. S. T., unpublished data; yielding the two *E74B* cDNAs, C-1 and E-1).

Drosophila growth and staging, larval organ culture, extraction of RNA, and Northern blotting were performed as described (Thummel et al., 1990).

Differential Hybridization with Labeled cDNA

Lambda DNA from the *E74* walk was digested with restriction enzymes, fractionated by agarose gel electrophoresis, and transferred onto nitrocellulose membranes by the bidirectional blotting method of Smith and Summers (1980). The genomic DNA was hybridized with cDNA probes prepared using 2 μg of poly(A)$^+$ RNA isolated from either uninduced or ecdysone-induced cultured larval organs. The RNA was added to a reaction mix containing 100 μCi of [α-^{32}P]dCTP (800 Ci/mmol; Amersham), 0.5 mM each dGTP, dATP, and dTTP, 80 mM Tris-HCl (pH 8.3 at 42°C), 10 mM $MgCl_2$, 70 mM KCl, 0.1 mM DTT, 1 μg/μl hexameric oligodeoxynucleotides (Pharmacia), and 10 U of reverse transcriptase (Seikagaku America). The reaction was allowed to proceed for 30 min at 42°C, heated to 60°C for 10 min in 20 mM EDTA, 0.1 N NaOH, and chromatographed on P60 (Biorad) to remove unincorporated deoxynucleotides.

Synthesis of Single-Stranded DNA Probes

Because of the low abundance of the *E74* transcripts, labeled single-stranded DNA probes were used for Northern blotting, S1 analysis, and primer extension. Originally, genomic DNA fragments were subcloned into a pBR322-based vector containing the φx174 RF to SS origin of replication (R. Mulligan, unpublished data). By using enzymes provided by the laboratory of A. Kornberg, it was possible to use the φx174 RF to SS replication reaction in vitro in the presence of [α-^{32}P]dCTP, cold dATP, dTTP, and dGTP, and the proteins SSB, gene A, rep, and holoenzyme III to create multiple ^{32}P-labeled single-stranded circles (Arai et al., 1980). These were used directly for Northern blot hybridization or, following gel purification, for S1 protection experiments such as those shown in Figure 5A.

More recently, probes were prepared in vitro from DNA subcloned in M13 vectors. DNA fragments were inserted into M13mp18 or M13mp19 vectors (Yanisch-Perron et al., 1985) and transformed into JM101, and single-stranded DNA was purified as described (Messing, 1983). These DNAs were used as templates for synthesizing radioactive complementary DNA. Single-stranded template (1–2 μg) was annealed to 15 ng of an oligonucleotide primer in 0.1 M NaCl, 5 mM $MgCl_2$, 50 mM Tris (pH 8.0), and 1 mM DTT for 1 hr at 37°C. The reaction was then supplemented with 50 μCi of [α-^{32}P]dCTP (Amersham), cold dATP, dTTP, and dGTP (final concentration, 150 μM each), and dCTP (final total concentration, 2 μM). After the addition of 5 U of Klenow DNA polymerase (BRL), the reaction was incubated for 10 min at 37°C. Cold dCTP was then added to 100 μM as a chase, and the reaction was continued for 5 min. The polymerase was inactivated by heating at 65°C for 5 min, the reaction was chilled, and the labeled DNA was digested with a restriction enzyme that would cut at an appropriate location downstream of the primer. The labeled DNA was ethanol precipitated, resuspended, denatured, and run on either a preparative urea-acrylamide gel or a low melting point agarose gel. The single-stranded labeled fragment (extending from the primer to the downstream restriction site) was located by autoradiography and excised from the gel, isolated by electroelution (or melting), and used as a probe.

End-labeled single-stranded DNA probes were used for some S1 and primer extension analyses (e.g., Figure 5B) as described (Weaver and Weissman, 1979). For 5′ end-labeling, DNA fragments were dephosphorylated with calf intestinal phosphatase (Boehringer Mannheim) and labeled with [γ-^{32}P]ATP as described (Maniatis et al., 1982). To label the 3′ end, the DNA was digested under controlled conditions with exonuclease III (Wu et al., 1976), and the recessed ends were filled in with ^{32}P-labeled nucleotides and Klenow DNA polymerase. The labeled DNA strands were then separated by denaturing with DMSO and running on a 5% acrylamide (50:1) gel (Maniatis et al., 1982). The bands were visualized by autoradiography and excised, DNA was extracted by electroelution, and the correct strand was identified by its ability to hybridize to an appropriate M13 clone. The single-stranded end-labeled probe was then used for either S1 or primer extension analyses.

S1 Analysis

Single-stranded DNA probes (50,000–500,000 cpm, prepared as described above) and 1–2 μg of poly(A)$^+$ RNA were ethanol precipitated together, resuspended in 2–5 μl of aqueous hybridization buffer (5× SSC, 50 mM PIPES [pH 6.4], 1 mM EDTA), and heated under mineral oil at 85°C for 5 min. They were then incubated at 50°C–60°C for 3–6 hr (the temperature was determined empirically as giving an optimum signal-to-noise ratio). S1 nuclease (1000–3000 U, Boehringer-Mannheim; 18–28 U, U. S. Biochemical) was added with 50 μl of S1 buffer (30 mM sodium acetate [pH 4.5], 0.25 M NaCl, 1 mM $ZnSO_4$, 5% glycerol), and the reaction was incubated at 15°C for 1 hr. S1 digestion was terminated by the addition of EDTA to 10 mM, and the products were ethanol precipitated with 5 μg of carrier tRNA. The precipitate was resuspended in formamide loading buffer, denatured at 90°C for 2 min, and run on a urea-acrylamide sequencing gel as described (Maniatis et al., 1982).

Primer Extension Analysis

Two different single-stranded DNA probes were used for primer extension. Either 5′ end-labeled DNA (Figure 5B) or uniformly labeled short (<100 bp) single-stranded DNA (Figure 5C) was prepared as described above. The DNA probe (10,000–100,000 cpm) was mixed with 1–2 μg of poly(A)$^+$ RNA, ethanol precipitated, resuspended in 2–5 μl of aque-

ous hybridization buffer (see S1 protocol above), covered with mineral oil, heated to 90°C for 2 min, and allowed to anneal from 5 hr to overnight at 50°C. The nucleic acids were then diluted into a primer extension mix (see differential hybridization protocol above) and reverse transcription was allowed to proceed at 42°C for 30 min. The cDNA was ethanol precipitated with 5 μg of carrier tRNA, resuspended in formamide loading buffer, denatured at 90°C for 2 min, and resolved on a urea-acrylamide sequencing gel as described (Maniatis et al., 1982).

In Situ Hybridization to Polytene Chromosomes

Tritium in situ hybridizations were done by the method of Bonner and Pardue (1976). Biotin in situ hybridizations were done by the method of Langer-Safer et al. (1982).

DNA Sequencing

DNA sequencing of *E74A* was performed by the methods of Maxam and Gilbert (1977) and Sanger et al. (1977), as modified by Biggin et al. (1983). Sequencing of *E74B* was performed using Sequenase (T7 DNA polymerase, U.S. Biochemical) and double-stranded miniprep DNA templates.

Acknowledgments

We would like to thank Ken Relloma for assistance with DNA sequencing and Larry Kauvar for the cDNA library. This work was supported by a National Science Foundation grant (to D. S. H.), a National Science Foundation Graduate Fellowship (to K. C. B.), an American Cancer Society postdoctoral fellowship (to C. S. T.), a Helen Hay Whitney postdoctoral fellowship (to C. W. J.), and the Howard Hughes Medical Institute (F. D. K. and C. S. T.).

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.

Received October 16, 1989; revised January 3, 1990.

References

Arai, K., Arai, N., Schlomai, J., and Kornberg, A. (1980). Replication of duplex DNA of phage φx174 reconstituted with purified enzymes. Proc. Natl. Acad. Sci. USA *77*, 3322–3326.

Ashburner, M., Chihara, C., Meltzer, P., and Richards, G. (1974). On the temporal control of puffing activity in polytene chromosomes. Cold Spring Harbor Symp. Quant. Biol. *38*, 655–662.

Bender, W., Spierer, P., and Hogness, D. S. (1983). Chromosomal walking and jumping to isolate DNA from the *Ace* and *rosy* loci and the bithorax complex in *Drosophila melanogaster*. J. Mol. Biol. *168*, 17–33.

Benson, M., and Pirrotta, V. (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.

Berk, A. J., and Sharp, P. A. (1978). Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. USA *75*, 1274–1278.

Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA *80*, 3963–3965.

Bonner, J. J., and Pardue, M. L. (1976). Ecdysone-stimulated RNA synthesis in imaginal discs of *Drosophila melanogaster*. Assay by *in situ* hybridization. Chromosoma *58*, 87–99.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. *50*, 349–383.

Burtis, K. C. (1985). Isolation and characterization of an ecdysone inducible gene from *Drosophila melanogaster*. Ph.D. thesis, Stanford University, Stanford, California.

Burtis, K. C., and Baker, B. S. (1989). Drosophila *doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell *56*, 997–1010.

Carlson, J. R., and Hogness, D. S. (1985). The *Jonah* genes: a new multigene family in *Drosophila melanogaster*. Dev. Biol. *108*, 341–354.

Cavener, D. R. (1987). Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. Nucl. Acids Res. *15*, 1353–1361.

Chao, A. T., and Guild, G. M. (1986). Molecular analysis of the ecdysterone-inducible 2B5 "early" puff in *Drosophila melanogaster*. EMBO J. *5*, 143–150.

Chelsky, D., Ralph, R., and Jonak, G. (1989). Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Mol. Cell. Biol. *9*, 2487–2492.

Cherbas, L., Fristrom, J. W., and O'Connor, J. D. (1984). The action of ecdysone in imaginal discs and K_c cells of *Drosophila melanogaster*. In Biosynthesis, Metabolism and Mode of Action of Invertebrate Hormones, J. Hoffmann and M. Porchet, eds. (Berlin/Heidelberg: Springer-Verlag).

Davis, R. W., Botstein, D., and Roth, J. R. (1980). Advanced Bacterial Genetics (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Evans, R. M. (1988). The steroid and thyroid hormone receptor superfamily. Science *240*, 889–895.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., and Noll, M. (1986). Structure of the segmentation gene *paired* and the Drosophila PRD gene set as part of a gene network. Cell *47*, 735–746.

Fujiwara, S., Fisher, R. J., Seth, A., Bhat, N. K., Showalter, S. D., Zweig, M., and Papas, T. S. (1988a). Characterization and localization of the products of the human homologs of the *v-ets* oncogene. Oncogene *2*, 99–103.

Fujiwara, S., Fisher, R. J., Bhat, N. K., Espina, S., and Papas, T. S. (1988b). A short-lived nuclear phosphoprotein encoded by the human *ets-2* proto-oncogene is stabilized by activation of protein kinase C. Mol. Cell. Biol. *8*, 4700–4706.

Goldberg, M. (1981). Sequence analysis of *Drosophila* histone genes. Ph.D. thesis, Stanford University, Stanford, California.

Haynes, S. R., Rebbert, M. L., Mozer, B. A., Forquignon, F., and Dawid, I. B. (1987). Pen repeat sequences are GGN clusters and encode a glycine-rich domain in a *Drosophila* cDNA homologous to the rat helix destabilizing protein. Proc. Natl. Acad. Sci. USA *84*, 1819–1823.

Hultmark, D., Klemenz, R., and Gehring, W. J. (1986). Translational and transcriptional control elements in the untranslated leader of the heat-shock gene *hsp22*. Cell *44*, 429–438.

Huynh, T. V., Young, R. A., and Davis, R. W. (1985). Constructing and screening cDNA libraries in λgt10 and λgt11. In DNA Cloning: A Practical Approach, Volume 1, D. Glover, ed. (Oxford: IRL Press), pp. 49–78.

Kalderon, D., Roberts, B. L., Richardson, W. D., and Smith, A. E. (1984). A short amino acid sequence able to specify nuclear location. Cell *39*, 499–509.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucl. Acids Res. *12*, 857–872.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44*, 283–292.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucl. Acids Res. *15*, 8125–8148.

Langer-Safer, P. R., Levine, M., and Ward, D. C. (1982). Immunological method for mapping genes on *Drosophila* polytene chromosomes. Proc. Natl. Acad. Sci. USA *79*, 4381–4385.

Laughon, A., and Scott, M. P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. Nature *310*, 25–31.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell *15*, 687–701.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Maroy, P., Dennis, R., Beckers, C., Sage, B. A., and O'Connor, J. D. (1978). Demonstration of an ecdysteroid receptor in a cultured cell line of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA *75*, 6035–6038.

Maxam, A. M., and Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA 74*, 560–564.

Mermod, N., O'Neill, E. O., Kelly, T. J., and Tjian, R. (1989). The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. Cell *58*, 741–753.

Messing, J. (1983). New M13 vectors for cloning. In Methods in Enzymology, Volume 101, R. Wu, L. Grossman, and K. Moldave, eds. (New York: Academic Press), pp. 20–78.

Möritz, T., Edström, J., and Pongs, O. (1984). Cloning of a gene localized and expressed at the ecdysteroid regulated puff 74EF in salivary glands of *Drosophila* larvae. EMBO J. *3*, 289–295.

Mueller, P. P., and Hinnebusch, A. G. (1986). Multiple upstream AUG codons mediate translational control of *GCN4*. Cell *45*, 201–207.

Papas, T. S., Bhat, N. K., Fujiwara, S., Ascione, R., and Fisher, R. J. (1987). Temporal regulation of *ets* gene expression in murine tissue. In Molecular Mechanisms in the Regulation of Cell Behavior, Modern Cell Biology, Volume 5, B. H. Satir and C. Waymouth, eds. (New York: Alan R. Liss, Inc.), pp. 123–129.

Pirrotta, V., Manet, E., Hardon, E., Bickel, S. E., and Benson, M. (1987). *Structure and sequence of the Drosophila zeste gene*. EMBO J. *6*, 791–799.

Poole, S. J., Kauvar, L. M., Drees, B., and Kornberg, T. (1985). The *engrailed* locus of Drosophila: structural analysis of an embryonic transcript. Cell *40*, 37–43.

Pribyl, L. J., Watson, D. K., McWilliams, M. J., Ascione, R., and Papas, T. S. (1988). The *Drosophila ets-2* gene: molecular structure, chromosomal localization, and developmental expression. Dev. Biol. *127*, 45–53.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature *335*, 683–689.

Rao, V. N., Huebner, K., Isobe, M., Ar-Rushdi, A., Croce, C. M., Reddy, E. S. P. (1989). *elk*, tissue-specific *ets*-related genes on chromosomes X and 14 near translocation breakpoints. Science *244*, 66–70.

Reddy, E. S, and Rao, V. N. (1988). Structure, expression and alternative splicing of the human c-*ets*-1 proto-oncogene. Oncogene Res. *3*, 239–246.

Reddy, E. S. P., Rao, V. N., and Papas, T. S. (1987). The *erg* gene: a human gene related to the *ets* oncogene. *Proc. Natl. Acad. Sci. USA 84*, 6131–6135.

Regulski, M., McGinnis, M., Chadwick, R., and McGinnis, W. (1987). Developmental and molecular analysis of *Deformed*; a homeotic gene controlling *Drosophila* head development. EMBO J. *6*, 767–777.

Richards, G. (1981). The radioimmune assay of ecdysteroid titres in *Drosophila melanogaster*. Mol. Cell. Endocrinol. *21*, 181–197.

Riddihough, G., and Pelham, R. B. (1987). An ecdysone response element in the Drosophila *hsp27* promoter. EMBO J. *6*, 3729–3734.

Sage, B. A., Tanis, M. A., and O'Connor, J. D. (1982). Characterization of ecdysteroid receptors in cytosol and naive nuclear preparations of *Drosophila* K_c cells. J. Biol. Chem. *257*, 6373–6379.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Schaltmann, K., and Pongs, O. (1982). Identification and characterization of the ecdysterone receptor in *Drosophila melanogaster* by photoaffinity labeling. Proc. Natl. Acad. Sci. USA *79*, 6–10.

Scott, M. P., and Carroll, S. B. (1987). The segmentation and homeotic gene network in early Drosophila development. Cell *51*, 689–698.

Segraves, W. A. (1988). Molecular and genetic analysis of the *E75* ecdysone-responsive gene of *Drosophila melanogaster*. Ph.D. Thesis, Stanford University, Stanford, California.

Segraves, W. A., and Hogness, D. S. (1990). The *E75* ecdysone-inducible gene responsible for the 75B early puff in *Drosophila* encodes two new members of the steroid receptor superfamily. Genes Dev., in press.

Smith, G. E., and Summers, M. D. (1980). The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzyloxymethyl-paper. Anal. Biochem. *109*, 123–129.

Staden, R., and McLachlan, A. D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucl. Acids Res. *10*, 141–156.

Tautz, D., Trick, M., and Dover, G. A. (1986). Cryptic simplicity in DNA is a major source of genetic variation. Nature *322*, 652–656.

Thummel, C. S. (1989). The Drosophila *E74* promoter contains essential sequences downstream from the start site of transcription. Genes Dev. *3*, 782–792.

Thummel, C. S., Burtis, K. C., and Hogness, D. S. (1990). Spatial and temporal patterns of *E74* transcription during Drosophila development. Cell, this issue.

Watson, D. K., McWilliams-Smith, M. J., Nunn, M. F., Duesberg, P. H., O'Brien, S. J., and Papas, T. S. (1985). The *ets* sequence from the transforming gene of avian erythroblastosis virus, E26, has unique domains on human chromosomes 11 and 21: both loci are transcriptionally active. Proc. Natl. Acad. Sci. USA *82*, 7294–7298.

Watson, D. K., McWilliams-Smith, M. J., Kozak, C., Reeves, R., Gearhart, J., Nunn, M. F., Nash, W., Fowle, J. R., III, Duesberg, P., Papas, T. S., and O'Brien, S. J. (1986). Conserved chromosomal positions of dual domains of the *ets* protooncogene in cats, mice, and humans. Proc. Natl. Acad. Sci. USA *83*, 1792–1796.

Weaver, R. F., and Weissman, C. (1979). Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucl. Acids Res. *7*, 1175–1193.

Wharton, K. A., Johansen, K. M., Xu, T., and Artavanis-Tsakonas, S. (1985). Nucleotide sequence from the neurogenic locus Notch implies a gene product that shares homology with proteins containing EGF-like repeats. Cell *43*, 567–581.

Wolfner, M. (1980). Ecdysone-responsive genes of the salivary gland of *Drosophila melanogaster*. Ph.D. thesis, Stanford University, Stanford, California.

Wu, R., Ruben, G., Siegel, B., Jay, E., Spielman, P., and Ru, C. D. (1976). Synchronous digestion of SV40 DNA by exonuclease III. Biochemistry *15*, 734–740.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene *33*, 103–119.

Yund, M. A., King, D. S., and Fristrom, J. W. (1978). Ecdysteroid receptors in imaginal discs of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA *75*, 6039–6043.