The *E75* ecdysone-inducible gene responsible for the 75B early puff in *Drosophila* encodes two new members of the steroid receptor superfamily

William A. Segraves[1] and David S. Hogness[2]

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305-5307 USA

A pulse of the steroid hormone ecdysone at the end of *Drosophila* larval development triggers coordinate changes in both larval and imaginal tissues that result in metamorphosis to the adult fly. In larval salivary glands, this pulse activates a genetic regulatory hierarchy manifested by the induction of two kinds of transcription puffs in the polytene chromosomes: a small set of "early" puffs representing a primary response to the hormone, and a complex set of "late" puffs whose delayed appearance is dependent on proteins synthesized during the primary response. We isolated a 50-kb ecdysone-inducible gene, *E75*, that occupies the early puff locus at 75B. *E75* contains two overlapping transcription units. The E75 A unit is coextensive with the *E75* gene and contains six exons: two 5′-proximal exons, A0 and A1, which are specific to this unit, and exons 2–5, which are common to both units. The E75 B unit is 20 kb long and contains five exons, a 5′-terminal exon, B1, located within the second intron of E75 A, and the common exons 2–5. Large open reading frames start within the first exon of each unit and continue into the last exon and therefore encode two different proteins. Both proteins exhibit sequence similarity to the conserved DNA-binding and hormone-binding domains of proteins in the steroid receptor superfamily. The two putative zinc fingers that characterize the DNA-binding domain are encoded by exon A1 and exon 2, so that the E75 A protein contains both fingers, whereas the E75 B protein contains only the second. Both proteins contain the same putative hormone-binding domain encoded by exon 4.

The steroid hormone 20-OH ecdysone (hereafter referred to as ecdysone) activates a genetic regulatory hierarchy in the salivary gland of the late third instar *Drosophila melanogaster* larva, manifested by the induction of a specific program of polytene chromosome puffing (Ashburner 1972). Within several minutes of ecdysone exposure, five large "early" ecdysone-inducible puffs appear. After several hours, these puffs begin to regress and a large, heterogeneous set of late ecdysone-inducible puffs begins to appear. Although the induction of the early puffs is insensitive to inhibition of protein synthesis, both the regression of the early puffs and the induction of the late puffs are blocked (Ashburner 1974). Ashburner et al. (1974) proposed that transcription of the early genes (i.e., those responsible for the early puffs) is directly induced by an ecdysone–receptor complex and that the proteins they encode act to repress early gene transcription and to induce late gene transcription.

Interaction of an ecdysone–receptor complex with the sites of the early puffs was subsequently indicated by immunofluorescent staining of polytene chromosomes with anti-ecdysone antibodies (Gronemeyer and Pongs 1980). This suggestion has been confirmed by the labeling of early puff loci with antibodies against the ecdysone receptor encoded by a recently cloned *Drosophila* gene, *EcR* [M. Koelle, W.A. Segraves, W. Talbot, M. Bender, and D.S. Hogness (in prep.); *EcR* is equivalent to the *DHR23* gene of Segraves (1988)]. Genetic data suggesting that at least some of the early genes encode proteins involved in the regulation of ecdysone response have also been reported. Thus, mutants within the 2B5 early puff region have specific effects on the salivary gland puff hierarchy (Belyaeva et al. 1981; Belyaeva and Zhimulev 1982; Zhimulev et al. 1982), and an analysis of puffing in aneuploids has demonstrated a dependence of the kinetics of both early puff regression and late puff induction on the gene dosage of a region including the 74EF and 75B early puffs (Walker and Ashburner 1981).

These studies have provided a foundation for the molecular elucidation of the genetic regulatory hierarchy

controlling ecdysone response in the late third instar salivary gland. By comparison, relatively little is known about the nature of the ecdysone response in other target tissues of the late third instar or at other stages of development that are similarly characterized by an ecdysone pulse (for a review of these stages, see Richards 1981). Through the study of several vertebrate steroid receptors and some of their target genes, the basic elements of the primary steroid response are well understood. In contrast, there has been little progress in the identification and characterization of those primary target genes encoding proteins that regulate the activity of secondary genes in regulatory hierarchies controlling the different tissue responses.

The cytogenetic criteria noted above strongly suggest that the genes responsible for the early puffs belong to this class of target genes. The isolation and analysis of these early genes therefore provide a likely means for addressing the problem of how a given steroid, in this case, ecdysone, is able to activate such a variety of developmental responses. Indeed, it is possible, and we think likely, that regulatory hierarchies in other tissues and at other stages of development will employ the same early genes, perhaps in different combinations, to activate

their responses. Two of the early genes have recently been isolated and partially characterized: the gene(s) responsible for the early puff at 2B in the X chromosome (Chao and Guild 1986), and the *E74* gene responsible for the early puff at 74 EF (Burtis 1985; K.C. Burtis, C.S. Thummel, C.W. Jones, F.D. Karim, and D.S. Hogness; C.S. Thummel, K.C. Burtis, and D.S. Hogness; both in prep.).

Here, we report on the isolation and characterization of a third early gene, *E75*, that occupies the early puff locus at 75B. The *E75* gene contains two overlapping transcription units that encode two similar, but different, proteins. Surprisingly, each of these contain sequences homologous to the DNA-binding and hormone-binding domains of proteins in the steroid receptor superfamily.

Results

Chromosomal walking yields 350 kb of genomic DNA spanning a region that includes the 75B puff

We have used the method of chromosomal walking (Bender et al. 1983) to isolate the genomic DNA encompassing the 75B puff region. The starting point for the
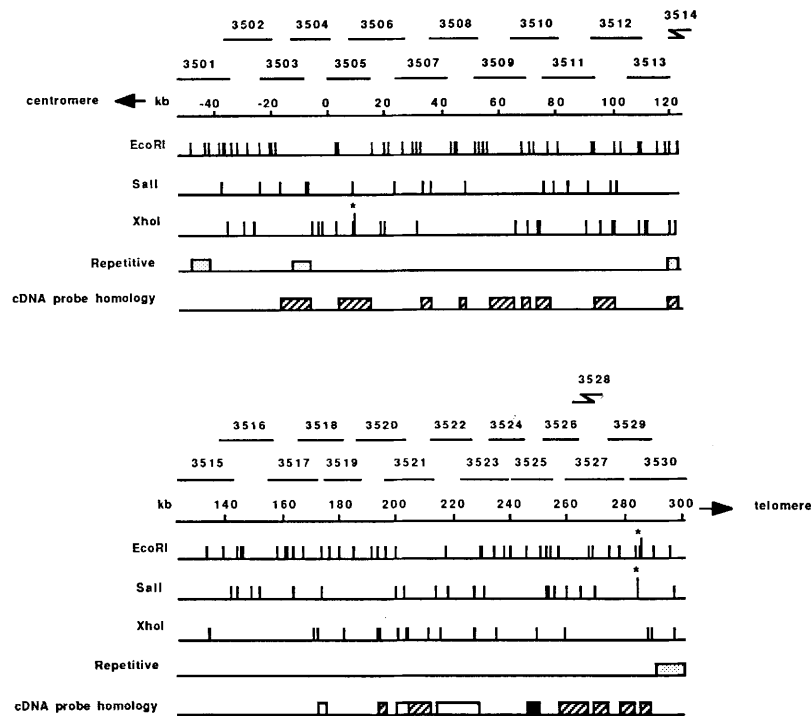
Figure 1. Composite restriction map of the 75B chromosomal walk, with genomic clones numbered 3501–3530 above it. All clones, except for λbDm3515, yDm3527, and λbDm3530, are λcDm clones isolated from the Cs Maniatis library. λbDm3515 and λbDm3530 are Or clones isolated from the Meyerowitz library, and yDm3527 is a subclone of an Or cosmid clone (see Methods). Coordinate 0 kb of the chromosomal walk is defined as the midpoint of the overlap of the λcDm3504 and λcDm3505 clones isolated by homology to the starting probe λ8253. Restriction sites for *EcoRI*, *SalI*, and *XhoI* are shown. A *XhoI* site present in λcDm3505, but absent from λcDm3506, and an *EcoRI* site and a *SalI* site present in λcDm3529, but absent from λbDm3530, are indicated by asterisks. Repetitive sequences within λcDm3514 and λcDm3528 (see text) are indicated as insertions; other regions shown to be repetitive by reverse genomic Southern analysis (see Methods) are indicated by stippled boxes. Hatched boxes indicate regions homologous to RNAs constitutively expressed in tissues of late third instar larvae. Open boxes indicate regions homologous to ecdysone-induced RNAs. The filled box indicates a region that appears to have a weak ecdysone-dependent signal in the cDNA Southern analysis but for which Northern blots failed to detect a message.

walk was a genomic clone, λ8253 (a gift of J. Burke), which had been localized by in situ hybridization to the proximal end of 75B. Isolated restriction fragments of λ8253 were used to screen a library of genomic DNA from the Canton S (C^s) strain of *D. melanogaster* (Maniatis et al. 1978). Genomic clones λcDm3504 and λcDm3505 were isolated by homology to λ8253. Figure 1 shows the position of these two cloned DNAs on the molecular map, where 0 kb on the map scale is the midpoint of the overlap between them and represents the start of the walk. The walk was then extended in both directions until ~100 kb of genomic DNA had been isolated, when the orientation of the walk was determined by in situ hybridization of the terminal segments to polytene chromosomes. Thereafter, the walk was extended in the rightward direction on the molecular map, or distally relative to the centromere. The 350 kb of genomic DNA encompassed by the walk corresponds to the chromosomal region between bands 75A6−7 and 75B11−13 (Fig. 2), as determined by in situ hybridization. This region includes the 75B puff, which appears to initiate by simultaneous decondensation of chromosomal bands 75B3−5 and then spread to surrounding bands (Semishin et al. 1985; W.A. Segraves, unpubl.). Our in situ hybridization data indicate that this puff is contained within the DNA between +120 kb and +280 kb on the molecular map (Fig. 2).

Dispersed repetitive DNA sequences within the C^s genomic clones λcDm3514 and λcDm3528 prevented continuation of the walk within this strain. Because such sequences frequently derive from transposable elements whose genomic locations are strain-specific, Bender et al. (1983) found that such a block to walking could usually be circumvented by switching to clones from another strain. Consistent with these findings, λbDm3515, isolated from a library of genomic DNA from the Oregon R (O^r) strain (Meyerowitz and Hogness 1982), and yDm3527, a subclone of a cosmid isolated from an O^r cosmid library (see Methods), did not contain repetitive DNA and allowed continuation of the chromosomal walk beyond these points.

In addition, when restriction fragment-digested clones representing the entire chromosomal walk were probed with nick-translated *Drosophila* genomic DNA, four other regions were shown to contain repetitive DNA not identified in the process of walking. These fall within λcDm3501, λcDm3503, λcDm3514 (proximal end), and λbDm3530 and are indicated in Figure 1.

Molecular mapping of deletion mutations further defines the DNA required for the 75B puff

Figure 2 shows the positions of three deletions, or deficiencies, on the molecular map of the walk and on the polytene chromosome map. Two of these deficiencies, W^{R4} and W^{R10}, were isolated as revertants of the *Wrinkled* mutation, and the third, $E75^{x48}$, is one of several lethal mutations isolated on the basis of their ability to complement W^{R4} but not W^{R10} (W.A. Segraves and D.S. Hogness, in prep.). Because the W^{R4} deletion extends from 75B8−11 to 75C5−7, and W^{R10} extends from
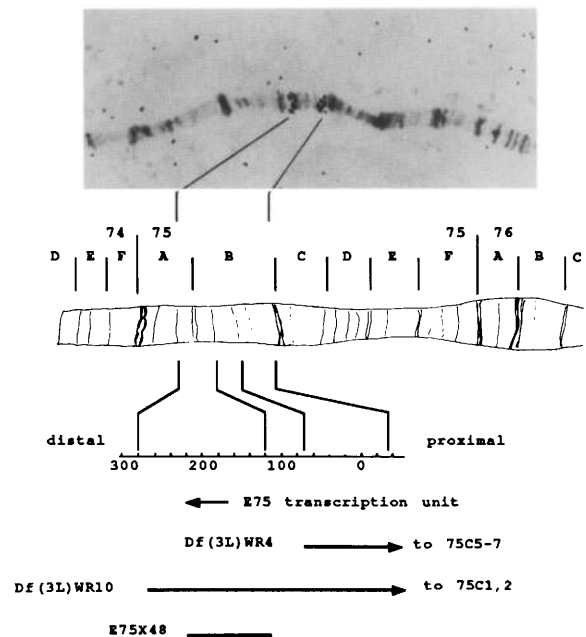
Figure 2. Correlation between molecular and cytogenetic maps of the 75B region. The positions on the molecular map of the breakpoints for the three deficiencies used for the cytogenetic definition of the 75B puff region are indicated graphically and have the following numerical values: (W^{R4}) +72 kb; (W^{R10}) ~ +270 kb; ($E75^{x48}$) +115 and +220 kb. Molecular reference points on the cytogenetic map were determined by in situ hybridization of probes near the ends of the molecular map and at other reference points, as indicated. The in situ hybridization shown was performed using fragment probes derived from λcDm3502 and λcDm3529. To maintain fidelity to cytological convention, the orientation of the molecular map is opposite that in Figs. 1, 3, and 4.

75A6−7 to 75C1,2, it was expected that $E75^{x48}$ would lie between the 75A6−7 and 75B8−11 breakpoints, or between approximately +270 kb and +72 kb, respectively, where these breakpoints have been located on the molecular map. (Note that the orientation of the molecular map in Fig. 2 is opposite that in Figs. 1, 3, and 4.) Molecular mapping of $E75^{x48}$ verified this expectation, demonstrating that it carries a deletion of 105 kb of genomic DNA between +115 kb and +220 kb.

Cytological examination of polytene chromosomes carrying these deletions revealed that the 75B puff was absent in $E75^{x48}$ and W^{R10} chromosomes but unaffected in W^{R4}. These observations demonstrate that sequences to the left (proximal to the centromere) of +72 kb on the molecular map are not required for puffing and suggests that those sequences responsible for the 75B puff are contained within the 105 kb deleted in $E75^{x48}$.

Differential screening of the walk with cDNA probes reveals a 50-kb region containing sequences homologous to ecdysone-induced transcripts

Figure 1 (bottom line) summarizes the results obtained when restriction fragments of the genomic clones were tested for their ability to hybridize with each of two cDNA probes: one derived from the RNA in ecdysone-
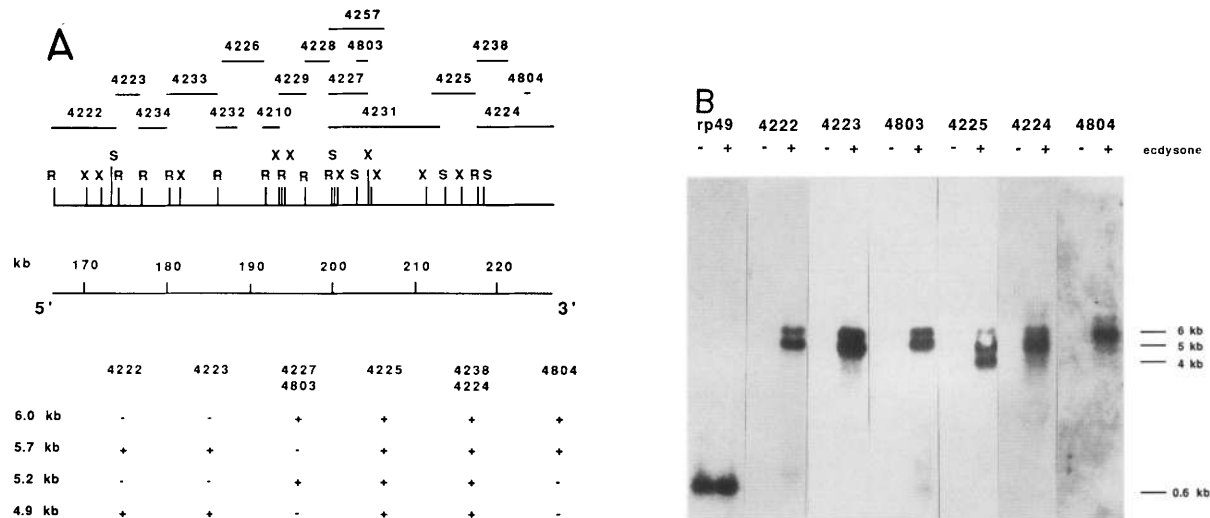
**Figure 3.** Northern blot analysis of the +166 to +226 region of the chromosomal walk. (*A*) The coordinates on the scale correspond to those on the molecular map shown in Fig. 1. A map of the *Eco*RI (R), *Sal*I (S), and *Xho*I (X) restriction sites is shown, along with the locations of the restriction fragments used to produce probes for the Northern analyses. These fragments were cloned in both orientations into the pφX plasmid vector, and ^{32}P-labeled single-stranded probes were prepared from these clones, as described in Methods. The clone for a given fragment, such as 4222, is designated in our stocks as pφDm4222, followed by a letter (A or B) indicating the orientation. The table summarizes the results shown in *B* for the indicated four size classes of ecdysone-inducible RNAs. (*B*) Autoradiographs of Northern blots hybridized with a subset of probes. Only probes for rightward transcription detect the ecdysone-inducible transcripts. The + and − lanes were loaded with 20 μg of total RNA prepared from late third instar tissues incubated, respectively, with ecdysone plus cycloheximide or with cycloheximide alone (see the text and Methods). The rp49 control probe is derived from a gene encoding a ribosomal protein (O'Connell and Rosbash 1984) and was labeled by nick translation. The 4225 and 4231 probes detect a 4-kb ecdysone-inducible RNA in addition to the four size classes shown in the table in *A*. This RNA appears to be specific to the O^r strain of *D. melanogaster*, to be colinear with the genomic DNA over most, if not all, of its length, to be AT rich, and not to contain a large open reading frame (W.A. Segraves, unpubl.).

induced cells, and the other from the RNA in noninduced cells. Two such differential screens were carried out. In the first, genomic DNA covering the entire 350-kb walk was examined with cDNA probes synthesized with reverse transcriptase from an oligo(dT) primer annealed to poly(A)$^+$ RNA. The poly(A)$^+$ RNA was prepared from total inner tissues that were mass-isolated from late third instar larvae and incubated in the presence of ecdysone plus cycloheximide, or cycloheximide alone. (See Methods. Cycloheximide was included in the expectation that higher levels of ecdysone-induced transcripts would accumulate in its presence. Fig. 6, below, shows the validity of this expectation.)

Each of the ^{32}P-labeled cDNA probes made from these two poly(A)$^+$ RNAs was applied to one of two duplicate Southern blots that contained, in addition to the genomic fragments from the walk, a control DNA consisting of sequences from the ribosomal protein 49 gene (O'Connell and Rosbash 1984), which was used to normalize the hybridization intensities of the duplicate blots. This screen revealed sequences specific to ecdysone-induced RNAs only within the λcDm3522 genomic clone that is centered at approximately +220 kb on the molecular map. Sequences homologous to constitutively expressed RNAs are indicated by the hatched boxes in Figure 1.

Because the above probes will preferentially detect sequences near the 3' termini of the RNAs, particularly in the case of long transcripts, a second differential screen
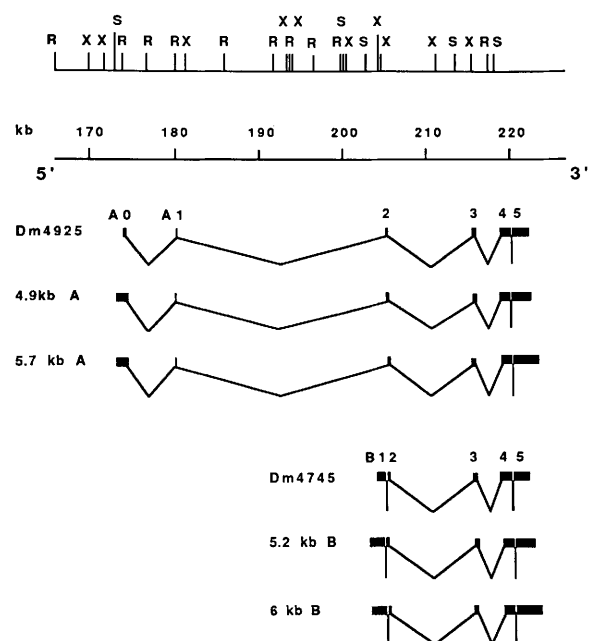
Figure 4. Molecular map of the Dm4925 and Dm4745 cDNAs and of the mRNAs that they represent, indicated by size class and the E75 transcription unit (A or B) from which they derive. The scale is the same as that in Fig. 3. Each cDNA stops short of full length at both the 5' and 3' ends. Exon 5 can be 2.0 or 2.8 kb long in both units (see text).
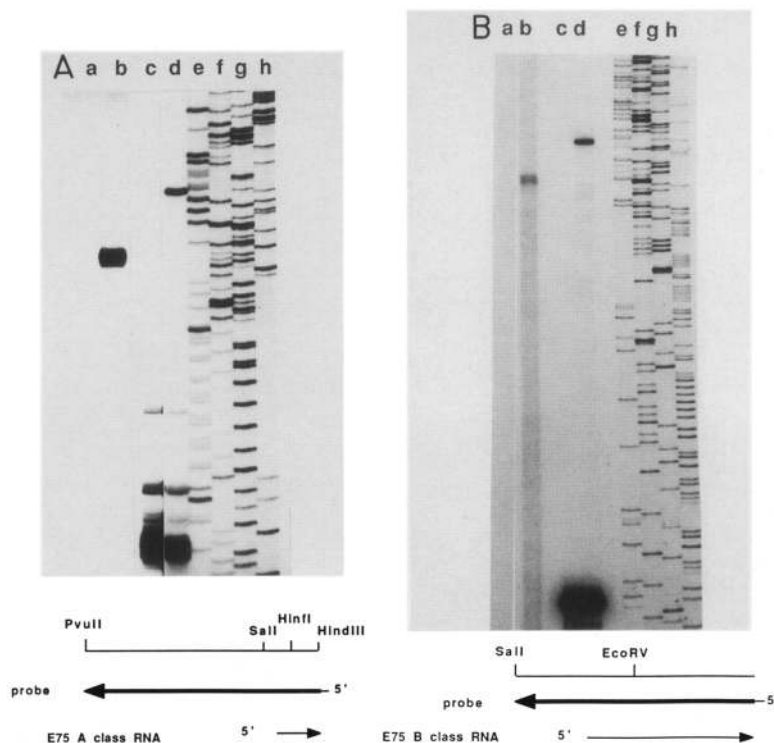
Figure 5. Mapping of the 5′ termini of the *E75* transcripts. (*A*) S1 nuclease protection and primer extension mapping of the 5′ end of the E75 A transcripts. A 700-bp *Hind*III–*Pvu*II fragment of 4222 was cloned into M13mp8 for synthesis of uniformly labeled, single-stranded probes complementary to the E75 A mRNAs. The thin extension at the 5′ end of the probe diagram represents the synthetic primer used for probe synthesis. The probe was cleaved beyond the *Pvu*II/M13mp8 junction for S1 probes and at the *Hinf*I site for primer extension analysis. (*a*) S1 nuclease protection with 1 µg of poly(A)⁺ RNA from mock-induced larval tissue. (*b*) S1 nuclease protection with 1 µg of poly(A)⁺ RNA from ecdysone-induced larval tissue. (*c*) Primer extension with no *Drosophila* RNA. (*d*) Primer extension with 1 µg of poly(A)⁺ RNA from ecdysone-induced larval tissue. (*e–h*) A, C, G, and T tracks of dideoxynucleotide chain-termination reactions performed on the same clone used for probe preparation. The difference in lengths of the S1 and primer extension products equals the length of the synthetic primer and the intervening M13 sequences present in the primer extension but not the S1 product, indicating coincident identification of the 5′ terminus, which is shown in Fig. 7A. (*B*) S1 nuclease protection and primer extension mapping of the 5′ end of the E75 B transcripts. A 420-bp fragment generated by exonuclease III digestion was cloned into pEMBL19 for synthesis of probes (see Methods). Cleavage was at the *Sal*I site for S1 probes and at the *Eco*RV site for primer extension analysis. (*a*) S1 protection with no *Drosophila* RNA. (*b–h*) as above. The difference in lengths of the S1 and primer extension products equals (as in *A*) the length of the synthetic primer and the intervening M13 sequences present in the primer extension but not the S1 product, indicating coincident identification of the 5′ terminus, which is shown in Fig. 7B.

was carried out with cDNA probes primed with random hexamers (Methods). This screen, which was restricted to the 135 kb of genomic DNA between +105 kb and +240 kb, revealed ecdysone-inducible sequences in fragments spread out over an ~50-kb region between +170 kb and +220 kb, as is indicated in Figure 1 by the open boxes. The finding that the additional sequences detected with the randomly primed probe are all to the left of those also detected by the oligo(dT) probe suggests that this region is occupied by an ecdysone-inducible unit of rightward transcription. This region is overlapped by the 105 kb of DNA deleted in *E75ˣ⁴⁸* and, as will become apparent, represents the *E75* gene.

The E75 gene contains two overlapping transcription units

Northern analysis Figure 3 shows the results of Northern blot analysis of ecdysone-induced and noninduced RNAs, prepared as described above and hybridized with strand-specific DNA probes derived from cloned restriction fragments in the 60-kb region (+166 to +226 kb) containing the *E75* gene. The ecdysone-induced RNAs hybridized only with probes detecting rightward transcription, confirming the orientation suggested in the preceding section. Four ecdysone-inducible RNAs (4.9, 5.2, 5.7, and 6.0 kb) were detected with each of the probes, 4224, 4225, and 4238, that derive from cloned fragments located near the 3′ end of the region. These four RNAs are divisible into two classes by the results obtained with probes located farther upstream. One class (the B class) consists of the 5.2- and 6.0-kb RNAs that are detected by the 4227 and 4803 probes and not by probes located farther upstream. The other, or A class, consists of the 4.9- and 5.7-kb RNAs detected by the 4222 and 4223 probes from the 5′ end of the region.

These results suggest that the A and B classes of ecdysone-inducible RNAs are initiated by different pro-
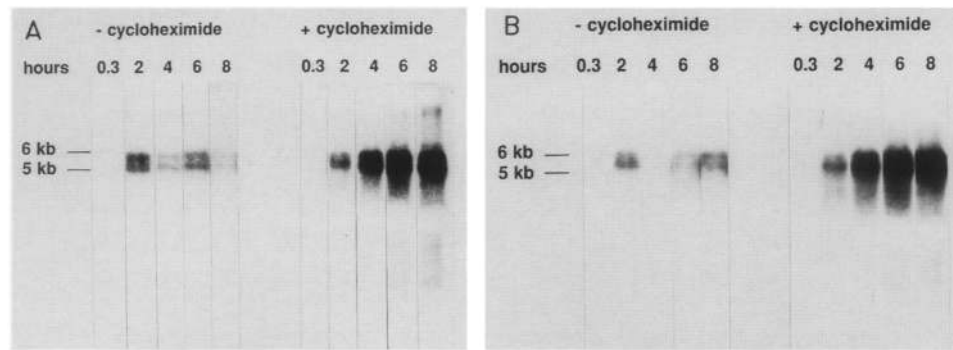
Figure 6. Time course of *E75* induction by ecdysone in salivary glands. Total RNA from 100 salivary glands per lane was used for Northern blots, and amounts of RNA loaded per lane were shown to be similar by ethidium bromide staining before transfer. The times indicate the period of exposure of the glands to ecdysone or ecdysone plus cycloheximide before preparation of the RNA. (*A*) Expression of the E75 A mRNAs, as detected with the 4222 probe (Fig. 3). Other data show a more continuous decrease in the level of E75 A mRNAs from a peak at ~2 hr; variability in the rate of disappearance of E75 A mRNAs may reflect heterogeneity in the purity or responsiveness of the isolated glands. (*B*) Expression of the E75 B mRNAs as detected with the 4803 probe (Fig. 3). The apparent reinduction of E75 B may reflect delayed induction of these RNAs in a subpopulation of cells in the isolated tissues (see text).

moters, located ~30 kb apart in the 4222 and 4227 fragments, respectively, and that the two transcription units defined by these promoters overlap in the region downstream from the B promoter. The observation that the short 4804 probe at the 3' end of the region detects only the longer member of each class further suggests that the difference in lengths between the two members of each class derives from alternative polyadenylation–cleavage sites common to both classes—a suggestion consistent with the ~0.8-kb length difference for the members of each class.

Cloned cDNA structures The organization of the *E75* gene suggested by the results of the Northern blot analysis has been confirmed by analysis of the structure of cloned cDNAs. Approximately 10^6 clones from an early pupal cDNA library (Poole et al. 1985) were screened with probes 4223, 4227, 4225, and 4224, defined in Figure 3. The 116 cDNA clones identified by this screen were analyzed by restriction digestion and hybridization to a panel of probes derived from the 60-kb (+166 to +226 kb) region shown in Figure 3. One of the clones, λDm4925, hybridized to the 4223, 4233, 4257, 4225, and 4224 fragments and is therefore a likely derivative of the A class of RNAs. A second clone, λDm4745, hybridized to 4227, 4257, 4225, and 4224, consistent with its derivation from the B class of RNAs.

The genomic regions homologous to these two cDNA clones were further localized by Southern blot analysis, and the nucleotide sequence of these regions and of both cDNA clones was determined. These sequences are given in Figure 7 and are examined in detail below. For present purposes, they have been used to define the exons in the two cDNAs and their position in the genomic DNA, with the results shown in Figure 4. The two cDNAs share four exons, labeled 2, 3, 4, and 5. The λDm4925 cDNA contains two additional exons (A0 and A1) from the 5' end of the region, which account for its hybridization with probes 4223 and 4233 and indicate

that it derives from a class A mRNA. Similarly, the λDm4745 cDNA contains an additional exon (B1) that accounts for its hybridization by the 4227 probe and indicates that it derives from a class B mRNA. Figure 4 also shows the structure of each member of class A and class B mRNAs, as deduced from the cDNA structures and the results of 5'- and 3'-terminal sequence determinations given in the next section.

Terminal sequence analysis The observation that class A mRNAs hybridize with the 4222 probe (Fig. 3) but that the λDm4925 cDNA does not indicates that this cDNA carries a truncated A0 exon and/or that the 4222 fragment contains additional class A exons upstream of A0. S1 nuclease protection experiments with the 4222 and 4223 probes and a probe spanning the 4222/4223 junction demonstrated that this region contains a single class A exon (A0) of 1.1 kb, which is truncated in the cDNA (data not shown). Fine-structure S1 and primer extension analyses were then used to determine the 5'-terminal nucleotide of this exon and to test whether it coincides with the 5' end of the class A mRNAs. Figure 5A shows that the two methods identify the same 5'-terminal nucleotide, indicating that the 5' end of the A0 exon and the class A mRNAs are coincident. This nucleotide is identified as the mRNA start site in Figure 7A.

Low-resolution S1 nuclease protection experiments, using the 4257 probe (Fig. 3) and probes derived from subclones of this fragment, demonstrated that the B1 exon is 1.5 kb in length and is truncated in the λDm4745 cDNA (data not shown). Figure 5B shows that high-resolution S1 and primer extension mapping of the 5' terminus of the B1 exon and the class B mRNAs, respectively, identify the same nucleotide, which is indicated in Figure 7B.

The 3' ends of the class A and B transcripts cannot be determined from their respective cloned cDNAs, λDm4925 and λDm4745, because their sequences reveal

A

```
ACTTACTAGTGAAAAACATGATAATAAACAACTTGCCAAAAAAAATCCAATGAAATTGACACTTATGTTAAAAAAATAGGTGAGATTGTAACCGTTGATGTACACTTACGAAGTACGTAA

                                                                                                                     -72
CAAGTTCATGAACTGATTTCGTGAGCAGGTCTCTCCATAATCGCCGTATCTGTGGGATCGCGCGCTCCTGCTCGCACTCGCTGGGTGGATGGCAGCACATGTTCGAAGTGCGAGAGAGTG

                                                                                                                      49
CAAAGCGGAGAGCGCCGACGTCGACGCCGAAAAAACTGAACAAGATCCGCCGCGAATGTTGATTTTCCTTTCATTGACTAACTGCCACTCGCAGCGCGCAGATCGTCGGCTCCGCTTGTT
                                                                 ━━━➤
                                                                  mRNA start site
                                                                                                                     169
CCGTTCCGTTCGTTTCGTTTCGTTTCGTTCGATCTACTTCGAGTCGCGAGTTTTAAGCAGTGTAGTGAGTGCCCCGTGAAAAGGATAACCCAAAAAGTGATTTCTACTATTTTCCAATAG

                                                                                                                     289
TTTTTATCAGTGTGAAGAAAACATGTAAACTTGGCTCAAAAAGGGCTTTAAAAGATACAAAGCTTCAATGCGAAGGATAAAATAATATCGCACCAGTGCTTCAAAAACCAAAACTATGCC

                                                                                                                     409
TAAGGCTGGAAATTTAAATTAAAATTTTTTTAATAAATATTCCAAAAATATTGCCCCTGAAAAGTGTTGATAAACCCCCAACCGAGCAAAATGTTAATGTCCGCGGACAGTTCAGATAGC
                                                                                        METLeuMETSerAlaAspSerSerAspSer
                                                                                                                     529
GCCAAGACTTCTGTGATCTGCAGCACGGTGAGTGCCAGCATGCTAGCACCACCAGCTCCAGAACAGCCCAGCACCACAGCACCACCCATTTTGGGGGTAACAGGTCGATCTCACCTGGAA
AlaLysThrSerValIleCysSerThrValSerAlaSerMETLeuAlaProProAlaProGluGlnProSerThrThrAlaProProIleLeuGlyValThrGlyArgSerHisLeuGlu
11
                                                                                                                     649
AATGCCCTGAAACTACCGCCAAACACAAGTGTTTCGGCTTACTACCAGCACAACAGCAAGCTGGGCATGGGCCAGAATTACAATCCGGAATTCAGGAGCCTGGTAGCACCTGTCACAGAT
AsnAlaLeuLysLeuProProAsnThrSerValSerAlaTyrTyrGlnHisAsnSerLysLeuGlyMETGlyGlnAsnTyrAsnProGluPheArgSerLeuValAlaProValThrAsp
51
                                                                                                                     769
CTGGATACTGTGCCACCCACAGGTGTGACCATGGCGAGTTCTTCGAATTCTCCCAACTCCTCCGTCAAGCTGCCCCACAGCGGCGTGATCTTTGTCAGCAAATCGAGTGCCGTCAGCACC
LeuAspThrValProProThrGlyValThrMETAlaSerSerSerAsnSerProAsnSerSerValLysLeuProHisSerGlyValIlePheValSerLysSerSerAlaValSerThr
91
                                                                                                                     889
ACCGATGGTCCCACTGCAGTGTTGCAACAGCAGCAGCCGCAGCAGCAAATGCCCCAGCACTTCGAGTCCCTGCCCCACCACCACCCCCAGCAGGAACACCAGCCACAGCAGCAGCAGCAA
ThrAspGlyProThrAlaValLeuGlnGlnGlnGlnProGlnGlnGlnMETProGlnHisPheGluSerLeuProHisHisHisProGlnGlnGluHisGlnProGlnGlnGlnGlnGln
131
                                                                                                                     1009
CAACATCACCTTCAGCACCACCCACATCCACATGTGATGTATCCGCACGGATATCAGCAGGCCAATCTGCACCACTCGGGTGGTATTGCTGTGGTTCCGGCGGATTCGCGTCCCCAGACT
GlnHisHisLeuGlnHisHisProHisProHisValMETTyrProHisGlyTyrGlnGlnAlaAsnLeuHisHisSerGlyGlyIleAlaValValProAlaAspSerArgProGlnThr
171
                                                                                                                     1099
CCCGAGTACATCAAGTCCTACCCAGTTATGGATACAACTGTGGCTAGTTCGGTAAAGGGGGAACCAGAACTCAACATAGgtgagttgtg.intron 1.ttctttgcagAATTCGATGGC
ProGluTyrIleLysSerTyrProValMETAspThrThrValAlaSerSerValLysGlyGluProGluLeuAsnIleG                              luPheAspGly
211
                                                                                                                     1189
ACCACAGTGCTGTGCCGCGTTTGCGGGGATAAGGCCTCCGGTTTCCATTACGGCGTGCATTCCTGCGAGGGTTGCAAGgtaagttcgt.intron 2.atcgtttcagGGATTCTTCCGC
ThrThrValLeuCysArgValCysGlyAspLysAlaSerGlyPheHisTyrGlyValHisSerCysGluCysLys                                GlyPhePheArg
241
                                                                                                                     1309
CGCTCCATCCAGCAAAAGATCCAGTATCGCCCGTGCACCAAGAATCAGCAGTGCAGCATTCTGCGCATCAATCGCAATCGTTGTCAATATTGCCGCCTGAAAAAGTGCATTGCCGTGGGC
ArgSerIleGlnGlnLysIleGlnTyrArgProCysThrLysAsnGlnGlnCysSerIleLeuArgIleAsnArgAsnArgCysGlnTyrCysArgLeuLysLysCysIleAlaValGly
271
                                                                                                                     1399
ATGAGTCGCGATGgtgagtacct.intron 3.ccaattgcagCTGTGCGTTTTGGACGGTGCCGAAGCGCGAAAAGGCGCGTATCCTGGCGGCCATGCAACAGAGCACCCAGAATCGC
METSerArgAspA                         laValArgPheGlyArgValProLysArgGluLysAlaArgIleLeuAlaAlaAlaMETGlnGlnSerThrGlnAsnArg
311
                                                                                                                     1519
GGCCAGCAGCGAGCCCTCGCCACCGAGCTGGATGACCAGCCACGGCCTCCTCGCCGCCGTGCTGCGCGCCCACCTCGAGACCTGTGAGTTCACCAAGGAGAAGGTCTCGGCGATGCGGCAG
GlyGlnGlnArgAlaLeuAlaThrGluLeuAspAspGlnProArgLeuLeuAlaAlaValLeuArgAlaHisLeuGluThrCysGluPheThrLysGluLysValSerAlaMETArgGln
341
                                                                                                                     1609
CGGGCGCGGGATTGCCCCTCCTACTCCATGCCCACACTTCTGgtaagtctca.intron 4.atttcttcagGCCTGTCCGCTGAACCCCGCCCCTGAACTGCAATCGGAGCAGGAGTTC
ArgAlaArgAspCysProSerTyrSerMETProThrLeuLeu                                   AlaCysProLeuAsnProAlaProGluLeuGlnSerGluGlnGluPhe
381
                                                                                                                     1729
TCGCAGCGTTTCGCCCACGTAATTCGCGGCGTGATCGACTTTGCCGGCATGATTCCCGGCTTCCAGCTGCTCACCCAGGACGATAAGTTCACGCTCCTGAAGGCGGGACTCTTCGACGCC
SerGlnArgPheAlaHisValIleArgGlyValIleAspPheAlaGlyMETIleProGlyPheGlnLeuLeuThrGlnAspAspLysPheThrLeuLeuLysAlaGlyLeuPheAspAla
411
                                                                                                                     1849
CTGTTTGTGCGCCTGATCTGCATGTTTGACTCGTCGATAAACTCAATCATCTGTCTAAATGGCCAGGTGATGCGACGGGATGCGATCCAGAACGGAGCCAATGCCCGCTTCCTGGTGGAC
LeuPheValArgLeuIleCysMETPheAspSerSerIleAsnSerIleIleCysLeuAsnGlyGlnValMETArgArgAspAlaIleGlnAsnGlyAlaAsnAlaArgPheLeuValAsp
451
                                                                                                                     1969
TCCACCTTCAATTTCGCGGAGCGCATGAACTCGATGAACCTGACAGATGCCGAGATAGGCCTGTTCTGCGCCATCGTTCTGATTACGCCGGATCGCCCCGGTTTGCGCAACCTGGAGCTG
SerThrPheAsnPheAlaGluArgMETAsnSerMETAsnLeuThrAspAlaGluIleGlyLeuPheCysAlaIleValLeuIleThrProAspArgProGlyLeuArgAsnLeuGluLeu
491
                                                                                                                     2089
ATCGAGAAGATGTACTCGCGACTCAAGGGCTGCCTGCAGTACATTGTCGCCCAGAATAGGCCCGATCAGCCCGAGTTCCTGGCCAAGTTGCTGGAGACGATGCCCGATCTGCGCACCCTG
IleGluLysMETTyrSerArgLeuLysGlyCysLeuGlnTyrIleValAlaGlnAsnArgProAspGlnProGluPheLeuAlaLysLeuLeuGluThrMETProAspLeuArgThrLeu
531
                                                                                                                     2209
AGCACCCTGCACACCGAGAAACTGGTAGTTTTCCGCACCGAGCACAAGGAGCTGCTGCGCCAGCAGATGTGGTCCATGGAGGACGGCAACAACAGCGATGGCCAGCAGAACAAGTCGCCC
SerThrLeuHisThrGluLysLeuValValPheArgThrGluHisLysGluLeuLeuArgGlnGlnMETTrpSerMETGluAspGlyAsnAsnSerAspGlyGlnGlnAsnLysSerPro
571
                                                                                                                     2329
TCGGGCAGCTGGGCGGATGCCATGGACGTGGAGGCGGCCAAGAGTCCGCTTGGCTCGGTATCGAGCACTGAGTCCGCCGACCTGGACTACGGCAGTCCGAGCAGTTCGCAGCCACAGGGC
SerGlySerTrpAlaAspAlaMETAspValGluAlaAlaLysSerProLeuGlySerValSerSerThrGluSerAlaAspLeuAspTyrGlySerProSerSerSerGlnProGlnGly
611
                                                                                                                     2449
GTGTCTCTGCCCCTCGCCGCCTCAGCAACAGCCCTCGGCTCTGCCAGCTCGGCTCCTCTGCTGGCGGCCACCCTCTCCGGAGGATGTCCCCTGCGCAACCGGGCCAATTCCGGCTCCAGC
ValSerLeuProSerProProGlnGlnGlnProSerAlaLeuAlaSerSerAlaProLeuLeuAlaAlaThrLeuSerGlyGlyCysProLeuArgAsnArgAlaAsnSerGlySerSer
651
                                                                                                                     2569
GGTGACTCCGGAGCAGCTGAGATGGATATCGTTGGCTCGCACGCACATCTCACCCAGAACGGGCTGACAATCACGCCGATTGTGCGACACCAGCAGCAGCAACAACAGCAGCAGCAGATC
GlyAspSerGlyAlaAlaGluMETAspIleValGlySerHisAlaHisLeuThrGlnAsnGlyLeuThrIleThrProIleValArgHisGlnGlnGlnGlnGlnGlnGlnGlnGlnIle
691
                                                                                                                     2659
Ggtgagtaatc.intron 5.tttcttacagGAATACTCAATAATGCGCATTCCCGCAACTTGAATGGGGGACACGCGATGTGCCAGCAACAGCAGCAGCACCCACAACTGCACCACCAC
G                                    lyIleLeuAsnAsnAlaHisSerArgAsnLeuAsnGlyGlyHisAlaMETCysGlnGlnGlnGlnGlnHisProGlnLeuHisHisHis
731
                                      G (Dm4 925)
                                                                                                                     2779
TTGACAGCCGGAGCTGCCCGCTACAGAAAGCTAGATTCGCCCACGGATTCGGGCATTGAGTCGGGCAACGAGAAGAACGAGTGCAAGGCGGTGAGTTCGGGGGGAAGTTCCTCGTGCTCC
LeuThrAlaGlyAlaAlaArgTyrArgLysLeuAspSerProThrAspSerGlyIleGluSerGlyAsnGluLysAsnGluCysLysAlaValSerSerGlyGlySerSerCysSer
761                                     Arg
                                                                                                                     2899
AGTCCGCCGTTCCAGTGTGGATGATGCGCTGGACTGCAGCGATGCCGCCGCCAATCACAATCAGGTGGTGCAGCATCCGCAGCTGAGTGTGGTGTCCGTGTCACCAGTTCGCTCGCCCCAG
SerProArgSerSerValAspAspAlaLeuAspCysSerAspAlaAlaAlaAlaAsnHisAsnGlnValValGlnHisProGlnLeuSerValValSerValSerProValArgSerProGln
801
                                                                                                                     3019
CCCTCCACCAGCAGCCATCTGAAGCGACAGATTGTGGAGGATATGCCCGTGCTGAAGCGCGTGCTGCAGGCTCCCCCTCTGTACGATACCAACTCGCTGATGGACGAGGCCTACAAGCCG
ProSerThrSerSerHisLeuLysArgGlnIleValGluAspMETProValLeuLysArgValLeuGlnAlaProProLeuTyrAspThrAsnSerLeuMETAspGluAlaTyrLysPro
841
                                                                                                                     3139
CACAAGAAATTCCGGGCCCTGCGCGCATCGCGAGTTCGAGACCGCCGAGGCGGATGCCAGCAGTTCCACTTCCGGCTCGAACAGCCTGAGTGCCGGCAGTCCGCGGCAGAGCCCAGTCCCG
HisLysLysPheArgAlaLeuArgHisArgGluPheGluThrAlaGluAlaAspAlaSerSerSerThrSerGlySerAsnSerLeuSerAlaGlySerProArgGlnSerProValPro
881
                                                                                                                     3259
AACAGTGTGGCCACGCCCCCGCCATCGGCGGCCAGCGCCGCCGCAGGTAATCCCGCCCAGAGCCAGCTGCACATGCACCTGACCCGCAGCAGCCCCAAGGCCTCGATGGCCAGCTCGCAC
AsnSerValAlaThrProProProSerAlaAlaSerAlaAlaAlaGlyAsnProAlaGlnSerGlnLeuHisMETHisLeuThrArgSerSerProLysAlaSerMETAlaSerSerHis
921
                                                                                                                     3379
TCGGTGCTGGCCAAGTCTCTCATGGCCGAGCCGCGCATGACGCCCGAGCAGATGAAGCGCAGCGATATTATCCAAAACTACTTGAAGCGCGAGAACAGCACAGCAGCCAGCAGCACCACC
SerValLeuAlaLysSerLeuMETAlaGluProArgMETThrProGluGlnMETLysArgSerAspIleIleGlnAsnTyrLeuLysArgGluAsnSerThrAlaAlaSerSerThrThr
961
                                                                                                                     3499
AATGGCGTGGGCAACCGCAGTCCCAGCAGCAGCTCCACACCGCCGCGTCGGCCGGTCCAGAATCAGCAGCGTTGGGGCAGCAGCTCGGTGATCACCACCACCTGCCAGCAGCGCCAGCAG
AsnGlyValGlyAsnArgSerProSerSerSerSerThrProProSerArgArgProValGlnAsnGlnGlnArgTrpGlySerSerSerValIleThrThrThrCysGlnGlnArgGlnGln
1001
                                                                                                                     3619
TCCGTGTCGCCGCACAGCAACGGTTCCAGCTCCAGTTCGAGCTCTAGCTCCAGCTCCAGTTCGTCATCCTCCTCCACATCCTCCAACTGCAGCTCCAGCTCGGCCAGCAGCTGCCAGTAT
SerValSerProHisSerAsnGlySerSerSerSerSerSerSerSerSerSerSerSerSerSerSerThrSerSerAsnCysSerSerSerSerAlaSerSerCysGlnTyr
1041
```

Figure 7A. *(See p. 212 for legend.)*

A (continued)

```
                                                                                                                                 3739
TTCCAGTCGCCGCACTCCACCAGCAACGGCACCAGTGCACCGGCGAGCTCCAGTTCGGGATCGAACAGCGCCACGCCCCTGCTGGAACTGCAGGTGGACATTGCTGACTCGGCGCAGCCT
PheGlnSerProHisSerThrSerAsnGlyThrSerAlaProAlaSerSerSerSerGlySerAsnSerAlaThrProLeuLeuGluLeuGlnValAspIleAlaAspSerAlaGlnPro
1081
                                                                                                                                 3859
CTCAATTTGTCCAAGAAATCGCCCACGCCGCCGCCCAGCAAGCTGCACGCTCTGGTGGCCGCCGCCAATGCCGTTCAAAGGTATCCCACATTGTCCGCCGACGTCACAGTGACAGCCTCC
LeuAsnLeuSerLysLysSerProThrProProProSerLysLeuHisAlaLeuValAlaAlaAlaAsnAlaValGlnArgTyrProThrLeuSerAlaAspValThrValThrAlaSer
1121
                                                                                                                                 3979
AATGGCGGGTCCTCCGTCGGCGGCGGCGAGTCCGGCCGCCAGCAGCAGTCCGCCGGCGAGTGTGGGCTCCCCCAATCCGGGCCTGAGCGCCGCCGTGCACAAGGTAATGCTGGAGGCGTA
AsnGlyGlySerSerValGlyGlyGlyGluSerGlyArgGlnGlnGlnSerAlaGlyGluCysGlyLeuProGlnSerGlyProGluArgArgArgAlaGlnGlyAsnAlaGlyGlyVal
1161
                                                                                                                                 4099
AGAGCGGGAGGAGGTAGGTGGTTTTACGCGGAGAAGTGGGAGAGACAGAGACTGGGAGTGGCAGTTCAGCGAAGCAGGAAGCAGGATCACTTGGAGCGGCGGGAGTTGAATTAAATTATT
ArgAlaGlyGlyGlyArgTrpPheTyrAlaGluLysTrpGluArgGlnArgLeuGlyValAlaValGlnArgSerArgGlnAspHisLeuGlnArgArgGluLeuAsn ·
1201
                                                                                                                                 4219
TTACCATTTAATTGAGACGTGTACAAAGTTTGAAAGCAAAACCAACATGCATGCAATTTAAAACTAATATTTAAAGCAACAACAAACAAAACAACTACAAGTTATTAATTTAAAAAACAA
                                                                                                                                 4339
ACAAACAAACAAACAACAAAAAACCCAAGCTTGAATGGTATTACAAAAGAAAAAGAAAAACAGAAAAAATATAAATATATTTTAGCAGTTAAACTTTAACGTAGCAAGAAACCAACAAAC
                                                                                                                                 4459
CCAAGGCAGCGCTCTGATTTCGCATTAACTTTTCTTCAGCTGCTACCGAAAACGCCCCTCACCTCCCCCCCACCCAACCCTTCCTCCACACACCAACCGTCTTTCGACCCCTGATTGTTT
                                                                                                                                 4579
TATAAGTTTTAAGCTCTTGTTGTACATATTAATTACGTTTATTGGTAACTATGTTTAGCGCTTTAGTTGTAGTTGGAGCAAAACTACTTTGCTTTTTTGGATGTTTTTTGAAAAAACTGC
                                                                                                                                 4699
AAAATTATTATTATTAAATTTTTAAATACCTAAAAACAAAACAATGTGTGTGAAATTTTTTATTGTGCGATCTCCAAGCAGAATGAAGTGCAGTTTGCAACAAATTTTAACTACGATTAAG
                   └──► polyA site cDm4927 and cDm4928
                                                                                                                                 4819
TTGATAACGATTCATTTTTTATGAATTTAACTAATTTTATGAATTTGTTATAGTTTTCCACCCTTCTATAGATCTTTCTATCTGATCATCTAGCTACCCGTATTCCTGATTTCTCCTTTG
                                                                                                                                 4939
GCACAAAGCTCTTCTCTATGCTAAAGAATCAAGTGGAATAAATATTGTTTTCTAATTTTAAAACTACCACAAAAATACGATTAAAATATACACGAAGTAATTGAAAATCAAACAAAATGC
                                                                                                                                 5059
TTAAAGTTTTAGCAGCAAGCAGTAAAACGACGATGAAGAAGAGAAACCCAACGTTAAATATATCTGTTGTGTACATAGTTAAATGTTAAATTAAACACAAAAACATATTTAAAGTACATA
                                                                                                                                 5179
TAAATACACATAATTATTAATGAAGAAACCTATGCTTAAAAGATTCAATGTTTGATTGGCATCTTAGAAAACCAAGCGAAAAATACAAAAAAAAATCAACAAACAAAAATTATGATATAT
                                                                                                                                 5299
TATTTAAAAGTAAAGTATACATTTACATTACAGAAAAACAAAAGAGAAAACTTGCGGTAGCAACAAAACTATTATATTAATTACATTTTAATTATGCTGTACTATTATGATTATTAATTA
                                                                                                                                 5419
TTATGATTAATTAATTACGATTTTTATGCTTAGACAAACCAACAAAAAACAAATATGCAAAAACCATTAAAAAAAAAACAAAAAACAAGCAAAAAATTACACTGGCGCAGAAATTTGTA
                                       └──► putative polyA signal for long mRNAs
                                                                                                                                 5539
TTGTCAAATCAAGAAAAAATTTGTTTAAAAATTTCAAAAGAAAGCCTCTTAAGTTTTTTCATTTCAATTTCCTTTTCAGTTGAACACTTTTTCTTAAAATGTTCAGTTTTACCGTTACTT
                                                                                                                                 5659
TGCTTTGGAGTGGTAAGATTTTGGGTTTCACTTGATTTCATTTTCGTTTTGACTATCGTGCTGTAAAAATATTTACTAAATTATATCGAAAGTATTTATATCATAATTAAATAAACAAGA
```

B

```
CGACGCGTTTGGAGTGAACGTCCTCAGTTGGCACACAAAAACAAAAACACAAAACGACAGCAACAACATCGTGGGGGGAGTACGAGCGGGATGGGGGTAATGGGGGGCACCGGGGGAGT
                                                                                                                                 -9
GGAGGCCGAGAGAGCGAGAGAGCGACCCGAAGCAACACAACACCAACACGAGGCCCAAAAAGACACTTCGGCTGGGTTCAGCTCGTGTTGCTCTGGGTCGTTTTGTATTGCTGGTGGACG
                                                                                                                                 112
CTGCTTTCATTCGCAAATTGCTCGTCGTTGGCAGCGGTTGTGCAGAGCAAGAAAAGCGCGCGAAAAACCAAGCAAAAAATTAATACAGCTGGATCAAGCGAAAGAGATAGAGAGCAGAGT
            └──► mRNA start site
                                                                                                                                 232
CAACAGCAACAAATGTTCAATAGCAAATGATATCGCATATTTTTGTTGGTGCCAGTGAAGTGAGATCAAAGTGAAGTGTGCAATGTTCCTTATTAGCAAATCGTAGAGCAACCAACAATC
                                                                                                                                 352
GAGAGTTCAAGTGTCATTTCGAAGCCAAAAAGCAAAATCTCTAATTCAAATATGGTTTGTGCAATGCAAGAGGTTGCTGCTGTGCAGCATCAGCAGCAGCAACAGCAACTCCAGTTGCCC
                                                                      METValCysAlaMETGlnGluValAlaAlaValGlnHisGlnGlnGlnGlnGlnGlnGlnLeuGlnLeuPro
                                                                                                                                 472
CAGCAGCAACAGCAGCAGCAGCAGACAACACAGCAGCAACATGCAACAACGATAGTGCTGCTGACGGGCAATGGCGGCGGTAATCTGCACATTGTCGCCACACCGCAACAGCATCAGCCG
GlnGlnGlnGlnGlnGlnGlnGlnThrThrGlnGlnGlnHisAlaThrThrIleValLeuLeuThrGlyAsnGlyGlyGlyAsnLeuHisIleValAlaThrProGlnGlnHisGlnPro
24
                                                                                                                                 592
ATGCATCAGCTCCACCATCAGCATCAGCATCAGCATCAGCACCAGCAGCAGGCCAAGAGCCAACAGCTGAAGCAACAACACTCGGCGCTGGTCAAGTTGCTGGAGTCGGCGCCCATCAAG
METHisGlnLeuHisHisGlnHisGlnHisGlnHisGlnHisGlnGlnGlnAlaLysSerGlnGlnLeuLysGlnGlnLeuSerAlaLeuValLysLeuLeuGluSerAlaProIleLys
64
                                                                                                                                 712
CAGCAACAGCAGACGCCCAAGCAAATTGTTTACCTGCAGCAGCAGCAGCAACCGCAACGCAAAAGACTGAAAAACGAAGCAGCAATCGTACAACAGCAACAACAAACACCTGCAACA
GlnGlnGlnGlnThrProLysGlnIleValTyrLeuGlnGlnGlnGlnGlnProGlnArgLysArgLeuLysAsnGluAlaAlaIleValGlnGlnGlnGlnGlnThrProAlaThr
104
                                                                                                                                 832
CTAGTAAAGACAACAACCACCAGCAACAGCAACAGCAACAACACCCAGACAACAAATAGTATTAGTCAGCAGCAACAGCAGCATCAGATTGTGTTGCAGCACCAGCAGCCAGCCGCGGCA
LeuValLysThrThrThrThrSerAsnSerAsnSerAsnAsnThrGlnThrThrAsnSerIleSerGlnGlnGlnGlnHisGlnIleValLeuGlnHisGlnGlnProAlaAlaAla
144
                                                                                                                                 952
GCAACACCAAAGCCATGTGCCGATCTGAGCGCCAAAAATGACAGCGAGTCGGGCATCGACGAGGACTGCCCCAACAGCGATGAGGATTGCCCCAATGCCAACCCGGCGGGCACATCGCTC
AlaThrProLysProCysAlaAspLeuSerAlaLysAsnAspSerGluSerGlyIleAspGluAspCysProAsnSerAspGluAspCysProAsnAlaAsnProAlaGlyThrSerLeu
184
                                                                                                                                 1072
GAGGACAGCAGCTACGAGCAGTATCAGTGCCCCTGGAAGAAGATACGCTATGCGCGTGAGCTCCTCAAGCAGCGCGAGTTGGAGCAGCAGCAGACCACCGGAGGCAGCAACGCGCAGCAG
GluAspSerSerTyrGluGlnTyrGlnCysProTrpLysLysIleArgTyrAlaArgGluLeuLeuLysGlnArgGluLeuGluGlnGlnGlnThrThrGlyGlySerAsnAlaGlnGln
224
                                                                                                                                 1192
CAAGTCGAGGCGAAGCCAGCTGCAATACCCACCAGCAACATCAAGCAGCTGCACTGTGATAGTCCCTTTTCGGCGCAGACCCACAAGGAAATCGCCAATCTCCTGCGCCAACAGTCCCAG
GlnValGluAlaLysProAlaAlaIleProThrSerAsnIleLysGlnLeuHisCysAspSerProPheSerAlaGlnThrHisLysGluIleAlaAsnLeuLeuArgGlnGlnSerGln
264
                                                                                                                                 1312
CAACAACAGGTTGTGGCCACGCAGCAGCAGCAGCAACAGCAGCAGCAGCACCAGCACCAGCAACAACGAAGGGATAGCTCCGACAGCAACTGCTCGCTGATGAGCAACTCGAGCAACTCC
GlnGlnGlnValValAlaThrGlnGlnGlnGlnGlnGlnGlnGlnGlnHisGlnHisGlnGlnGlnArgArgAspSerSerAspSerAsnCysSerLeuMETSerAsnSerSerAsnSer
304
                                                                                                                                 1432
AGTGCGGGCAATTGTTGCACCTGCAACGCTGGCGACGACCAGCAGCTGGAGGAGATGGACGAGGCCCACGATTCGGGCTGCGACGATGAACTTTGCGAGCAGCATCACCAGCGACTGGAC
SerAlaGlyAsnCysCysThrCysAsnAlaGlyAspAspGlnGlnLeuGluGluMETAspGluAlaHisAspSerGlyCysAspAspGluLeuCysGluGlnHisHisGlnArgLeuAsp
344
                                                                                                                                 1552
TCCTCCCAACTGAATTACCTGTGCCAGAAGTTCGATGAGAAACTGGACACGGCGCTGAGCAACAGCAGCGCCAACACGGGGAGGAACACGCCAGCTGTAACAGCTAACGAAGATGCCGAT
SerSerGlnLeuAsnTyrLeuCysGlnLysPheAspGluLysLeuAspThrAlaLeuSerAsnSerSerAlaAsnThrGlyArgAsnThrProAlaValThrAlaAsnGluAspAlaAsp
384
```

Figure 7. (*See following page for legend.*)

neither poly(A) tails nor consensus polyadenylation–cleavage signals. Low-resolution S1 nuclease protection experiments, using the 4224 and 4238 probes (Fig. 3), detected the 1-kb exon 4 found in both cDNAs and putative alternative 3'-terminal exons of 2.0 or 2.8 kb (data not shown), either of which may be represented in truncated form by the exon 5 of the cDNAs. The sequence of two short cDNA clones, λDm4927 and λDm4928, isolated from an ecdysone-induced salivary gland cDNA library (C.W. Jones and D.S. Hogness, unpubl.), shows that both of these clones are polyadenylated at a site located 2045 bp from the acceptor splice site for exon 5 (+4616 in Fig. 7A). These data demonstrate that the alternate consensus polyadenylation–cleavage signal AT-TAAA, located 19 bp upstream from this site, is active and confirm the proposed structure of the 2.0-kb exon. Figure 7A also shows an ATTAAA sequence located 774 bp downstream (at +5365 in Fig. 7A) from the above polyadenylation signal at the position expected if the 2.8 kb exon is also a 3'-terminal exon representing a simple extension of the 2.0-kb exon. Clearly, the model for the 3'-terminal exons of class A and B mRNAs shown in Figure 4 is the simplest that will account for the above data, for the consistent ~0.8-kb difference between the lengths of these two exons and those of the members of each mRNA class, and for the observation that only the longer members of each class hybridize to the 4804 probe (Fig. 3), which is specific to sequences in the interval between the above ATTAAA sequences.

We call the two transcription units producing the class A and B mRNAs the E75 A and E75 B units, respectively, and henceforth refer to their mRNAs and the proteins they encode as the E75 A and E75 B mRNAs and proteins.

Ecdysone-induced expression of E75 mRNAs in the presence and absence of cycloheximide correlates with the behavior of the 75B puff

Ashburner et al. (1974) have carefully characterized the appearance of the 75B and other early ecdysone-inducible puffs in the salivary gland. The 75B puff is rapidly induced by ecdysone and begins to regress within a few hours. Its induction is insensitive to cycloheximide, but its regression is blocked. We have observed the expression of E75 A and B mRNAs under similar experimental conditions, using probes specific for each. Figure 6 demonstrates that transcripts of both classes are induced by the addition of ecdysone to the salivary gland, peak within a few hours, and subsequently disappear. In contrast to the E75 A mRNAs, the E75 B mRNAs appear to be reinduced upon extended culture. An alternative interpretation of this unexpected result is suggested by preliminary experiments indicating that expression of E75 B mRNAs may be delayed in some tissues (W.A. Segraves, unpubl.). The apparent reinduction of these transcripts in isolated salivary glands may thus reflect a delayed induction in contaminating tissues or in a subpopulation of salivary gland cells superimposed on the earlier pattern of induction and regression.

In the presence of cycloheximide, the initial induction of the E75 A and B mRNAs is similar to that in the ab-

Figure 7. Sequences of the *E75* exons and flanking DNA. The sequence is that of the C^s genomic DNA, which was identical to the Dm4925 and Dm4745 cDNAs, except for the T→G change in Dm4925 indicated at position +2691. This change would convert a leucine to an arginine in the protein sequences (for a description of sequencing strategy, see Methods). The Dm4925 cDNA extends from just 5' of the *Eco*RV site at +939 to +4267 in *A*. The Dm4745 cDNA extends from +804 in *B* to a point near the *Hin*dIII site at +4246 in *A*. (*A*) The E75 A exons and flanking DNA. The sequences of the A0, A1, and common exons 2–5 (Fig. 4) are interrupted by intron sequences (lowercase), which are limited to those near the splice sites and are in agreement with consensus sequences for donor (5') and acceptor (3') splice sites (Mount 1982). Numbers at the right end of each line refer to the number of base pairs upstream of the E75 A initiation site if negative; positive numbers refer to positions in the E75 A mRNAs, continuing into the 3' flanking DNA. Numbers at the left end of each line refer to amino acid residues in the E75 A protein. The underlined 14 bp sequence at −159 to −172 exhibits a 13/14 bp match to a sequence (CGTAGCGGGTCTC) found 47 bp upstream of the ecdysone-inducible E74 A transcription unit responsible for the early puff at 74EF (K.C. Burtis, C.S. Thummel, C.W. Jones, F.D. Karim, and D.S. Hogness, in prep.). This sequence represents the proximal part of a 19-bp sequence in the E74 A promoter that binds the protein encoded by the *D. melanogaster zeste* gene (Thummel 1989). Another underlined sequence in the E75 A promoter at −74 to −82 is also found in the E75 B promoter, where it is part of a tandemly repeated octanucleotide (GAGAGAGC) located at −106 to −121 in *B*. This repeat matches the consensus sequence for the binding sites of the GAGA transcription factor (Biggin and Tjian 1988), which also binds to the E74 A promoter (Thummel 1989). Other underlined sequences represent, at −27 to −33, the best match to the TATA box consensus at an appropriate position (Goldberg 1979; Breathnach and Chambon 1981), three AUG codons that are closely followed by in-frame stop codons in the 5'-leader sequence of the E75 mRNAs, and alternative polyadenylation–cleavage signals at 4591 and 5365 that are used by both E75 A and E75 B mRNAs. Homopolymeric repeats of glutamine, most of which are related to the *opa* nucleotide repeat (Wharton et al. 1985), are found at amino acid residues 139–146 and 164–171 in the E75 A-specific region and at 721–729 and 749–753 in the common region. Closer to the carboxyl terminus at residues 1049–1065, a homopolymeric repeat of 17 serine residues is seen, primarily encoded by a repeating hexanucleotide with a TCCAGC consensus. (*B*) The B1 exon and its 5'-flanking DNA. The numbering at the right and left ends of the lines follows the same convention as in *A*. Exons 2–5 shown in *A* are also used in E75 B, but the amino acid residues and base pair numbers shown in *A* must be increased by 157 and 375, respectively, to apply to the E75 B protein and mRNA. The first 10 nucleotides of the 136-nucleotide E75 B-specific intron linking the B1 exon to exon 2 are gtaggttag, whereas the last 10 are shown upstream of nucleotide 1178 in *A*. The underlined sequences represent, in order, the region of homology to a sequence upstream of E75 A, noted above, the best match to the TATA box consensus at −21 to −27, and three AUG codons followed by in-frame stop codons in the 5' leader of the E75 B mRNA. Homopolymeric repeats of at least five glutamine residues are seen at 14–19, 24–31, 116–121, 135–139, 166–170, and 311–325 which, in combination with a histidine/glutamine repeat at 69–78, contribute to 23% overall glutamine content of the E75 B-specific region.

sence of cycloheximide. However, both classes continue to accumulate for several hours after ecdysone addition, as if repression of their synthesis is blocked.

These results demonstrate a strong correlation between the expression of the *E75* mRNAs and puffing at the 75B locus. This correlation, the position of the *E75* gene relative to that of the bands that initially decondense during formation of the 75B puff and to the DNA required for that formation (Fig. 2), and the observation that *E75* probes exhibit in situ hybridization to RNA within this puff (Segraves 1988) strongly imply that *E75* transcription is responsible for the 75B puff.

E75 *transcripts encode two proteins with specific and common domains*

Figure 7 presents a composite of the genomic and cDNA sequence data for *E75*, showing the coding and non-coding sequences of the E75 A and B mRNAs, their splice junctions, and the 5'- and 3'-flanking sequences. Certain sequences of potential interest within the 5'-flanking DNA and in the 5'-leader mRNA sequences are indicated in the legend to Figure 7. Here, we focus on the large open reading frames of the E75 A and B mRNAs that begin at 380 and 284 bp downstream from their respective mRNA start sites, each continuing into the common final exon (exon 5; Fig. 4). The termination codon in exon 5 lies upstream of both alternative polyadenylation sites; thus, the sequence of the encoded protein is not affected by which site is selected. Because the open reading frames in the E75 A and B mRNAs begin in the A0 and B1 exons and merge at the beginning of exon 2, the proteins encoded by the two transcription units differ in the amino-terminal region and are the same in the carboxy-terminal region. The specific amino-terminal regions contain 266 and 423 amino acid residues in the E75 A and B proteins, respectively, whereas their common carboxy-terminal region consists of 971 residues. The predicted molecular weights of the A and B proteins are thus 132,000 and 151,000. The open reading frames display characteristic *Drosophila melanogaster* codon usage (O'Connell and Rosbash 1984), and their extents have been confirmed by in vitro translation of mRNAs transcribed in vitro from cDNA constructs and by expression of fusion proteins in *Escherichia coli* (Segraves 1988). The predicted protein sequence for each protein is punctuated by homopolymeric tracts of amino acids, which are noted in Figure 7 and its legend.

The most interesting and significant result of this amino acid sequence analysis is the finding that the E75 proteins exhibit striking homology with the hormone receptors that belong to the steroid receptor superfamily (for review, see Evans 1988). Comparison of E75 protein sequences to the existing data bases and to individual members of this family reveals that similarities are essentially restricted to two regions, labeled C and E, according to the convention of Krust et al. (1986), and correspond, respectively, to the DNA-binding and hormone-binding domains of these receptors. Figure 8A shows the positions of these two regions in the E75 A

protein and the percent identity that each exhibits to the corresponding regions in members of the steroid receptor superfamily. We relate a detailed structural comparison of these regions and a consideration of their possible functions in the E75 proteins in the Discussion.

Discussion

We have identified and characterized the *E75* ecdysone-inducible gene that occupies the early puff locus at 75B. *E75* contains two overlapping transcription units, E75 A and B, that produce mRNAs exhibiting different 5'-terminal exons and common 3'-terminal exons. Translation of these mRNAs results in the production of proteins having different amino-terminal, but common carboxy-terminal, regions.

Analysis of the sequences of the E75 proteins and comparison to the sequences of known proteins have revealed similarity between the E75 proteins and members of the steroid receptor superfamily (Evans 1988). Figure 8A shows a schematic alignment of the E75 A protein with the glucocorticoid, estrogen, thyroid hormone, vitamin D, and retinoic acid receptors, along with the *Drosophila* knirps (Nauger et al. 1988) and human ear-1 (Miyajima et al. 1989) proteins. We have used the nomenclature of Krust et al. (1986) in dividing the proteins into six regions, lettered A–F, in the amino- to carboxy-terminal direction.

Similarity between E75 A and these proteins is strongest in the C region, a cysteine-lysine-arginine-rich region that is necessary and sufficient for the binding of these receptors to DNA (for review, see Evans 1988; Green and Chambon 1988). The C region consists of 66–68 amino acids, of which 20 residues are invariant within this family. Among these are nine invariant cysteine residues, eight of which are believed to coordinate zinc in the formation of two zinc finger-like structures (Miller et al. 1985; Freedman et al. 1988; Severne et al. 1988). Within the C region, E75 A contains all of the highly conserved residues and is approximately as closely related to other members of the steroid receptor superfamily as they are to one another. Interestingly, E75 A is no more closely related to the knirps (Nauber et. al. 1988) and knirps-related proteins (Oro et al. 1988) than to most of the known vertebrate receptors. Indeed, the closest relative of *E75* appears to be the human *ear-1* gene, which has nearly 80% amino acid identity to E75 A in the DNA-binding domain.

The other region conserved among members of the steroid receptor superfamily is the E region, which is required for steroid binding and for the linkage of steroid-binding and *trans*-activation functions (for review, see Evans 1988; Green and Chambon 1988). Although overall E region similarity (Fig. 8A) is clearly significant for the comparison of E75 A to the thyroid hormone, vitamin D, and retinoic acid receptors, and ear-1, similarity to the glucocorticoid and estrogen receptors is considerably lower. However, the plots of local similarities given in Figure 8B show a clear similarity to each of these proteins within three subregions of the E region,

Figure 8. Similarity between the E75 A protein and members of the steroid receptor superfamily. The FASTP (Lipman and Pearson 1985) protein sequence comparison program was used to align E75 A with proteins of the steroid receptor superfamily. (hGR) Human glucocorticoid receptor (Hollenberg et al. 1985); (hER) human estrogen receptor (Green et al. 1986; Greene et al. 1986); (cERBA) chicken c-*erbA* thyroid hormone receptor (Sap et al. 1986); (rVDR) rat vitamin D receptor (Burmester et al. 1988); (hRAR) human retinoic acid receptor (Giguere et al. 1987); (hEAR1) human *erbA*-related gene product (Miyajima et al. 1989); (knirps) *Drosophila* knirps protein (Nauber et al. 1988). (*A*) The location of conserved regions within the proteins. A, B, C, D, E, and F domains of Krust et al. (1986) are indicated, as are the E75 A exons (Fig. 4) encoding them. The numbers within the C and E domains are the percent identity between E75 A and the indicated protein within those regions. Similarity between E75 A and human ear-1 is also seen in the D region (19 of the first 23 residues). E75 A is not aligned (n.a.) with knirps within the E region. (*B*) The alignment of E75 A with each indicated receptor was used to generate a plot of the number of identities per 10 amino acids at 1-amino-acid intervals. The abscissa is defined as the first E75 A amino acid of the 10-amino-acid window, starting with residue 241, 4 amino acids from the beginning of the C region. The values for the four comparisons shown have been summed to give a total similarity index for the E75 A comparison. Conserved regions C1, C2, E1, E2, and E3 are indicated. (*C*) Amino acid sequence of the five regions of greatest similarity between the E75 A protein and members of the steroid receptor superfamily. The position of the first amino acid of each of the regions of similarity is given for each of the protein sequences, and conserved residues have been boxed. The following amino acid substitutions have been allowed: L-I, L-M, I-V, D-E, K-R, R-H, and F-Y-W (for discussion of criteria, see Schulz and Schirmer 1979). The estrogen receptor was not aligned with E75 in the right-hand portion of region E2 by the FASTP program.

which we call E1, E2, and E3. The E1 subregion is the most highly conserved and corresponds to a region shown by in vitro mutagenesis to be essential for steroid binding and steroid-dependent *trans*-activation (Giguere et al. 1986; Danielson et al. 1987). Region E2 is less highly conserved in primary amino acid sequence but can, in part, be seen as a conserved hydrophobic region in the hydropathy plots of several of these proteins (data not shown). A deletion of 14 amino acids within this region abolished steroid binding (Rusconi and Yamamoto 1987). E3 falls close to the end of the region that is

absolutely required for steroid binding (Rusconi and Yamamoto 1987).

The sequences of the C regions and of each of the highly conserved E subregions are shown in Figure 8C, where conserved residues have been boxed to indicate similarity. Based on these results, we believe that it is likely that E75 A (and probably E75 B; see below) is a bona fide member of the steroid receptor superfamily and contains DNA-binding and hormone-binding domains functionally related to those of the other receptors. The finding of both characteristics in a *Droso-*

phila protein clearly establishes the evolutionary age of the conserved structural unit encompassing the DNA- and hormone-binding domains of these receptors. Indeed, at least two other *Drosophila* genes that encode proteins with both characteristics have recently been cloned and characterized (Segraves 1988; M. Koelle, W.A. Segraves, W. Talbot, M. Bender, and D.S. Hogness, in prep.). In contrast to E75 and the other receptor-like proteins, the *Drosophila* knirps and knirps-related proteins contain only the conserved DNA-binding domain. Structural divergence of the carboxyl termini of these molecules from the other receptors and from one another suggests that they may, in fact, be non-ligand-regulated transcription factors.

Although the characteristic structural features of the steroid receptor superfamily are well conserved in *E75*, two novel variations are seen. The first of these concerns the structure of E75 B protein, which contains a major alteration within its putative DNA-binding domain. The steroid receptor superfamily DNA-binding domain appears to consist of two DNA-binding zinc fingers, C1 and C2, separated by a less conserved linker region (Fig. 8B,C). In *E75*, as in nearly all other genes of this family, an intron is found between the two fingers. This intron occurs either one codon after the final cysteine residue of the first finger—in E75 A, the chicken c-*erbA* thyroid hormone receptor gene (Zahraoui and Cuny 1987), the human retinoic acid receptor β-gene (Dejean et al. 1986; de The et al. 1987), *knirps* (Nauber et al. 1988) and *knirps-related* (Oro et al. 1988) genes—or within the tenth codon after this cysteine residue, in the progesterone (Jeltch et al. 1986; Huckaby et al. 1987), mineralicorticoid (Arriza et al. 1987), and estrogen receptors (Ponglikitmongkol et al. 1988).

In *E75*, this splice marks the beginning of the region held in common between the E75 A and B proteins. This region causes the E75 A protein to have two fingers, whereas the E75 B protein has unrelated B-specific sequences in place of the first finger. Mutagenesis of the estrogen, glucocorticoid, and thyroid hormone receptors has shown that the first finger includes residues required for the recognition of specific response elements by these receptors (Danielson et al. 1989; Mader et al. 1989; Umesono and Evans 1989). Thus, a seemingly critical part of the E75 B DNA-binding domain has been removed. It is possible that other sequences within the B-specific amino terminus provide this recognition function, although we detect no significant similarity between the predicted E75 B-specific sequences and any reported protein sequence. (There are a number of cysteine and histidine residues in the last 100 amino acids of the B region, but this region is more acidic than most finger-like structures.) Alternatively, the B protein may bind DNA with only one finger, as GAL4 appears to do (Laughon and Gesteland 1984; Keegan et al. 1986; Johnston and Dover 1987). Although the possibility that the E75 B protein does not bind DNA cannot be ignored, we think it more likely that these structural differences imply a functional difference in the DNA-binding properties of the E75 A and B proteins that may allow them

to differentially regulate secondary ecdysone response. Although demonstration that both the E75 A and B proteins are functional awaits detailed mapping of *E75* mutants, the genetic complementation patterns and phenotypes of *E75* mutants indicate the existence of two genetically separable *E75* functions (Segraves 1988).

The second unusual feature of the E75 proteins is the presence of a large F region, encompassing nearly one-half of the proteins (Fig. 8A). Many of the other receptors have very small F regions, and no function has yet been ascribed to this region. However, preliminary experiments indicate that this region may play an important role in the regulation of target genes by the E75 proteins (W.A. Segraves, unpubl.).

The finding that the E75 proteins are likely hormone receptors raises a number of interesting questions. What is the ligand for the E75 proteins, and how does it affect the function of *E75* in ecdysone-triggered regulatory hierarchies? Because ecdysteroids are the only known steroid hormones in *Drosophila*, the most obvious candidate for an E75 ligand would be ecdysone itself. However, neither binding studies using partially purified E75 proteins and radiolabeled ecdysteroids nor cotransfection studies examining the regulation of target genes by E75 support this idea (W.A. Segraves, unpubl.). Furthermore, a *Drosophila* ecdysone receptor gene, *EcR*, has recently been isolated as a homolog of *E75* that encodes a protein with sequence similarities to both the DNA- and hormone-binding domains of the steroid receptor superfamily (Segraves 1988; M. Koelle, W.A. Segraves, W. Talbot, M. Bender, and D.S. Hogness, in prep.). This protein was identified as an ecdysone receptor on the basis of its ecdysone-binding, DNA-binding, and regulatory activities (M. Koelle, W.A. Segraves, P. Cherbas, and D.S. Hogness, in prep.). The putative hormone-binding domains of the EcR and E75 proteins do not exhibit the high sequence similarity that one might expect if E75 were also an ecdysone receptor. In the absence of other known small hydrophobic regulatory molecules, it therefore seems likely that the E75 proteins would bind either a terpenoid juvenile hormone or a novel *Drosophila* hormone.

For a number of reasons, we consider it likely that *E75* and its ligand are intimately involved in the regulation of ecdysone response. *E75* is clearly induced in the presence of ecdysone, and genetic evidence suggests that a gene, or genes, within a small region of the chromosome, including *E75*, is involved in the regulation of ecdysone-induced puffing patterns (Walker and Ashburner 1981). The localization of E75 A protein to the sites of several ecdysone-regulated genes is consistent with just such a model (R.J. Hill, W.A. Segraves, W.S. Talbot, E. MacAvoy, and D.S. Hogness, unpubl.). Finally, the level of an ecdysone-inducible enzyme, arginine kinase, is sensitive to the dosage of the small deletion, $E75^{x48}$, which removes the *E75* gene (Collier and James 1987; James and Collier 1987).

These observations suggest that *E75* is the type of regulatory early gene predicted by the Ashburner model (Ashburner et al. 1974), that is, a gene encoding regula-

tory proteins involved in the repression of early gene response and/or in the activation of late gene response. Indeed, through the production of two different E75 proteins that are each ligand-regulated, several different types of secondary response could be generated. In a variety of systems, it is now apparent that a single hormone, such as thyroid hormone or retinoic acid, may have multiple receptors whose differential patterns of expression may be involved in the generation of temporally or spatially restricted hormone responses (Giguere et al. 1987; Thompson et al. 1987; Brand et al. 1988; Zelent et al. 1989). The E75 gene displays an elegant mechanism for the production of two different receptor molecules. These receptors carry the same hormone-binding domain and, thus, should bind the same ligand, but have different DNA-binding domains and amino termini, potentially allowing them to differentially regulate subsequent gene expression. The differential expression patterns of the mRNAs encoding these receptors during development (W.A. Segraves and D.S. Hogness, in prep.) suggests that these differences may be significant in the generation of diversity of ecdysone response and that E75 may indeed be a branchpoint at which more general primary responses diverge into more stage- or tissue-specific secondary responses.

Methods

Genomic and cDNA libraries

C^s genomic DNAs were isolated from a library of sheared, EcoRI-linkered C^s DNA cloned into the Charon 4 λ phage vector (Maniatis et al. 1978). O^r genomic DNAs were isolated from a library of sheared DNA, GC-tailed into the sep6 λ vector (Meyerowitz and Hogness 1982). One step in the chromosomal walk was taken using a cosmid library (prepared in collaboration with S. Germeraad) of SauIIIA partially digested O^r DNA cloned into the cosmid p14B1 (Segraves 1988) by the method of Ish-Horowicz and Burke (1981).

The λDm4925 and λDm4745 cDNAs were isolated from an O^r early pupal cDNA library in λgt10 (Poole et al. 1985). The two cDNAs (λDm4927 and λDm4928) that were used for 3'-end mapping were isolated from an ecdysone-induced salivary gland cDNA library in λ607 prepared by C.W. Jones. (Our strain collection names for the cDNA clones used in these studies are λfDm4925, λfDm4745, λeDm4927, and λeDm4928.)

In situ hybridization

In situ hybridization to polytene chromosomes was carried out with DNA probes that were nick-translated in the presence of ³H-labeled TTP (NEN), as described by Bonner and Pardue (1976), with the following modifications: Heat and RNase treatments of the slides were omitted, and hybridization and washing were at 63°C in 2× SSPE for 18 and 2 hr, respectively.

Organ culture and RNA isolation

Late third instar O^r larvae were harvested, washed in 0.7% NaCl, resuspended in Robb's phosphate-buffered saline (PBS) (Robb 1969; preaerated with a blender), and passed through a set of rollers to extrude the organs. This 'grindate' was filtered through a course Nitex screen to remove carcasses and settled five times (3–5 min per settling) by gravity to remove floating

and microscopic debris. Isolated tissues (primarily salivary glands, imaginal discs, gut, and Malphigian tubules) were cultured at 25°C in plastic petri dishes in aerated Robb's PBS. β-Ecdysone (Sigma) (0.2 μ/ml of 10 mg/ml) in ethanol and/or cycloheximide (2 μl/ml of 35 mM) in water was added to the appropriate cultures. Incubations in the presence of cycloheximide (other than the time course in Fig. 6) were for ~8 hr. Isolated tissues were homogenized in 10 volumes of 6 M guanidine-HCl/0.6 M sodium acetate (pH 5.2), centrifuged at 5000g for 10 min to remove debris, and layered onto a 5.7 M CsCl shelf, as described previously (Chirgwin et al. 1979). Poly(A)+ RNA was purified by oligo(dT) chromatography.

Southern and Northern blot analysis

Southern blots were performed on nitrocellulose and Northern blots on APT paper (Seed 1982), as described previously (Segraves et al. 1984). Reverse genomic Southern analysis was performed by hybridizing genomic clones with nick-translated genomic DNA; this technique readily identifies repetitive sequences. cDNA probes were prepared by incubation of 2 μg of poly(A)+ RNA with 700 ng of oligo(dT)[12-18] (Collaborative Research) or 15 μg of random hexamers (Pharmacia) in a 20-μl reaction mixture containing 80 mM Tris-Cl (pH 8.3 at 42°C), 10 mM MgCl$_2$, 100 mM KCl, 0.4 mM DTT, 0.25 mM each of dATP, dGTP, and dTTP, and 100 μCi of [³²P]dCTP (800 Ci/mole; Amersham). After incubation at 37°C for 45 min, 80 μl of 10 mM EDTA and 2 μl of 5 N NaOH were added before incubation at 70°C for 10 min to denature the products and hydrolyze the RNA. After the addition of 10 μl of 1 M Tris-Cl (pH 7.5) and 5 μl of 1 N HCl, unincorporated label was removed by chromatography on Biogel P60. Probes to be used for Northern blots were cloned into the vector pφX (R. Mulligan, unpubl.), containing the φX174 origin of replication cloned in between the HindIII and BamHI sites of pBR322. This allowed the synthesis of single-stranded probe DNA (Arai et al. 1980), which was performed by the incubation of supercoiled plasmid DNA with gene A protein, rep and ssb proteins, and DNA polymerase III holoenzyme (all generously provided by the A. Kornberg laboratory) in a reaction containing 20 mM Tris-Cl (pH 7.5), 80 μg/ml BSA, 4% glycerol, 20 mM DTT, 1 mM ATP, 16 mM concentrations of the three unlabeled deoxynucleotides and 1.6 mM concentrations of the labeled deoxynucleotide for 1 hr at 30°C. EDTA was then added to 20 mM, SDS to 0.1%, and proteinase K to 50 μg/ml. The reactions were digested for 30 min at 37°C, and unincorporated label was removed by gel filtration.

S1 nuclease protection and primer extension analysis

Single-stranded probes, prepared as described above by the φX in vitro replication system, were purified by electrophoresis on low melting point agarose gels for use as S1 probes. All other probes were prepared by extension of the −20, 17-mer sequencing primer (New England Biolabs) on single-stranded M13mp (Messing 1983) or pEMBL (Dente et al. 1983) recombinant templates using ³²P-labeled nucleotides, followed by cleavage with the appropriate restriction enzyme and purification of the probe on denaturing polyacrylamide gels. Labeled probe (100,000–300,000 cpm) was incubated with 1 μg of poly(A)+ RNA in a 5-μl reaction mixture containing 5 μg of yeast tRNA, 0.4 M NaCl, 40 mM PIPES (pH 6.8), and 1 mM EDTA at 60°C under oil. Reactions were cooled and diluted 1 : 10 into either S1 digestion or primer extension buffer. S1 nuclease digestions were performed in 50 mM acetate buffer (Na), 400 mM NaCl, and 4 mM ZnSO$_4$ at 20°C for 1 hr with ~15–150 Vogt units of S1 nuclease (Boehringer) per 50-μl reac-

tion. Primer extensions were performed at 42°C in 50 mM Tris-Cl (pH 8.3 at 42°C), 80 mM KCl, 2 mM DTT, 1 mM each of dATP, dCTP, dGTP, and dTTP, with 20 units of AMV reverse transcriptase (Seikagaku) per 50-μl reaction. Reactions were terminated by the addition of EDTA, tRNA carrier was added to the S1 nuclease digestions, and samples were ethanol-precipitated and either electrophoresed directly on 5% or 6% denaturing polyacrylamide gels or glyoxalated (McMaster and Carmichael 1977) and electrophoresed on 1% agarose gels run in 10 mM sodium phosphate buffer (pH 6.8).

DNA sequence analysis

The cDNA clones λDm4927 and λDm4928 were sequenced by chemical degradation (Maxam and Gilbert 1980). All other sequencing was performed using the dideoxynucleotide chain termination method (Sanger et al. 1980). Fragments were cloned into M13mp (Messing 1983) or pEMBL (Dente et al. 1983) vectors and sequenced directly or following the generation of a set of overlapping deletions using exonuclease III (Henikoff 1984). Sequencing was performed on both strands of the λDm4925 cDNA, the B-specific region of λDm4745 cDNA, the A- and B-specific 5' genomic regions not represented in the cDNAs, and the 3'-flanking region. The remaining exon boundaries of λDm4745 and genomic regions represented within the cDNA clones were sequenced on one strand. Sequence data were compiled using the Bionet system. Protein sequence comparison was performed using FASTP (Lipman and Pearson 1985) and Bionet IFIND programs.

Acknowledgments

Note added in proof

Sequence data described in this paper have been submitted to the EMBL/GenBank Data Libraries.

References

Arai, K., N. Arai, J. Schlomai, and A. Kornberg. 1980. Replication of duplex DNA of phage ϕX174 reconstituted with purified enzymes. *Proc. Natl. Acad. Sci.* **77:** 3322–3326.

Arriza, J.L., C. Weinberger, G. Cerelli, T.M. Glaser, B.L. Handelin, D.E. Housman, and R.M. Evans. 1987. Cloning of human mineralicorticoid receptor cDNA: Structural and functional kinship with the glucocorticoid receptor. *Science* **237:** 268–275.

Ashburner, M. 1972. Patterns of puffing activity in the salivary glands of *Drosophila*. VI. Induction by ecdysone in salivary glands of *D. melanogaster* cultured in vitro. *Chromosoma* **38:** 255–281.

———. 1974. Sequential gene activation by ecdysone in polytene chromosomes of *Drosophila melanogaster*. II. The effects of inhibition of protein synthesis. *Dev. Biol.* **39:** 141–157.

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

Belyaeva, E.S., I.E. Vlassova, Z.M. Biyasheva, V.T. Kakpakov, G. Richards, and I.F. Zhimulev. 1981. Cytogenetic analysis of the 2B3-4–2B11 region of the X chromosome of *Drosophila melanogaster*. II. Changes in the 20-OH ecdysone puffing caused by genetic defects of puff 2B5. *Chromosoma* **84:** 207–219.

Belyaeva, E.S. and I.F. Zhimulev. 1982. Cytogenetic analysis of the X chromosome region 2B3-4–2B11. IV. Mutation at the *swi* (*singed wings*) locus interfering with the late 20-OH ecdysone puff system. *Chromosoma* **86:** 251–263.

Bender, W., P. Spierer, and D.S. Hogness. 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.

Biggin, M.D. and R. Tjian. 1988. Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. *Cell* **53:** 699–711.

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

Brand, N., M. Petkovich, A. Krust, P. Chambon, H. de The, A. Marchio, P. Tiollais, and A. Dejean. 1988. Identification of a second human retinoic acid receptor. *Nature* **332:** 850–853.

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

Burmester, J.K., R.J. Wiese, N. Maeda, and H.F. DeLuca. 1988. Structure and regulation of the rat 1,25-dihydroxyvitamin D_3 receptor. *Proc. Natl. Acad. Sci.* **85:** 9499–9502.

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

Chao, A.T., and G.M. Guild. 1986. Molecular analysis of the ecdysterone-inducible 2B 'early' puff in *Drosophila melanogaster*. *EMBO J.* **5:** 143–150.

Chirgwin, J.M., A.E. Przbyla, R.J. MacDonald, and W.J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18:** 5294–5299.

Collier, G.E. and J.M. James. 1987. Arginine kinase and pyruvate kinase are coordinately regulated by the 75B 'early' gene in *Drosophila* imaginal discs. *Genetics* **116:** s12.

Danielson, M., J.P. Northrop, J. Jonklaas, and G.M. Ringold. 1987. Domains of the glucocorticoid receptor involved in specific and nonspecific deoxyribonucleic acid binding, hormone activation and transcriptional enhancement. *Mol. Endocrinol.* **1:** 816–822.

Danielson, M., L. Hinck, and G.M. Ringold. 1989. Two amino acids within the knuckle of the first zinc finger specify DNA response element activation by the glycocorticoid receptor. *Cell* **57:** 1131–1138.

Dejean, A., L. Bougueleret, K. Grzeschik, and P. Tiollais. 1986. Hepatitis B virus DNA integration in a sequence homologous to v-*erbA* and steroid receptor genes in a hepatocellular carcinoma. *Nature* **322:** 70–72.

Dente, L., G. Cesareni, and R. Cortes. 1983. pEMBL: A new family of single-stranded plasmids. *Nucleic Acids Res.* **11:** 1645–1654.

de The, H., A. Marchio, P. Tiollais, and A. Dejean. 1987. A novel steroid thyroid hormone receptor-related gene inappropriately expressed in human hepatocellular carcinoma. *Nature* **330:** 667–670.

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

Freedman, L.P., B.F. Luisi, Z.R. Korszun, R. Basavappa, P.B. Sigler, and K.R. Yamamoto. 1988. The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. *Nature* **334:** 543–546.

Giguere, V., S.M. Hollenberg, M.G. Rosenfeld, and R.M. Evans. 1986. Functional domains of the human glucocorticoid receptor. *Cell* **46:** 645–652.

Giguere, V., E.S. Ong, P. Segui, and R.M. Evans. 1987. Identification of a receptor for the morphogen retinoic acid. *Nature* **220:** 624–629.

Goldberg, M.L. 1979. 'Sequence analysis of *Drosophila* histone genes.' Ph.D. thesis, Stanford University, Stanford, California.

Green, S., P. Walter, V. Kumar, A. Krust, J. Bornert, P. Argos, and P. Chambon. 1986. Human estrogen receptor cDNA: Sequence, expression and homology to v-*erbA*. *Nature* **320:** 134–139.

Green, S. and P. Chambon. 1988. Nuclear receptors enhancer our understanding of transcription regulation. *Trends Genet.* **4:** 309–314.

Greene, G.L., P. Gilna, M. Waterfield, A. Baker, Y. Hort, and J. Shine. 1986. Sequence and expression of human estrogen receptor complementary DNA. *Science* **231:** 1150–1154.

Gronemeyer, H. and O. Pongs. 1980. Localization of ecdysterone on polytene chromosomes of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **77:** 2108–2112.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28:** 351–359.

Hollenberg, S.M., C. Weinberger, E.S. Ong, G. Cerelli, A. Oro, R. Lebo, E.B. Thompson, M.G. Rosenfeld, and R.M. Evans. 1985. Primary structure and expression of a functional human glucocorticoid receptor cDNA. *Nature* **318:** 635–641.

Huckaby, C.S., O.M. Conneely, W.G. Beattie, A.D.W. Dobson, M. Tsai, and B.W. O'Malley. 1987. Structure of the chromosomal chicken progesterone receptor gene. *Proc. Natl. Acad. Sci.* **84:** 8380–8384.

Ish-Horowicz, D. and J.F. Burke. 1981. Rapid and efficient cosmid cloning. *Nucleic Acids. Res.* **9:** 2989–2998.

James, J.M. and G.E. Collier. 1987. Developmental regulation of arginine kinase in *Drosophila* imaginal discs. *ICSU Short Rep.* **7:** 81.

Jeltsch, J.M., Z. Krozowski, C. Quirin-Stricker, H. Gronemeyer, R.J. Simpson, J.M. Garnier, A. Krust, F. Jacob, and P. Chambon. 1986. Cloning of the chicken progesterone receptor. *Proc. Natl. Acad. Sci.* **83:** 5424–5428.

Johnston, M. and J. Dover. 1987. Mutations that inactivate a yeast transcriptional regulatory protein cluster in an evolutionarily conserved DNA binding domain. *Proc. Natl. Acad. Sci* **84:** 2401–2405.

Keegan, L., G. Gill, and M. Ptashne. 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* **231:** 699–704.

Krust, A., S. Green, P. Argos, V. Kumar, P. Walter, J. Bornert, and P. Chambon. 1986. The chicken oestrogen receptor sequence: Homology with v-*erbA* and the human oestrogen and glucocorticoid receptors. *EMBO J.* **5:** 891–897.

Laughon, A. and R.F. Gesteland. 1984. Primary structure of the *Saccharomyces cerevisiae* GAL4 gene. *Mol. Cell. Biol.* **4:** 260–267.

Lipman, D.J. and W.R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* **227:** 1435–1441.

Mader, S., V. Kumar, H. de Verneuil, and P. Chambon. 1989.

Three amino acids of the oestrogen receptor are essential to its ability to distinguish an oestrogen from a glucocorticoid-responsive element. *Nature* **338:** 271–274.

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

Maxam, A.M. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavage. *Methods Enzymol.* **65:** 499–560.

McMaster, G.K. and G.G. Carmichael. 1977. Analysis of single and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci.* **74:** 4835–4838.

Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* **101:** 20–78.

Meyerowitz, E.M. and D.S. Hogness. 1982. Molecular organization of a *Drosophila* puff site that responds to ecdysone. *Cell* **28:** 165–176.

Miller, J., A.D. McLachlan, and A. Klug. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4:** 1609–1614.

Miyajima, N., R. Horiubuchi, Y. Shibuya, S. Fukushige, K. Matsubara, K. Toyoshima, and T. Yamamoto. 1989. Two *erbA* homologs encoding proteins with different T_3 binding capacities are transcribed from opposite strands of the same genetic locus. *Cell* **57:** 31–39.

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

Nauber, U., M.J. Pankratz, A. Kienlin, E. Seifert, U. Klemm, and H. Jäckle. 1988. Abdominal segmentation of the *Drosophila* embryo requires a hormone receptor-like protein encoded by the gap gene *knirps*. *Nature* **336:** 489–492.

O'Connell, P. and M. Rosbash. 1984. Sequence, structure and codon preference of the *Drosophila* ribosomal protein 49 gene. *Nucleic Acids Res.* **12:** 5495–5513.

Oro, A.E., E.S. Ong, J.S. Margolis, J.W. Posakony, M. McKeown, and R.M. Evans. 1988. The *Drosophila* gene *knirps-related* is a member of the steroid receptor superfamily. *Nature* **336:** 493–496.

Ponglikitmongkol, M., S. Green, and P. Chambon. 1988. Genomic organization of the human oestrogen receptor gene. *EMBO J.* **7:** 3385–3388.

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

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

Robb, J.A. 1969. Maintenance of imaginal discs of *Drosophila melanogaster* in chemically defined media. *J. Cell. Biol.* **41:** 876–885.

Rusconi, S. and K.R. Yamamoto. 1987. Functional dissection of the hormone and DNA binding activities of the glucocorticoid receptor. *EMBO J.* **6:** 1309–1315.

Sanger, F., A.R. Coulson, B.F. Barrell, A.J.H. Smith, and B.A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143:** 161–178.

Sap, J., A. Munoz, K. Damm, Y. Goldberg, J. Ghysdael, A. Leutz, H. Beug, and B. Vennstrom. 1986. The c-*erb-A* protein is a high-affinity receptor for thyroid hormone. *Nature* **324:** 635–640.

Schulz, G. and R. Schirmer. 1979. Amino acids. In *Principles of protein structure*. pp. 14–16. Springer-Verlag, New York.

Seed, B. 1982. Diazotizable arylamine cellulose papers for the

coupling and hybridization of nucleic acids. *Nucleic Acids Res.* **10:** 1799–1810.

Segraves, W.A., C. Louis, S. Tsubota, P. Schedl, J.M. Rawls, and B.P. Jarry. 1984. The *rudimentary* locus of *Drosophila melanogaster. J. Mol. Biol.* **175:** 1–17.

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

Semishin, V.F., E.M. Baricheva, E.S. Belyaeva, and I.F. Zhimulev. 1985. Electron microscopal analysis of *Drosophila* polytene chromosomes. II. Development of complex puffs. *Chromosoma* **91:** 210–233.

Severne, Y., S. Wieland, W. Schaffner, and S. Rusconi. 1988. Metal binding 'finger' structure of the glucocorticoid receptor defined by site-directed mutagenesis. *EMBO J.* **7:** 2503–2508.

Thompson, C.C., C. Weinberger, R. Lebo, and R.M. Evans. 1987. Identification of a novel thyroid hormone receptor abundantly expressed in the mammalian central nervous system. *Science* **237:** 1610–1614.

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

Umesono, K. and R.M. Evans. 1989. Determinants of gene specificity for steroid/thyroid hormone receptors. *Cell* **57:** 1139–1146.

Walker, V. and M. Ashburner. 1981. The control of ecdysterone-regulated puffs in *Drosophila* salivary glands. *Cell* **26:** 269–277.

Wharton, K.A., B. Yedvobnick, V.G. Finnerty, and S. Artavanis-Tsakonas. 1985. opa: A novel family of transcribed repeats share by the *Notch* locus and other developmentally regulated loci in *D. melanogaster. Cell* **40:** 55–62.

Zahraoui, A. and G. Cuny. 1987. Nucleotide sequence of the chicken proto-oncogene c-*erbA* corresponding to domain 1 of v-*erbA. Eur. J. Biochem.* **166:** 63–69.

Zelent, A., A. Kurst, M. Petkovich, P. Kastner, and P. Chambon. 1989. Cloning of murine α and β retinoic acid receptors and a novel receptor γ predominantly expressed in skin. *Nature* **339:** 714–717.

Zhimulev, I.F., I.E. Vlassova, and E.S. Belyaeva. 1982. Cytogenetic analysis of the 2B3-4–2B11 region of the X chromosome of *Drosophila melanogaster*. III. Puffing disturbance in salivary gland chromosomes of homozygotes for mutation 1(1) ppl[t10]. *Chromosoma* **85:** 659–672.

The E75 ecdysone-inducible gene responsible for the 75B early puff in Drosophila encodes two new members of the steroid receptor superfamily.

W A Segraves and D S Hogness

References	This article cites 71 articles, 17 of which can be accessed free at: **http://genesdev.cshlp.org/content/4/2/204.full.html#ref-list-1**
License	
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**