Developmentally regulated alternative splicing of transcripts from the *Drosophila* homeotic gene *Antennapedia* can produce four different proteins

John R. Bermingham, Jr and Matthew P. Scott

Department of Molecular, Cellular and Developmental Biology,
University of Colorado at Boulder, Boulder, CO 80309-0347, USA

Communicated by V. Pirrotta

***Antennapedia* (*Antp*) is a *Drosophila* homeotic gene that controls differentiation of the thoracic segments. *Antp* transcripts are produced from either of two promoters that are independently regulated in temporally and spatially distinct patterns. In addition, *Antp* transcripts utilize either of two major polyadenylation sites. *Antp* primary transcripts contain the same protein coding sequences. Alternative RNA splicing at two positions within the primary transcripts produces mRNAs that can encode four slightly different *Antp* proteins. Different classes of alternatively spliced transcript predominate early and late in *Drosophila* development, indicating that the *Antp* gene is regulated by the processing of its transcripts as well as by controlling their transcription. Alternative splicing appears to be independent of which promoter and which polyadenylation site is used.**

Key words: *Antennapedia*/*Drosophila*/homeotic/mRNA/splicing

Introduction

The *Drosophila* body consists of segments, each different from the others. Many of the genes controlling development in *Drosophila* have been identified; they act to divide the developing embryo into segments and to specify the differentiation of the segments into distinct tissues and structures (reviewed in Gehring and Hiromi, 1986; Scott and O'Farrell, 1986; Akam, 1987; Scott and Carroll, 1987). The homeotic genes control segment differentiation. How are the homeotic genes regulated and what do their products do? The genes are expressed in specific regions along the anterior–posterior axis of the embryo (reviewed in Gehring and Hiromi, 1986; Scott and O'Farrell, 1986; Akam, 1987; Scott and Carroll, 1987). It appears that the patterns of expression of the homeotic genes are set up by at least some of the segmentation genes (Ingham and Martinez-Arias, 1986; Duncan, 1986; White and Lehmann, 1986); the patterns are then maintained by regulating interactions among the homeotic genes themselves. Currently, it is hypothesized that the products of homeotic genes control the activities of other genes. Many of the homeotic genes have been shown to share a common DNA sequence, the homeobox (McGinnis *et al.*, 1984; Scott and Weiner, 1984; reviewed in Gehring, 1987). The homeobox encodes a putative DNA binding domain (Laughon and Scott, 1984; Shephard *et al.*, 1984), and homeodomain-containing proteins have been shown to be capable of sequence-specific DNA binding *in vitro* (Desplan

et al., 1985). These findings suggest that the products of these genes may act as transcriptional activators or repressors.

The *Antennapedia* (*Antp*) gene is one of the genes that controls the identity of the thoracic segments. *Antp* dominant and/or recessive mutations produce homeotic transformations involving thoracic structures. Most of the dominant *Antp* mutants produce antenna-to-second thoracic leg transformations. The dominant *Antp* phenotypes result from ectopic expression of the gene (Denell, 1973; Struhl, 1981; Hazelrigg and Kaufman, 1983; Frischer *et al.*, 1986; Schneuwly *et al.*, 1987a,b). However, the majority of *Antp* mutations are recessive lethals with no dominant phenotype (Lewis *et al.*, 1980; Struhl, 1981). Individuals homozygous for most *Antp* mutations die as late embryos or early larvae with their second and third thoracic segments transformed into first thoracic segments (Denell *et al.*, 1981; Wakimoto and Kaufman, 1981).

The *Antp* gene has been cloned (Scott *et al.*, 1983; Garber *et al.*, 1983), and partially sequenced (Schneuwly *et al.*, 1986; Strocher *et al.*, 1986; Laughon *et al.*, 1986). The structure of the gene is diagrammed in Figure 1. The gene is large; exons span 102 kb of DNA. Transcripts are initiated at either of two different promoters and are polyadenylated at two major polyadenylation sites. Thus, the gene consists of two transcription units, one nested inside the other. Transcription from the upstream promoter, P1, produces mRNAs of 3.4 kb and 4.9 kb, depending upon which polyadenylation site is used. Likewise, transcripts initiated at the internal promoter, P2, produce mRNAs of 3.6 kb and 5.1 kb. All four classes of transcripts contain the same coding regions and produce proteins of ~42 kd.

The P1 and P2 *Antp* promoters are controlled independently. *In situ* hybridization of promoter-specific probes to section embryos revealed that the two promoters are expressed in different patterns, and both promoters are strongly expressed in the thorax (A. Martinez-Arias, J.R. Bermingham, Jr and M.P. Scott, submitted). Both promoters are also expressed in mesothoracic imaginal discs, and again the patterns are different (Jorgensen and Garber, 1987). During early embryogenesis, the P2 promoter is regulated by the segmentation gene *fushi tarazu* (*ftz*), while the P1 promoter is not (Ingham and Martinez-Arias, 1986), and presumably other *trans*-acting genes differently control the two promoters. The results of complementation studies using *Antp* mutations indicate that mutations disrupting P1-derived transcripts can complement mutations disrupting P2-derived transcripts (Kaufman and Abbott, 1984; Abbott and Kaufman, 1986). These results indicate that the two *Antp* promoters are controlled separately and differently, and raise the possibility that the gene could perform different functions in different parts of the fly, both during embryogenesis and metamorphosis. The phenotypes of somatically induced clones of *Antp* mutant tissue also suggest that *Antp* plays

Fig. 1. Schematic diagram of the *Antp* gene. Exons are represented as boxes; the shaded portions depict the coding region. Introns are represented as single lines connecting the exons. The figure is to scale, but those introns that are broken are 10-fold longer than depicted in the figure. P1, P2: promoters. A1, A2: polyadenylation sites. HB: homeobox. Coordinates in kb refer to the chromosome walk of Scott *et al.* (1983).

Fig. 2. (a) An internal discontinuity exists between two *Antp* cDNAs, G1100 and cYE12. Pairs of ssDNAs containing cDNAs G1100 and cYE12 or cYE10, were hybridized to each other. *Antp* cDNA G1100 (Laughon *et al.*, 1986), is derived from the P1 promoter and contains part of exon A and exons B, D, E, F, G and H (Fig. 1). *Antp* cDNAs cYE10 (Laughon *et al.*, 1986) and cYE12 are derived from the P2 promoter and contain part of exon C, as well as the exons D, E, F, G and H that are common to transcripts derived from the two promoters. The hybrids were digested with nuclease S1, electrophoresed on a 1.5% agarose gel, and viewed after ethidium bromide staining. First lane: cYE12⁺, G1100⁺. (In this figure, '+' orientation indicates that the 5' end of the cDNA adjoins the polylinker of the pEMBL vector). Since the cDNAs are in the same orientation, no hybridization occurs. Second lane: cYE12⁺, G1100⁻. Third lane: cYE12⁻, G1100⁺. Fourth lane: cYE10⁻, G1100⁺. G1100 and cYE10 appear to be identical throughout the coding region. M: λ HindIII-digested DNA + ϕ X174 HaeIII-digested DNA. (b) *Antp* transcripts utilize two alternative splice donors. Genomic sequence, and sequence from cDNAs G1100 and cYE12 are shown for the region around the G-H splice junction. Exon sequences are highlighted by underlining and overlining; optional exon sequence between the G_S and G_L splice donors is indicated by dotted lines on the genomic sequence. The four amino acids whose presence depends on use of the G_L splice donor are shown in italics. The genomic sequence is derived from Canton-S flies. Seventy-three bases of genomic sequence from iso-1 isogenic flies, including the exon G splice donors and 15 bases of adjacent intron, are identical. Sequence from the 5' end of exon H, derived from an iso-1 cDNA, is also identical. cDNA G1100 was isolated from the Oregon-R cDNA library of M.Goldschmidt-Clermont; cDNA cYE12 was isolated from the Oregon-R library of L.Kauvar. Note the unusual G_L splice donor and the unusual splice acceptor.

multiple roles in development, both temporally (Struhl, 1981; Kaufman and Abbott, 1984; Abbott and Kaufman, 1986) and in dorsal thorax and leg tissue (Abbott and Kaufman, 1986).

Could the different genetic functions performed by the *Antp* gene be mediated by different proteins? Alternative splicing is an important mechanism for generating diverse products from a single gene, and in many cases, particular alternatively spliced products are restricted to specific tissues or are developmentally regulated (reviewed in Leff *et al.*, 1986; Breitbart *et al.*, 1987). Two of the *Antp* protein-coding exons (Figure 1) are noteworthy: exon G uses an unusual splice donor, even though a more normal splice donor exists 12 bases upstream of it (Schneuwly *et al.*, 1986; Laughon *et al.*, 1986). Differential use of the two splice donors would add or delete four amino acids from the *Antp* protein. Exon F contains 39 bases; its removal would not alter the reading frame. Because the P1 and P2 promoters are independently regulated, we asked whether or not transcripts derived from the two promoters are spliced differently. In addition, we

investigated the relationship of *Antp* alternative splicing to polyadenylation site usage. By S1 nuclease analysis of *Antp* transcripts and cDNAs, we demonstrate the use of alternative splice donors, and the optional use of exon F to produce transcripts encoding four distinct but related *Antp* proteins. Alternative splicing of *Antp* transcripts changes during development, and alternative splicing occurs in transcripts derived from either promoter and utilizing either polyadenylation site.

Results

Antp cDNAs are derived from alternatively spliced transcripts

To determine whether *Antp* transcripts are alternatively spliced, the structure of *Antp* transcripts was investigated by conventional S1 analysis, and previously isolated *Antp* cDNAs (Scott *et al.*, 1983; Laughon *et al.*, 1986) were examined. *Antp* cDNAs were analyzed by S1 nuclease

Fig. 3. Developmental use of alternative splice donors at the 3' end of exon G. (a) Diagram of probe and expected S1 digestion products. The top of the diagram illustrates the probe used in the experiment. Sequences derived from *Antp* cDNA G1100 are black; within the *Antp* sequence, the G_L -H splice junction is represented by a solid white line, while the G_S splice donor is shown as a dotted white line. The probe contained 174 nt of *Antp* cDNA from the *Ava*II site at position 1843 on the *Antp* sequence (Laughon *et al.*, 1986) to the *Xba*I site in exon H. The remainder of the probe was pEMBL 18 sequence, shown as a striped bar. The inclusion of pEMBL 18 sequences permitted discrimination between bands resulting from reannealing of the probe DNA and bands indicative of alternatively spliced RNAs. The bold 5' and 3' refer to the orientation of the coding strand. The probe was end-labeled using [α^{32} P]dATP, [α^{32} P]dGTP and DNA polymerase I Klenow fragment at the *Ava*II site. End-labeled fragments protected from S1 nuclease digestion are shown at the bottom of the diagram, along with their sizes in nucleotides. 116 nt: G_S transcripts; 128 nt: precursor; 174 nt: G_L transcripts; 294 nt: reannealed probe. (b) Stage-specific use of G_S and G_L splice donors. G_L bands resulting from use of the G_L splice donor, 174 nt. G_S : bands resulting from use of the G_S splice donor 116 nt. P: bands resulting from precursor RNA 128 nt. 0-4: 0-4 h staged embryonic RNA. 4-8: 4-8 h staged embryonic RNA. 8-12: 8-12 h staged embryonic RNA. 12-16: 12-16 h staged embryonic RNA. 16-20: 16-20 h staged embryonic RNA. 20-24: 20-24 h staged embryonic RNA. L3: third instar larval RNA. P: pupal RNA. t: *Escherichia coli* tRNA. Larval RNA from iso-1 isogenic strain. All other RNAs from Canton-S. 10 μ g A^+ RNA used in each experiment. (c) Summary of exon G splice donor utilization. On the X-axis, 0-4, 4-8, 8-12, 12-16, 16-20, 20-24, L3, P refer to stages as in (b). On the Y axis, G_S and G_L refer to the G_S and G_L splice donors, respectively. P: precursor.

protection experiments to detect discontinuities between any two overlapping cDNAs. Pairs of oppositely oriented single-stranded cDNAs were annealed to form DNA-DNA duplexes, and the duplexes were digested with S1 nuclease (see Materials and methods). For any two cDNAs derived from a single class of transcript, only a single double-stranded DNA fragment should be protected from S1 nuclease. However, should the two cDNAs forming a duplex be derived from differently spliced transcripts, then an S1-sensitive discontinuity would be produced, and two smaller double-stranded fragments would be observed. The duplexes formed by annealing two *Antp* cDNAs, G1100 and cYE12, were cleaved into two fragments by S1 nuclease digestion, indicating differences between the cDNA sequences at one internal site (Figure 2a). Another P2-derived cDNA, cYE10, produced only a single band when annealed to G1100 and digested with S1 nuclease. The different 5' ends of the P1 and P2-derived cDNAs only reduce the length of the duplex formed between them; the 5' differences cannot account for the cleavage of the duplex into the two fragments seen in the cYE12-G1100 case. The

cYE12-G110 discontinuity was located by restriction mapping of the duplexes remaining after S1 nuclease digestion. Comparison of the sequences revealed that transcripts corresponding to cDNA cYE12 utilize a splice donor in exon G 12 bp upstream from that utilized by transcripts corresponding to the G1100 cDNA (Figure 2b). Use of the cYE12-like splice donor, G_S (for short exon G), avoids the extremely unusual AG/GAAAGT splice donor, G_L (long exon G), used by G1100-like transcripts. The use of alternative splice donors at the 3' end of exon G includes or omits four codons immediately adjacent to the homeobox in exon H (Figure 2b).

Additional *Antp* cDNAs that use the G_S splice donor, but lack exon F, were isolated from a cDNA library derived from iso-1 flies, a strain made isogenic for all four chromosomes.

Use of alternative splice donors in exon G is developmentally regulated

To determine whether or not the alternative splice donors at the 3' end of exon G are utilized in a stage-specific

Fig. 4. Developmental use of exon F. **(a)** Diagram of probe and expected S1 digestion products. As in Figure 3, *Antp* sequences are black, while pEMBL 18 sequences are striped. The probe contained 211 nt of *Antp* cDNA sequence between the *Ava*I site at position 1097 and the *Dde*I site at position 1817 in the genomic sequence (Laughon *et al.*, 1986). The fragment was end-labeled using [α^{32} P]dCTP, [α^{32} P]dTTP and Klenow fragment at the *Ava*I site. The resulting S1 digestion products are diagrammed below the probe. 405 nt: reannealed probe; 211 nt: F^+ transcripts; 98 nt: F^- transcripts. **(b)** Stage specific use of exon F. F^+ : bands resulting from transcripts containing exon F; F^- : bands resulting from transcripts lacking exon F. Stages as in Figure 3b. **(c)** Summary of exon F utilization. Axes as in Figure 3c. Experiments from which precursor data were derived are not shown.

manner, S1 nuclease protection experiments were performed using RNAs prepared from embryos at successive stages of development, and from larvae and pupae. The probe used (Figure 3a) was derived from the G1100 cDNA. Because both strands of the DNA fragment are present during the hybridization, some reannealing of the DNA may occur. The presence of the pEMBL 18 sequences in the probe permits the S1 digestion products due to reannealing to be distinguished from those resulting from hybridization of RNA to the DNA probe. Transcripts using the long form of exon G (G_L) form a continuous heteroduplex along the entire cDNA portion of the probe, while transcripts using the G_S exon form heteroduplexes with an S1 nuclease-sensitive discontinuity consisting of a 12 bp loop containing the DNA between the splice donors. The results of this experiment are shown in Figure 3b. The unusual G_L splice donor appears to be used much more frequently than the G_S splice donor during early embryonic development. The G_L splice donor is present in ~80% of transcripts up until about 12 h after fertilization, after which time the frequency of G_L -containing transcripts decreases steadily to ~23% in pupae (Figure 3c).

What appears to be alternative splicing could be due to polymorphism in the use of splice donors: transcripts made from one variant could use the G_S splice donor while transcripts made from another variant could use the G_L splice donor. S1 protection experiments were performed using unstaged embryonic RNA (data not shown) and larval

RNA (Figure 3b), prepared from flies isogenic for all four chromosomes (iso-1). Because both splice donors are used in the isogenic flies, the results are clearly due to alternative splice donors and not due to polymorphism. The genomic sequence around the splice donors (Figure 2b) is identical in both the iso-1 and Canton S strains. Sequence obtained from an iso-1-derived cDNA indicates that the two strains contain identical exon sequences adjacent to the exon H splice acceptor. Together with the S1 data, these results prove that the AG/GAAAGT splice donor is used.

Use of the 39 bp exon F is developmentally regulated

Omission of the 39 bp exon F from *Antp* transcripts deletes 13 codons from the mRNA without altering the reading frame. To determine whether or not use of the *Antp* 39 bp exon is developmentally regulated, S1 nuclease protection experiments similar to those described above were performed. RNA from different developmental stages was hybridized to an end-labeled fragment derived from G1100 (Figure 4a). Nuclease S1 digestion produced three bands: a 405 bp band resulting from the reannealed probe, a 211 bp band due to transcripts containing exon F and a 98 bp band corresponding to transcripts which either lack exon F or still contain the intron between exons E and F, I_{EF} (Figure 4b).

To show that the 98 bp band was due to mature transcripts lacking exon F and not due to precursors containing the I_{EF}

Fig. 5. (a) Diagram of S1 experiment to determine relationship between use of exon F and use of the G_L splice donor. The top of the diagram illustrates the structure of the probe used. The bold letters refer to *Antp* exon sequence, shown in black. Splice junctions are white lines between the exon sequences. The G_S splice donor is shown as a broken white line. As in previous figures, pEMBL sequences are striped, and the bold 5' and 3' refer to the orientation of the coding strand. The probe was labeled at the 5' end of the non-coding strand. Possible S1 digestion products are presented below the probe, with their sizes in nucleotides to the right. Exon G contains two in-frame splice acceptors the use of which would remove 75 or 108 bases from the 5' end of the exon. Transcripts spliced at these internal splice acceptors would produce bands of 163 and 196 bases in this S1 experiment. These bands were not observed, and are represented by broken lines. (b) G_L transcripts can contain or lack exon F; splice acceptors in exon G are not used. Ten micrograms poly(A)⁺ RNA from 4–8 h embryos (4–8), third instar larvae (L3) or 10 µg *E. coli* tRNA (t) were hybridized to the probe, digested with nuclease S1, and electrophoresed on a 6% acrylamide gel. After autoradiography, bands from F^+G_L , F^-G_L and G_S transcripts were detected, but no bands corresponding to use of splice acceptors within exon G were detected.

intron, S1 nuclease experiments were performed using a probe containing the 3' end of exon E, the 5' end of the I_{EF} intron, and pEMBL 18 sequences. These results (data not shown) indicate that precursors containing the I_{EF} intron never exceed 5% of *Antp* transcripts at any time during the course of development. Therefore, the 98 bp band shown in Figure 4b is principally due to transcripts lacking exon F; such mRNAs comprise 6–8% of *Antp* transcripts early

in development (until ~12 h after fertilization), after which time the percentage increases. About 40% of pupal transcripts lack exon F (Figure 4c). The use of RNA from isogenic flies demonstrates that the optional use of exon F is *bona fide* alternative splicing, and not the result of polymorphism.

Which of the possible alternative protein forms can be made?

Alternative splicing at exons F and G could create transcripts encoding four different *Antp* proteins; are all four transcripts actually made? To answer this question, a probe derived from cDNA G1100 was made, containing pEMBL 18 sequences, part of exon E, exons F, G_L and 44 bp of exon H (Figure 5a). cDNA G1100 contains exon F and uses the G_L splice-donor; such transcripts are symbolized F^+G_L . After hybridization to poly(A)⁺ RNA and S1 digestion, the probe protected three bands (Figure 5b): a 45 base band due to G_S -containing transcripts; a 271 base band resulting from G_L transcripts lacking exon F and a 368 base band due to the G_L transcripts that contain exon F. In conjunction with the cDNAs, this experiment demonstrates that all of the four possible *Antp* proteins can be made: F^+G_L , F^-G_L and G_S transcripts are observed, while F^+G_S and F^-G_S transcripts are represented as cDNAs.

To distinguish F^-G_S and F^+G_S transcripts, a similar experiment was attempted using a probe derived from cDNA cYE12 (F^+G_S) (data not shown). This experiment differed from the experiment using G1100 in a subtle but important way: in the G1100 experiment, G_S transcripts caused 12 bases of DNA to be looped out, where it was easily cut by nuclease S1. The cYE12 probe caused the G_L transcript—the RNA rather than the DNA—to be looped out; the probe itself was inefficiently cut by nuclease S1, rendering the experiment uninterpretable. This result indicates that if there are additional *Antp* sequences not represented in our cDNAs, such as another optional exon, these experiments will not detect them. Any additional *Antp* exons must be small, however, since the sums of the lengths of the known exons correspond well with the lengths of *Antp* transcripts.

Antp promoter usage and alternative splicing patterns vary independently during development

Does alternative splicing occur in *Antp* transcripts derived from both the P1 and P2 promoters? We have shown that alternative splicing within the *Antp* coding region produces transcripts which could encode four distinct proteins. The two *Antp* promoters are active in distinct spatial patterns (Jorgensen and Garber, 1987; A. Martinez-Arias, J.R. Bermingham, Jr and M.P. Scott, submitted). Do the transcripts produced from the two promoters produce distinct proteins as well? To determine whether or not the alternative splicing within the coding sequences of *Antp* is related to promoter usage, an S1 nuclease protection experiment was performed that permitted discrimination between alternatively spliced RNAs initiated at either of the two *Antp* promoters. The strategy for this experiment is illustrated in Figure 6a. Single-stranded DNA containing most of the G1100 *Antp* cDNA was hybridized to RNA, digested with S1 nuclease, electrophoresed, blotted and then probed with specific *Antp* sequences. The G1100 cDNA represents transcripts initiated at the P1 promoter. G1100 contains 55 nucleotides of exon A and all of exons B, C, E and F, and is derived from an

Fig. 6. (a) Diagram of S1 protection experiment examining the relationship between promoter usage and alternative splicing. The top of the diagram illustrates a hypothetical heteroduplex between *Antp* mRNA, shown as a squiggly black line, and ssDNA containing most of cDNA G1100. cDNA sequences are represented in black. In the construction used here, cDNA G1100 is truncated at a BglII site at position 903 in exon H (Laughon *et al.*, 1986) and cloned into pEMBL 19; pEMBL sequences are striped. Exons are labeled with bold letters; the G1100 cDNA contains exon F and exon G_L. R: EcoRI; Bg: BglII. Below the heteroduplex are shown possible S1 digestion products resulting from different RNA splicing patterns of transcripts derived from each of the *Antp* promoters. P1: P1-derived S1 digestion products; P2: P2-derived S1 digestion products. The sizes in nucleotides of the S1 digestion products are listed to the right. The probe consisted of 723 bp of exons D and E sequence for cDNA G1100 between the BspM1 sites at positions 114 and 989 on the *Antp* genomic sequence (Laughon *et al.*, 1986). (b) Relationship of alternative splicing to promoter usage. The autoradiograph illustrates the relationship of alternative splicing to promoter usage for RNAs from different developmental stages and tissues. 4-8: 4-8 h embryonic RNA; L3: 3rd instar isogenic larval RNA; P: pupal RNA; t: *E. coli* tRNA; D: undigested ssDNA; M: marker. Markers consisted of labeled ΦX174 *Hae*III digested DNA, and three additional labeled DNAs: the 2322 and 2027 bp DNAs from phage λ DNA digested with *Hind*III, and a 1733 bp *Eco*RI fragment from phage 2015 of the *Antennapedia* complex walk (Scott *et al.*, 1983). Ten micrograms RNA was used for each lane. *Drosophila* RNA was poly(A)⁺-selected. (c) *Antp* alternative splicing and promoter usage in embryonic nerve cords. 4-8: 25 µg total RNA from 4-8 h isogenic embryos. NC: 10 µg total RNA from nerve cords enriched from 12-16 h Oregon-R embryos. 12-16: 25 µg total RNA from 12-16 h Oregon-R embryos. t: 10 µg *E. coli* tRNA. D: undigested ssDNA. M: markers as in Figure 6b.

Table 1. Observed frequencies of specific *Antp* transcripts, and expected frequencies if promoter usage, polyadenylation and RNA splicing are independent

Transcript	4-8 h RNA		Larval RNA	
	Expected	Observed	Expected	Observed
P1 F ⁺ G _L	0.34	0.30	0.08	0.04
F ⁺ G _S	0.05	0.09	0.16	0.08
F ⁻	0.03	0.08	0.12	0.23
P2 F ⁺ G _L	0.37	0.31	0.14	0.07
F ⁺ G _S	0.06	0.10	0.29	0.27
F ⁻	0.03	0.11	0.21	0.28
A1 F ⁺ G _L	0.47	0.44	0.08	0.09
F ⁻ G _L	0.04	0.03	0.04	0.07
G _S	0.08	0.10	0.24	0.16
A2 F ⁺ G _L	0.24	0.17	0.14	0.09
F ⁻ G _L	0.02	0.03	0.07	0.08
G _S	0.04	0.10	0.44	0.40

The observed frequencies of specific classes of transcript were determined by densitometry of the autoradiographs similar to those shown in Figures 6b and 7b. The expected values were calculated assuming no relationship between promoter usage, alternative RNA splicing and polyadenylation site usage, using the frequencies listed in Table II.

mRNA in which the G_L splice donor was used. Thus, transcripts initiated at P1 should minimally protect a continuous stretch of DNA including the 55 nucleotides of exon A, exons B, D, E and, if the splicing pattern is like that represented in G1100, exons F, G_L and H as well. Transcripts initiated at P2, however, will not protect exons A and B, but they will protect DNA from exons D through H, if their splicing pattern within the coding region is

identical to that of G1100. Since the A and B exons are protected only by P1-derived transcripts, P2 transcripts protect shorter fragments of DNA than do P1 transcripts. RNAs that lack exon F and/or use the exon G_S splice donor will produce distinct bands protected from S1 nuclease digestion and, in each case, bands indicative of P2-derived transcripts are shorter than bands indicative of P1-derived transcripts. Regardless of which promoter is used, transcripts

Table II. Frequencies of alternative splices, promoter and polyadenylation site usage from end-labeled and probed S1 experiments

Experiment	P1		P2		F ⁺		F ⁻		G _S		G _L		A1		A2	
	4-8	L	4-8	L	4-8	L	4-8	L	4-8	L	4-8	L	4-8	L	4-8	L
P1-P2 probed S1	0.48	0.36	0.52	0.64	0.81	0.47	0.19	0.53	0.24	0.77	0.76	0.23				
A1-A2 probed S1					0.92	0.55	0.08	0.46	0.23	0.64	0.77	0.36	0.66	0.35	0.34	0.65
Exon F end-labeled S1					0.92	0.67	0.08	0.33								
Exon G end-labeled S1									0.12	0.68	0.77	0.32				

The frequencies of transcripts (from Table I) using a given sequence element (i.e. promoter, alternative splice or polyadenylation site) were summed up to determine the frequency of use of that sequence element. This table shows that the proportion of *Antp* transcripts using the P1 promoter is smaller in larvae than in 4-8 h embryos. While most *Antp* transcripts terminate at the A1 polyadenylation site early in development, the reverse is true later in development. For comparison, the frequencies of transcripts containing or lacking exon F from the end-labeled S1 experiment in Figure 4, and the frequencies of transcripts using the G_L or G_S splice donors for the end-labeled S1 experiment in Figure 3 are also shown. The good agreement between the data derived from the two types of S1 nuclease protection experiment indicates that potential sources of error in the probed S1 experiments introduced by transfer to zetaBind filters and probing, etc. are not significant. The discrepancy between probed and end-labeled S1 experiments concerning the frequency of G_S transcripts can be attributed to the fact that precursor RNAs detected in the end-labeled S1 experiments would be detected as G_S transcripts in the probed S1 experiments. Similarly, precursor RNAs containing either of the introns flanking exon F may be detected as F⁻ transcripts in the probed S1 experiments.

lacking exon F, and/or using the G_S splice donor, will protect identical fragments between the site(s) of alternative splicing and the 3' end of the ss cDNA. These downstream sequences are therefore uninformative regarding the relationship between alternative splicing and promoter usage. The protected DNAs, after electrophoresis and blotting, were detected by probing with a fragment derived from exons D and E, located upstream of the alternative splice sites. Only fragments that are informative with respect to the relationship between alternative splicing and promoter usage are detected by this probe (Figure 6a).

RNA purified from different stages and tissues were used in this probed S1 experiment; the results are shown in Figure 6b. Transcripts from the P1 and P2 promoters which contain exon F and utilize the G_L splice donor (P1F⁺G_L; P2F⁺G_L) protect bands of 2023 and 1886 nt respectively. These bands are prominent in the embryonic RNA lane, but fainter in the larval RNA lane, as expected since use of both exon F and the G_L splice donor decreases as development proceeds. P1 and P2 transcripts containing exon F but using the G_S splice donor (P1F⁺G_S; P2F⁺G_S) protect 1419 and 1282 nt bands respectively. Because use of the G_S splice donor is more frequent in larvae than in early embryos, the P1F⁺G_S and P2F⁺G_S RNAs should constitute a greater proportion of *Antp* transcripts in larvae than in 4-8 h embryos. This is clearly the case for P2F⁺G_S transcripts, but is less obvious for P1F⁺G_S transcripts. As expected, P1F⁻ transcripts which protect an 1166 nt band, and P2F⁻ transcripts which protect a 1029 nt band, are more abundant in larval RNA than in 4-8 h embryonic RNA.

The intensities of the bands in each lane were measured by densitometry. Table I shows the observed frequencies for the six types of *Antp* transcript detected in the experiment using 4-8 h embryonic and larval RNA. Two important pieces of information are revealed by these experiments. First, each promoter produces at least three forms of alternatively spliced transcripts (this experiment does not discriminate F⁻G_L and F⁻G_S forms of the RNAs). Second, the relative levels of *Antp* promoter usage appear to change during development: while mRNA derived from the P1 and P2 promoters are almost equally abundant in 4-8 h embryos, P2-derived mRNAs are twice as abundant as P1-derived mRNAs in third instar larvae (see Table II). In pupae, P1 transcripts appear to be more abundant (Figure 6b). Table I also lists expected frequencies for the different

types of *Antp* transcripts. These values were calculated assuming that promoter usage, optional use of exon F, and alternative splicing at the 3' end of exon G are independent events. The good agreement between observed and expected values indicates that this assumption is valid. Discrepancies between observed and expected frequencies of specific transcripts are probably due to experimental error. Note that the data do not eliminate the possibility that in specific cells or tissues a single promoter and/or alternative splice pattern is used exclusively. Additionally, while the relative frequencies of promoter usage and the patterns of alternative splicing are changing in concert, the magnitudes of the changes are different, indicating that the events are not tightly coupled to one another. Table II displays data for the frequency of F⁺, F⁻, G_S, and G_L transcripts, derived from both the probed (Figures 6b and 7b) and end-labeled (Figures 3b and 4b) S1 nuclease protection experiments. Because the two types of S1 experiments are in good agreement with each other, the probed S1 experiments probably present an accurate picture of *Antp* alternative splicing.

Antp splicing patterns in RNA from embryonic nerve cords were examined to determine whether *Antp* is spliced differently in nervous tissue. RNA was purified from 12-16 h embryonic tissue highly enriched in nerve cords. For comparison, RNA was also purified from whole 12-16 h embryos. At this stage, *Antp* is predominantly expressed in nervous tissue (Levine *et al.*, 1983; Carroll *et al.*, 1986; Wirz *et al.*, 1986). Figure 6c shows that minimal differences exist between *Antp* splicing patterns in nervous tissue and in embryos as a whole, indicating that the splicing pattern observed for whole embryos in Figure 6c is predominantly due to nerve cord expression. These data suggest that the observed changes in *Antp* RNA splicing could result from the development of tissues that splice *Antp* transcripts differently.

Alternative splicing choice is independent of polyadenylation site usage

Antp transcripts are polyadenylated at two major sites, A1 and A2 (Laughon *et al.*, 1986; Figure 1). The functional significance (if any) of the differential sites is obscure, but the downstream (A2) site is preferentially used in cultures of primary neuroblasts (O'Connor *et al.*, 1988; J.R. Bermingham, Jr, unpublished results). We wished to determine whether alternative spliced *Antp* transcripts can use either

Fig. 7. (a) Diagram of S1 protection experiment examining the relationship between polyadenylation site usage and alternative splicing. As in Figure 6a, the top of the diagram illustrates a hypothetical heteroduplex between *Antnp* mRNA and ssDNA. In this case, sequences from *Antnp* cDNA G1100 were ligated to genomic sequences containing most of exon H, to produce a hybrid cDNA containing the A1 polyadenylation site (see Fig. 1). A 658 bp *Bst*XI-*Xba*I fragment from G1100 containing part of exon E, exons F and G, and part of exon H was fused to a 1377 bp *Xba*I-*Eco*RI fragment consisting entirely of part of exon H. The *Xba*I-*Eco*RI fragment was subcloned from phage 1815 of the *Antennapedia* complex walk (Scott *et al.*, 1983). The resulting 2035 bp *Antnp* cDNA was cloned into pEMBL 18 to produce ssDNA. Exons A, B, D and part of E were excluded from the ssDNA and therefore transcripts derived from either promoter will produce the same pattern of bands. This was done to prevent unduly complicating the experiment. No information was lost however, because RNA blot analysis has shown that both promoters utilize both polyadenylation sites (Laughon *et al.*, 1986). BX: *Bst*XI; R: *Eco*RI; A1: A1 polyadenylation site; *: putative polymorphism. Possible S1 digestion products are diagrammed below the heteroduplex, with their sizes listed on the right. The probe XR contains 778 bp of exon H sequence from cDNA G1100 between the *Xba*I site at position 356 on the *Antnp* sequence of Laughon *et al.* (1986) and an *Eco*RI site at the 3' end of the cDNA. To identify F⁻G₅ transcripts, and to determine the origin of polymorphism-derived bands, the filter was stripped and reprobed with a DNA fragment, DB, containing 366 bp of sequence from cDNA G1100 between a *Dde*I site at position 1817 in exon G (Laughon *et al.*, 1986) and a *Bam*HI site at position 516 in exon H. (b) Relationship of alternative splicing to polyadenylation site usage. The RNAs used in each lane were identical to those used in Figure 6b. The splice patterns and sizes of bands are given to the left of the autoradiograph. The blot was hybridized to probe XR, stripped, and rehybridized to probe DB. As indicated above the autoradiographs, the left-hand lanes were probed with DB while the right-hand lanes were probed with XR. The DB fragment detects a 214 nt band that must be derived from F⁻G₅ transcripts. Control lanes were identical to those in Figure 6b except for D, undigested ssDNA. (c) *Antnp* alternative splicing and polyadenylation in embryonic nerve cords. RNA lanes as in Figure 6c. Selected bands are labeled to the left of the autoradiograph. The blot was hybridized to probe DB. Control lanes were identical to those in Figure 7b.

polyadenylation site. As in the experiment just described, ssDNA was hybridized to RNA, digested with nuclease S1, and the digestion products were analyzed by electrophoresis, blotting and hybridization to a probe chosen to detect the informative fragments. To perform this experiment, an *Antnp* cDNA containing at least the upstream (A1) polyadenylation site was required. *Antnp* genomic DNA consisting entirely of exon H sequences was substituted for the 3' end of cDNA G1100 to provide the A1 polyadenylation site, and enough DNA downstream of it to detect transcripts using the A2 polyadenylation site. The experiment and its expected S1 nuclease digestion products are diagrammed in Figure 7a.

The ssDNA was hybridized to RNAs from different stages and tissues, digested with nuclease S1, electrophoresed and blotted. The protected DNAs were probed with labeled DNA derived from exon H of cDNA G1100. This labeled probe, XR (Figure 7a), lies between the 5' end of exon H and the A1 polyadenylation site; it will hybridize with equal efficiency to fragments resulting from use of either polyadenylation site and will not hybridize to fragments that provide no information about the relationship between *Antnp* alternative splicing and polyadenylation.

The results from the experiment are shown in Figure 7b. Because the ssDNA extends further downstream than the A1 polyadenylation site, transcripts using the A2 polyadenylation site will protect longer segments of DNA than those using the A1 site. Transcripts containing exon

F and using the G_L splice donor are, as expected, more abundant in embryos than in larvae. These transcripts protect a 2032 nt band if they utilize the A2 polyadenylation site, and a 1491 nt band if they utilize the A1 polyadenylation site. Depending upon which polyadenylation site is used, transcripts lacking exon F but using the G_L splice donor protect 1648 nt (F⁻G_LA2) or 1107 nt (F⁻G_LA1) bands. Neither band was easily detectable in embryonic RNA, but both were present in larval RNA lanes. Transcripts using the G_S splice donor protect bands of 1422 nt (G_SA2) or 881 nt (G_SA1). The abundance of each type of transcript was measured by densitometry of the bands in the autoradiograph shown in Figure 7b. These data are summarized in Table I. We find that all three detectable classes of alternatively spliced transcript use both polyadenylation sites. By summing up the frequencies of transcripts due to each of the polyadenylation sites (Table II), it can be seen that the A1 polyadenylation site is used twice as frequently as the A2 site in 4-8 h embryos, while in larvae the reverse is true.

Bands of 1140 nt, 600 nt, 520 nt and 280 nt were unexpected. These bands were generally of lower intensity than the expected bands. We believe that the four unexpected bands originate from a discontinuity located in the 3' untranslated region, about 280 nt downstream of the 5' end of exon H. The ssDNA in this region contained genomic sequences derived from phage 1815 of the *Antennapedia* complex walk (Scott *et al.*, 1983). Because similar bands

were not observed in the previous S1 experiment, in which the ssDNA consisted entirely of cDNA sequences, it is likely that the discontinuity results from the genomic DNA. The DB fragment (Figure 7a) hybridized to the 1107 nt F⁻G_L A1 band, but not to the 1140 nt band. This result confirms the identity of the F⁻G_L A1 band, and demonstrates that the 1140 nt band derives from the 3' end of the ss cDNA probe, specifically sequences protected by transcripts using the A2 polyadenylation site. The 280 nt band must derive from the region of overlap of the XR and DB fragments because it is detected by both. Hence it originates from the 5' end of exon H. These observations are consistent with the hypothesis that a polymorphism between the cloned genomic DNA and the RNA causes some S1 nuclease cleavage in exon H of the resulting heteroduplexes. The 280 nt and 1140 nt bands are created by cleavage of the 1422 nt G_S A2 band; similarly, the 280 nt band and the 600 nt band seen with probe XR correspond to the 881 nt G_S A1 band. The 520 nt band is probably derived from F⁻G_L transcripts while an expected 890 nt band derived from F⁺G_L transcripts is obscured by the 881 nt G_S A1 band. Bands X and Y in the pupal lanes (Figure 7b) appear to result from as yet uncharacterized pupal- and adult-specific splicing patterns, because they appear in experiments using isogenic adult RNA (data not shown) but not in experiments using isogenic larval RNA (Figure 7b). Both bands contain homeobox sequences at the 5' end of exon H because they are detected by both the XR and DB probes.

The relationship between alternative splicing and polyadenylation site usage in nerve cord RNA was also examined. Figure 7c shows that the A2 polyadenylation site is preferentially used in nerve cord RNA. A similar result was obtained using RNA from cultured neuroblasts (data not shown). As in the previous experiment, no differences were observed between the splicing and polyadenylation patterns of nerve cord RNA and RNA from whole embryos.

Discussion

Control of *Antp* alternative splicing

The experiments presented here have measured promoter usage, use of exon F, the use of alternative splice donors at the 3' end of exon G, and polyadenylation site utilization in embryos and larvae. *Antp* promoter and polyadenylation site utilization and alternative splicing all appear to vary independently during development. Four possible models could explain the control of alternative splicing of *Antp* transcripts. First, splicing could be primary transcript-specific, possibly controlled by different RNA secondary structures of transcripts with different 5' or 3' ends. Since alternative splicing appears to be independent of promoter or polyadenylation site choice, this model is unlikely to apply. Second, the alternative splicing of *Antp* transcripts could be stochastic; the splicing apparatus would randomly choose among the different splicing pathways. Changes during development in the patterns of RNA splicing render this explanation unlikely. Third, it is possible that the pattern of *Antp* alternative splicing is simply the by-product of developmental changes in the splicing apparatus itself. *Antp* splice site choice would not be specifically regulated, but because the splicing apparatus itself is changing, the choices would not be random. A fourth possibility is that *Antp* alternative splicing is regulated differently in different tissues.

Some tissues may express one or more forms of alternatively spliced *Antp* transcripts, while other tissues express other forms, and the pattern seen in the experiments presented here are the sums of a variety of tissue-specific expression patterns. The patterns of alternative splicing at exons F and G probably result from factors interacting with *Antp* primary transcripts. It is not clear whether such factors regulate *Antp* alternative splicing *per se*, or perform more general functions.

The optional use of exon F may involve specific sequences within *Antp* primary transcripts. Sequestering an exon within the loop of a hairpin structure can result in skipping of that exon (Solnick and Lee, 1987). Several pairs of small inverted repeats (the largest is 14/17 base pairs, five of which are G-T) within the introns flanking exon F could potentially sequester the optional exon within hairpin-loop structures. However, if such secondary structures are important for alternative splicing of exon F, their stability or their effects must be altered by *trans*-acting factor(s) that are themselves regulated in developmental or tissue-specific patterns. Sequences within an optional exon of the human fibronectin gene are necessary for inclusion of that exon in mature mRNA (Mardon *et al.*, 1987); similar sequences could exist within *Antp* exon F.

Specific sequences may be involved in dictating the choice of exon G splice donors. The bizarre AG/GAAAGU splice donor used by the long form of exon G is, to our knowledge, the first reported case of a naturally occurring AG/GA being used as a splice donor instead of the canonical AG/GU (Mount, 1982; Shapiro and Senapathy, 1987). *In vitro*, the engineered sequence GA- at the 5' end of the large intron of rabbit β -globin has been found to permit 5' cleavage at reduced efficiency, but to block 3' cleavage, resulting in the accumulation of lariat intermediates (Aebi *et al.*, 1986, 1987). It should be noted that the pyrimidine tract at the 3' end of the intron between exons G and H is permeated by an unusual number of purine residues. Recently, it has been shown that increasing the proportion of purine residues in the polypyrimidine stretch of the SV40 early pre-mRNA splice acceptor influences the choice of alternative splice donors (Fu *et al.*, 1988). Perhaps a special splicing apparatus is required to handle the unusual branch structure produced by splicing from this splice donor and/or to recognize the unusual splice acceptor.

Functions of multiple *Antp* proteins and/or transcripts

What biological role might alternative splicing play in the function(s) of the *Antp* gene? Alternative splicing produces *Antp* transcripts encoding four possible proteins differing from one another by 4, 13 or 17 amino acids. It is conceivable that alternative splicing of *Antp* transcripts has no function, that it merely reflects developmental changes in the *Drosophila* splicing apparatus, and that all four proteins are functionally equivalent. Another possibility is that only one or two of the four proteins is functional, and that controlled splicing determines whether or not an active protein is produced. The *Drosophila tra* gene uses a sex-specific alternative splice acceptor to produce functional *tra* protein only in females (Boggs *et al.*, 1987). A third possibility is that all four proteins are functional but in different ways.

It is possible that the alternatively spliced transcripts themselves function differently. For example, alternative RNA structures could impart differences in transcript stabi-

ty. Possibly the alternatively spliced transcripts differ in their rates of transport from the nucleus, or in their rates of translation. Such differences could result in altered levels of *Antp* protein.

The four *Antp* proteins may have significant structural differences. One of the optional four amino acids encoded by the G_L exon is a cysteine. Since the *Antp* protein has only three other cysteine residues, the presence or absence of the fourth Cys residue could rearrange intramolecular disulfide bonds, drastically altering the shape of the protein. The optional cysteine could also affect the formation of intermolecular disulfide bonds. For example, the secreted form of IgM, but not the alternative membrane form, contains a cysteine used to form a disulfide bond with the J segment (Alt *et al.*, 1980; Rogers *et al.*, 1980). The *Ultrabithorax* (*Ubx*) gene also uses alternative splice donors to encode alternative proteins that differ by nine amino acids, one of which is a cysteine (Beachy, 1986; O'Connor *et al.*, 1988). *Ubx* also contains two 51 bp 'microexons' (Beachy *et al.*, 1985) that are optionally used (O'Connor *et al.*, 1988). Significant alterations in *Antp* and *Ubx* protein structures need not, of course, be dependent on disulfide bond formation or just on the cysteines in the optional sequences; other residues may be equally or more important. Exon F, for example, contains two threonine residues and a tyrosine residue which could serve as sites for polyadenylation.

The structural differences in the alternative *Antp* proteins could affect *Antp* function(s). Proteins that contain homeodomains, including *Antp*, have been hypothesized to be sequence-specific DNA binding proteins *in vivo* (Laughon and Scott, 1984) and have been shown to be capable of DNA sequence recognition *in vitro* (Desplan *et al.*, 1985; A. Laughon, S. Hayashi and M.P. Scott, unpublished data). *Antp* DNA binding may be used to regulate transcription of (unknown) genes; the different *Antp* proteins could possess different DNA binding affinities. Perhaps the different *Antp* proteins engage in distinct protein-protein interactions. The various *Antp* proteins could interact differently with transcription complexes to selectively control subsets of the genes regulated by *Antp* during development. Activation of transcription of the *GAL1* gene in yeast requires acidic 'activating regions' on the *GAL4* protein (Ma and Ptashne, 1987). The optional use of exon F adds or deletes three acidic amino acids from the middle of an otherwise uncharged amino acid region of the protein, and thereby could provide an acidic contact point.

The *Antp* gene performs multiple functions during development (Struhl, 1981; Kaufman and Abbott, 1984; Abbott and Kaufman, 1986). Could multiple proteins produced by alternative splicing be responsible for the different functions? While much of the genetic complexity of *Antp* can be explained by the different tissue-specific patterns of expression of the two *Antp* promoters, the multiple products produced by alternative splicing of *Antp* transcripts could provide another important level of control. P1 and P2 transcripts are expressed in temporally and spatially distinct patterns, but their patterns of expression are not perfectly correlated with their putative genetic functions (Jorgensen and Garber, 1987; Abbott and Kaufman, 1986), suggesting the existence of additional mechanisms of *Antp* regulation. Whether multiple *Antp* functions, defined genetically, are related to alternative

splicing of *Antp* transcripts awaits improved understanding of any distinct molecular functions of the different *Antp* proteins and/or transcripts.

The time during development at which the patterns of *Antp* alternative splicing start to shift corresponds with the differentiation of specific neurons (Thomas *et al.*, 1984; Campos-Ortega and Hartenstein, 1985; Ghysen *et al.*, 1986). This correspondence suggests a possible role for the alternatively spliced forms of *Antp* in neural development. Several examples of neural-specific splicing have been found, including the *Ddc* gene in *Drosophila* (Morgan *et al.*, 1986), the calcitonin-CGRP gene in mammals (Leff *et al.*, 1987; Crenshaw *et al.*, 1987), and the *Aplysia* R15 polypeptide RNA (Buck *et al.*, 1987). We find that no *Antp* splicing pattern is specifically included or excluded from nervous tissue. However, our results are consistent with the possibility that the ratios of the various alternatively spliced *Antp* transcript differ between neural and non-neural tissue, and that the observed patterns of *Antp* alternative splicing result from the development of tissues expressing differing ratios of alternatively spliced *Antp* transcripts. It appears that the A2 polyadenylation site is preferentially used in neural tissue. Similarly, the downstream polyadenylation site of *Ultrabithorax* is preferentially used in cultured neuroblasts (O'Connor *et al.*, 1988). While the significance of these observations is presently unclear, it is noteworthy that downstream polyadenylation sites are also used in neural tissue for two other genes with multiple polyadenylation sites: the calcitonin-CGRP gene (Amara *et al.*, 1984) and the mouse NCAM gene (Barbas *et al.*, 1988).

Expression of the *Antp* gene is potentially controlled at multiple levels

Studies on the organization and expression of *Antp* have revealed a variety of mechanisms by which the gene could be controlled. First, the gene utilizes two promoters. The promoters, which are separated by 68 kb of DNA, are controlled differently because they can function independently (Jorgensen and Garber, 1987), and because their patterns of expression are spatially different (Jorgensen and Garber, 1987; A. Martinez-Aria, J.R. Bermingham, Jr and M.P. Scott, submitted) and temporally different (this paper). Second, the gene contains a number of large introns; ~6 kb of exon DNA is distributed over 102 kb of genomic DNA. It has been suggested by D. Hogness (personal communication) that large introns could serve to control the timing of expression of genes that possess them, by increasing the time needed to transcribe the entire gene. Third, transcripts from each promoter contain long 5' untranslated regions containing many AUGs (Schneuwly *et al.*, 1986; Stroehrer *et al.*, 1986; Laughon *et al.*, 1986). The existence of such long leader sequences suggests that *Antp* transcripts could be subject to translational control, and because transcripts derived from the two promoters possess different 5' exons, translational control could be transcript-specific. Fourth, the gene utilizes alternative splicing to create at least four distinct proteins. The challenge remains to show what roles these mechanisms (or a subset thereof) play in *Antp* gene function during development.

Materials and methods

S1 nuclease detection of cDNA heterogeneity

Antp cDNAs (Laughon *et al.*, 1986) were cloned in each orientation into

the plasmid pEMBL 8⁺, a vector which permits recovery of single-stranded plasmid DNA upon infection with phage f1 (Dente *et al.*, 1983). Approximately 660 ng of each of two opposite oriented ssDNAs was added to 20× SSC (3 M NaCl, 0.3 M NaCitrate, pH 7) to give a mixture that was 6× SSC in a total volume of 23.6 µl. The mixture was heated to 90°C for 5 min, then incubated at 65°C for 2 h. After the incubation, 215 µl of 1× S1 buffer (30 mM NaAcetate, pH 4.5, 250 mM NaCl, 10 mM ZnSO₄, 5% glycerol) was added to the hybridization mixture, followed by 1 µl of Boehringer Mannheim nuclease S1 diluted to 23.3 units/µl with 1× S1 buffer. The DNAs were digested for 30 min at 37°C. Following S1 digests, DNAs were ethanol precipitated, resuspended, electrophoresed on a 1.5% agarose gel, ethidium bromide stained and viewed under UV light.

Nerve cord enrichment

Nerve cords were enriched from 12–16 h Oregon-R embryos using a method developed by N. Patel and C. Goodman (Stanford University). Embryos were disrupted by grinding between two frosted glass plates. Nerve cords were separated from other tissues by several settlings through cold PBS (16 mM K⁺, 10 mM PO₄²⁻, 280 mM NaCl, pH 7.2) and were stored frozen at –80°C prior to RNA purification.

RNA purification

RNAs were purified from Canton-S flies (unless otherwise noted) or from iso-1, a strain made isogenic for all four chromosomes by Dr James Kennison. The extraction procedure used was similar to that of Cathala *et al.* (1983). Embryos, larvae or pupae were homogenized in 5 M guanidine thiocyanate, 50 mM Tris (pH 7.5), 10 mM EDTA, 1.4 M β-mercaptoethanol and then centrifuged at 16 000 g for 10 min at 4°C. Lithium chloride (6 M) was added to the supernatant until the final concentration was 3.3 M. After storage overnight at 4°C, the RNA solution was centrifuged at 16 000 g for 30 min at 4°C. The pellet was washed by resuspension in 3 M lithium chloride–4 M urea and centrifugation at 16 000 g for 30 min at 4°C. The pellet was resuspended for an hour or more at room temperature in 10 mM Tris (pH 8)–1 mM EDTA, 0.1% SDS and then extracted two to three times with 24:24:1 phenol–chloroform–isoamyl alcohol (vol/vol/vol). The long resuspension time was necessary for good yields. The RNA was precipitated at –20°C after addition of 0.05 vol 3 M sodium acetate and 2.5 vol of ethanol. Poly(A)⁺ RNA was selected using oligo(dT) cellulose as described in Maniatis *et al.* (1982).

S1 nuclease analysis of alternatively spliced transcripts

S1 nuclease analysis was performed either by using double-stranded fragments according to the procedure of Berk and Sharp (1977), or by using single-stranded DNAs according to the procedure of Nasmyth *et al.* (1980).

To detect the use of alternative splice donors at the 3' end of exon G, the optional use of exon F, or both, α³²P-labeled DNA fragments containing both *Antp* cDNA and pEMBL 18 vector sequences were used as probes. The inclusion of pEMBL 18 sequences permitted discrimination between bands resulting from reannealing of the probe DNA and bands indicative of alternatively spliced RNAs. Each probe was mixed with 10 µg of poly(A)⁺ selected RNA, phenol–chloroform extracted, resuspended in 80% formamide, 400 mM NaCl, 40 mM Pipes, pH 6.4, 1 mM EDTA, heated to 65°C for 5 min, then hybridized at 56°C for the exon G fragment, 57.5°C for the exon F fragment, or 55°C for the exon F + G fragment. These temperatures were empirically determined to maximize RNA–DNA hybridization, while minimizing DNA–DNA reannealing. After 12–15 h of hybridization, the samples were mixed with 250 µl cold S1 buffer containing 75–100 units S1 nuclease, and digested at 37°C for 30–40 min. After ethanol precipitation, the reaction products were electrophoresed on 7.8 M urea 0.5× TBE (90 mM Tris, 90 mM borate, 2.5 mM EDTA, pH 8.3) 5% acrylamide gels. After electrophoresis, the gels were dried onto Whatman 3MM paper, and autoradiographed using Kodak X-AR film at –80°C with an enhancing screen for up to 8 days.

S1 nuclease protection experiments to examine the relationship between alternative splicing at exons F and G and promoter or polyadenylation site choice were performed using unlabeled ssDNA as a probe. The ssDNA was mixed with 10 µg of poly(A)⁺ RNA. After phenol–chloroform extraction and ethanol precipitation, the nucleic acids were resuspended in 17.5 µl 0.1% diethylpyrocarbonate-treated distilled water. Eight microliters of 2.5 M NaCl, 31.3 mM Pipes pH 7.7, and 31.3 mM EDTA were added as described by Nasmyth *et al.* (1980). The solution was heated to 85°C for 5 min, and hybridized at 60°C for 12–15 h. S1 nuclease digestion was performed as described above. The S1 digestion products were electrophoresed on a sequencing gel that was 2.5% acrylamide on the upper half and a 2.5–12% gradient on the lower half. The choice of gel system permitted high resolution of bands in the 1800–2000 nt range, while retaining bands of lower mol. wt. By a variation of the method of Church and Gilbert (1984),

the DNAs were electrophoretically transferred from the gel onto zeta-bind filters (AMF–CUNO) in 1× TBE at 2 A constant current for 1 h using a Hoefer model TE Western blotting apparatus. The filters were pre-hybridized at 65°C in 6× SSC, 0.5% SDS, 0.5 mg/ml sheared salmon sperm DNA, 5× Denhardt's and hybridized at 65°C, in 6× SSC, 0.5% SDS, 0.5 mg/ml sheared salmon sperm DNA, 5× Denhardt's, 10 mM EDTA, 10% dextran sulfate, similar to Maniatis *et al.* (1982). The fragments protected from S1 nuclease digestion were detected using DNA probes labeled to high specific activity (1 × 10⁸–2 × 10⁸ c.p.m./µg) using Klenow fragment, [α³²P]dCTP (3000 Ci/mmol; NEN), and random primers (Pharmacia) by the method of Feinberg and Vogelstein (1983). After hybridization, filters were washed several times in 0.1× SSC, 0.1% SDS for ~4 h at 65°C. Filters were autoradiographed using Kodak X-AR film at –80°C for 12–60 h using an enhancing screen. Signals on the autoradiographs were quantified using a Hoefer model GS300 densitometer, and analyzed using Hoefer software on an IBM PC, courtesy of Dr Tom Cech.

Acknowledgements

We wish to thank Drs Allen Laughon and John Tamkun for valuable discussions and criticisms during the course of this work. We thank Dr James Kennison for providing the iso-1 isogenic flies, Dr John Tamkun for the iso-1 genomic and cDNA libraries, Dr Hugh Brock for isolating iso-1 cDNAs, Dr Anthony Mahowald for cultured neuroblast RNA, and Drs Mike O'Connor and Welcome Bender for communication of results prior to publication. We also wish to thank Drs John Tamkun, Allen Laughon and Karla Kirkegaard for critical readings of the manuscript, and Cathy Inouye for her expert secretarial assistance. This work was supported by NIH grant No. 18163 to M.P.S.

References

- Abbott, M.K. and Kaufman, T.C. (1986) *Genetics*, **114**, 919–942.
- Aebi, M., Hornig, H., Padgett, R.A., Reiser, J. and Weissmann, C. (1986) *Cell*, **47**, 555–565.
- Aebi, M., Hornig, H. and Weissmann, C. (1987) *Cell*, **50**, 237–246.
- Akam, M. (1987) *Development*, **101**, 1–22.
- Alt, F.W., Bothwell, A.L.M., Knapp, M., Siden, E., Mather, E., Koshland, M. and Baltimore, D. (1980) *Cell*, **20**, 293–301.
- Amara, S.G., Evans, R.M. and Rosenfeld, M.G. (1984) *Mol. Cell. Biol.*, **4**, 2151–2160.
- Beachy, P.A., Helfand, S.L. and Hogness, D.S. (1985) *Nature*, **313**, 545–551.
- Beachy, P. (1986) Ph.D. Thesis, Stanford University.
- Barbas, J.A., Chaix, J.-C., Steinmetz, M. and Goriadis, C. (1988) *EMBO J.*, **7**, 625–632.
- Berk, A.J. and Sharp, P.A. (1977) *Cell*, **12**, 721–732.
- Boggs, R.T., Gregor, P., Idriss, S., Belote, J.M. and McKeown, M. (1987) *Cell*, **50**, 739–747.
- Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987) *Annu. Rev. Biochem.*, **56**, 467–495.
- Buck, L.B., Bigelow, J.M. and Axel, R. (1987) *Cell*, **51**, 127–133.
- Campos-Ortega, J.A. and Hartenstein, V. (1985) *The Embryonic Development of Drosophila melanogaster*. Springer, Berlin.
- Carroll, S.B., Laymon, R.A., McCutcheon, M.A., Riley, P.D. and Scott, M.P. (1986) *Cell*, **47**, 113–122.
- Cathala, J.A., Savouret, J.-F., Mendez, B., West, B.L., Karin, M., Martial, J.A. and Baxter, J.D. (1983) *DNA*, **2**, 329–335.
- Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
- Crenshaw, E.B., III, Russo, A.F., Swanson, L.W. and Rosenfeld, M.G. (1987) *Cell*, **49**, 389–398.
- Denell, R.E. (1973) *Genetics*, **75**, 279–297.
- Denell, R.E., Hummels, K.R., Wakimoto, B.T. and Kaufman, T.C. (1981) *Dev. Biol.*, **81**, 43–50.
- Dente, L., Cesareni, G. and Cortese, R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.
- Desplan, C., Theis, J. and O'Farrell, P.H. (1985) *Nature*, **318**, 630–635.
- Duncan, I.M. (1986) *Cell*, **47**, 297–309.
- Feinberg, A. and Vogelstein, B. (1983) *Anal. Biochem.*, **132**, 6–13.
- Frischer, L.E., Hagen, F.S. and Garber, R.L. (1986) *Cell*, **47**, 1017–1023.
- Fu, X.-Y., Ge, H. and Manley, J.L. (1988) *EMBO J.*, **7**, 809–817.
- Garber, R.L., Kuroiwa, A. and Gehring, W.J. (1983) *EMBO J.*, **2**, 2027–2034.
- Gehring, W.J. (1987) *Science*, **236**, 1245–1252.
- Gehring, W.J. and Hiromi, Y. (1986) *Annu. Rev. Genet.*, **20**, 147–173.

- Ghysen, A., Dambly, C., Aceves, E., Jan, L.Y. and Jan, Y.N. (1986) *Roux's Arch. Dev. Biol.*, **195**, 281–289.
- Hazelrigg, T. and Kaufman, T.C. (1983) *Genetics*, **105**, 581–600.
- Ingham, P.W. and Martinez-Arias, A. (1986) *Nature*, **324**, 592–597.
- Jorgensen, E.M. and Garber, R.L. (1987) *Genes Dev.*, **1**, 544–555.
- Kaufman, T.C. and Abbott, M.K. (1984) In Malacinski, G.M. and Klein, W.H. (eds), *Molecular Aspects of Early Development*. Plenum, New York, pp. 189–218.
- Kaufman, T.C., Lewis, R. and Wakimoto, B. (1980) *Genetics*, **94**, 115–133.
- Laughon, A. and Scott, M.P. (1984) *Nature*, **310**, 25–31.
- Laughon, A., Boulet, A.M., Bermingham, J.R., Jr., Laymon, R.A. and Scott, M.P. (1986) *Mol. Cell. Biol.*, **6**, 4676–4689.
- Leff, S.E., Rosenfeld, M.G. and Evans, R.M. (1986) *Annu. Rev. Biochem.*, **55**, 1091–1117.
- Leff, S.E., Evans, R.M. and Rosenfeld, M.G. (1987) *Cell*, **48**, 517–524.
- Levine, M., Hafen, E., Garber, R.L. and Gehring, W.J. (1983) *EMBO J.*, **2**, 2037–2046.
- Lewis, R.A., Wakimoto, B.T., Denell, R.E. and Kaufman, T.C. (1980) *Genetics*, **95**, 383–397.
- Ma, J. and Ptashne, M. (1987) *Cell*, **51**, 113–119.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Mardon, H.J., Sebastio, G. and Baralle, P.E. (1987) *Nucleic Acids Res.*, **15**, 7725–7733.
- McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984) *Nature*, **308**, 428–433.
- Morgan, B.A., Johnson, W.A. and Hirsh, J. (1986) *EMBO J.*, **5**, 3335–3342.
- Mount, S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
- Nasmyth, K.A., Tatchell, K., Hall, B.D., Astell, C. and Smith, M. (1980) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 961–981.
- O'Connor, M.B., Binari, R., Perkins, L.A. and Bender, W. (1988) *EMBO J.*, **7**, 435–445.
- Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L. and Wall, R. (1980) *Cell*, **20**, 303–312.
- Schneuwly, S., Kuroiwa, A., Baumgartner, P. and Gehring, W.J. (1986) *EMBO J.*, **5**, 733–739.
- Schneuwly, S., Klemenz, R. and Gehring, W.J. (1987a) *Nature*, **325**, 816–818.
- Schneuwly, S., Kuroiwa, A. and Gehring, W.J. (1987b) *EMBO J.*, **6**, 201–206.
- Scott, M.P. and Carroll, S.B. (1987) *Cell*, **51**, 689–698.
- Scott, M.P. and O'Farrell, P.H. (1986) *Annu. Rev. Cell Biol.*, **2**, 49–80.
- Scott, M.P. and Weiner, A.J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4115–4119.
- Scott, M.P., Weiner, A.J., Polisky, B.A., Hazelrigg, T.I., Pirrotta, V., Scalenghe, F. and Kaufman, T.C. (1983) *Cell*, **35**, 763–776.
- Shapiro, M.B. and Senapathy, P. (1987) *Nucleic Acids Res.*, **15**, 7155–7174.
- Shephard, J.C.W., McGinnis, W., Carrasco, A.E., DeRobertis, E.M. and Gehring, W.J. (1984) *Nature*, **310**, 70–71.
- Solnick, D. and Lee, S.J. (1987) *Mol. Cell. Biol.*, **7**, 3194–3198.
- Stroehrer, V.L., Jorgensen, E.M. and Garber, R.L. (1986) *Mol. Cell. Biol.*, **6**, 4667–4675.
- Struhl, G. (1981) *Nature*, **262**, 635–638.
- Thomas, J.B., Bastiani, M.J., Bate, M. and Goodman, C.S. (1984) *Nature*, **310**, 203–207.
- Wakimoto, B.T. and Kaufman, T.C. (1981) *Dev. Biol.*, **81**, 51–64.
- White, R.A.H. and Lehmann, R. (1986) *Cell*, **47**, 311–321.
- Wirz, J., Fessler, L.I. and Gehring, W.J. (1986) *EMBO J.*, **5**, 3327–3334.

Received on May 10, 1988; revised on July 12, 1988

Note added in proof

Bands X and Y (Figure 7b) are present in RNA derived from adult iso-1 males, but not in RNA derived from adult iso-1 females. The 3' ends of these bands do not correspond to any recognizable splice donors, suggesting that they result from either (i) male specific splicing at non-consensus splice donors, (ii) male specific mRNA termination or (iii) the presence of one or two other genes, expressed only in adult males, with extensive homology to *Antennapedia* throughout the homeobox and downstream translated and untranslated regions.