MOLECULAR AND CELLULAR BIOLOGY, Dec. 1986, p. 4667–4675
0270-7306/86/124667-09$02.00/0

Vol. 6, No. 12

Multiple Transcripts from the Antennapedia Gene of *Drosophila melanogaster*

VIRGINIA L. STROEHER,[1] ERIK M. JORGENSEN,[2] AND RICHARD L. GARBER[2]*

Department of Biological Structure[1] and Department of Genetics,[2] University of Washington, Seattle, Washington 98195

Received 14 May 1986/Accepted 15 August 1986

The structures of four major transcripts from the homeotic gene Antennapedia of *Drosophila melanogaster* were determined. These transcripts constitute two RNA classes, each class initiating from a unique promoter but sharing 3′ exons. Within the shared sequences is a major open reading frame encoding a 378-amino-acid protein as well as alternative polyadenylation sites. Although the RNA classes differ in their 5′ sequences, both leaders contain many AUGs upstream of the major open reading frame. For the two RNA classes, neither gross tissue nor temporal specificity was observed. However, the second poly(A) site is preferred in neural tissue. The structural diversity of the RNAs is discussed in relation to biological functions of the Antennapedia locus.

As development proceeds from the zygote to the multicellular organism, determinative events occur which commit groups of cells to discrete developmental pathways. The underlying molecular mechanisms that determine and maintain tissue identity are still largely unknown. Development in *Drosophila melanogaster* involves the interactions between many different regulatory genes. These genes function to establish the polarity of the embryo (35), divide the animal into repeated segments (36), and define the morphological identity of each segment (21). Establishment and maintenance of the identity of tissues in the thoracic segments is the central role of the gene we are studying, the homeotic locus Antennapedia (*Antp*). Recessive lethal mutations of the *Antp* locus transform cells of the larval and adult ectoderm in an anterior direction (21, 46, 54, 55, 58). To understand how the *Antp* gene controls tissue identity, the following questions must be answered. How is *Antp* regulated in a cell-specific manner? How does it regulate the expression of other genes to shape subsequent development? What structural features of the *Antp* gene specify the biological functions attributed to it?

How is the *Antp* gene regulated during development? *Antp* exhibits a pattern of transcript accumulation which is not only tissue specific but is confined to specific subpopulations of cells within a given tissue (27, 29). Genetic and molecular analyses demonstrate that the actions of other genes lead to this restricted pattern. For example, activities of bithorax complex genes appear to repress the transcription of *Antp* in the third thoracic and first seven abdominal segments (12). Mutations in other genes, such as Sex combs reduced, extra sex combs, and Polycomb, also influence *Antp* expression either directly or indirectly (46, 56, 59).

How does *Antp* control other genes? Several observations suggest that some regulatory genes interact with target genes by binding to the DNA of the target loci. A number of these regulatory genes, including *Antp*, contain the homeobox sequence (31, 40, 42, 51). This encodes a peptide domain which shares homology with yeast transcription regulatory proteins (*MATα2* and *MATa1*), as well as bacterial DNA-binding proteins (26, 53). Consistent with a DNA-binding role is the finding that the protein products of several homeobox-containing genes are localized to the nucleus (2, 4, 7, 61). Finally, crude *Escherichia coli* extracts containing homeobox portions of the engrailed protein bind in vitro to DNA sequences upstream of the engrailed gene itself as well as to the fushi tarazu gene (6). If these proteins in fact bind to DNA in vivo, it is likely that they regulate transcription of their target genes. It is unknown whether the target genes are themselves regulatory loci or genes encoding structural products for cell differentiation.

How do features of the *Antp* locus and its transcripts facilitate its functions during development? We present the sequences of several *Antp* cDNAs. We also present evidence for the existence of at least four major RNAs arising from two promoters and ending at alternative polyadenylation sites. These RNAs differ by unusually large regions of 5′ or 3′ untranslated sequences. These features of the *Antp* gene and its RNA products are discussed in relation to the genetically defined functions of the locus.

MATERIALS AND METHODS

RNA isolation. Total RNA was prepared from *D. melanogaster* 4- to 8-h-old embryos or late pupae (40) with the following modifications. For each RNA preparation, 2 g of powdered embryos or pupae was homogenized in a 15-ml Dounce homogenizer containing 10 ml of guanidinium thiocyanate homogenization buffer and centrifuged at 9,500 × *g* for 20 min at 20°C in a Sorvall SS-34 rotor (Ivan Sorvall, Inc., Norwalk, Conn.). To the supernatant was added 0.25 volumes of CsCl solution (5.7 M CsCl, 10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES], 1 mM EDTA [pH 7.6]) and Sarkosyl (CIBA-GEIGY Corp., Summit, N.J.) to 4%. The mixture was split and layered over 2-ml cushions of CsCl solution, and RNA was pelleted by centrifugation at 110,000 × *g* for 20 h at 20°C in an SW41 rotor (Beckman Instruments, Inc., Fullerton, Calif.). Poly(A)$^+$ RNA was isolated by using oligo(dT) cellulose chromatography (type II; Collaborative Research, Inc., Waltham, Mass.) (28).

Imaginal disks were isolated from wandering third-instar larvae (Sevelin strain) by hand-dissection in Ringer insect solution (38) containing 5% horse serum to prevent tissue clumping. Minipreparations (20) of RNAs were made from 103 brain lobes; 40 ventral nerve cords; and 123 eye-antennal, 100 first-leg, 100 second-leg, 100 third-leg, 116 humeral, 25 wing, and 209 haltere imaginal disks.

[^{32}P]RNA probes. Appropriate sequences were subcloned into either pSP or pGEM vectors (Promega Biotec) for SP6

* Corresponding author.

or T7 RNA polymerase transcription. Linearized templates were treated with 0.2 mg of proteinase K per ml for 30 min at 37°C, followed by extraction twice with phenol-chloroform-isoamyl alcohol and once with chloroform, and then were ethanol precipitated (Vector Cloning Systems protocol).

RNA probes were transcribed (32) with 12 μM [α-^{32}P]UTP at either 600 or 3,000 Ci/mmol, yielding a final UTP concentration of 24 μM. These precursors yielded RNAs at specific activities of greater than 3×10^8 or 3×10^9 dpm/μg, respectively. For RNase protections, the DNA template was removed by adding DNase I (RNase free) to 0.1 U/μl and incubating for 30 min at 37°C. RNA probes were then extracted once with phenol-chloroform-isoamyl alcohol and once with chloroform and ethanol precipitated.

RNA blot analysis. Whole-animal poly(A)$^+$ RNA (5 to 10 μg per lane) or larval-tissue total RNA was separated on 1.0% agarose, 2.2 M formaldehyde gels in MOPS buffer (28). The RNA was capillary transferred to either nitrocellulose (Schleicher & Schuell, Inc., Keene, N.H.) or nylon membrane (Hybond-N; Amersham Corp., Arlington Heights, Ill.) with 20× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate [pH 7.0]) for 18 to 20 h (57) and then baked for 2 h at 80°C. Nylon membranes were UV irradiated 2 to 5 min before baking (Amersham Corp. 1985. Membrane transfer and detection methods. Booklet R851196-5/85).

Low-specific-activity ^{32}P-labeled RNA probes were used for hybridization to whole-animal RNA, whereas high-specific-activity probes were used for larval-tissue RNAs. Nitrocellulose filters were hybridized (32) with a probe concentration of 1 ng/ml per kilobase (kb) of probe length for 24 to 48 h at 56°C. Nylon filters were prehybridized in 50% formamide–5× SSPE (0.9 M NaCl, 50 mM NaH_2PO_4, 0.5 mM EDTA [pH 7.7])–5× Denhardt solution–10% polyethylene glycol 6000–1% sodium dodecyl sulfate–0.1 mg of denatured DNA per ml–0.05 mg of tRNA per ml at 60°C for 4 h. Hybridization was carried out in the same solution with a probe concentration of 1 ng/ml per kb of probe length for 20 h at 60°C. Filters were rinsed once in 5× SSPE at room temperature, then washed at 65°C for 20-min periods, once in 1× SSPE–0.1% sodium dodecyl sulfate, and three times in 0.1× SSPE–0.1% sodium dodecyl sulfate (1; New England Nuclear Corp. 1984. GeneScreenPlus hybridization transfer membrane. Protocol booklet; Amersham booklet).

Subsequent hybridization of the filters with the ribosomal protein 49 probe (37), provided by Michael Rosbash, demonstrated that RNA preparations from eye-antennal disks were intact.

S1 nuclease protection. Genomic DNA fragments containing exon regions plus flanking sequences were subcloned into M13 vectors. Single-stranded DNA was isolated (44) and purified by agarose gel electrophoresis. Single-stranded DNA (50 ng) was protected from S1 nuclease digestion by annealing first to poly(A)$^+$ RNA (5 to 10 μg) and then to single-stranded DNA of the opposite orientation (16). The double-stranded DNA fragments were fractionated on an agarose gel, transferred to nitrocellulose, and hybridized by using 10^6 cpm of nick-translated DNA probe per ml (1×10^8 to 3×10^8 cpm/μg) representing the original genomic DNA (13).

Isolation of cDNA clones. New cDNA clones were isolated from several *D. melanogaster* libraries constructed in λ*gt10* and kindly made available by M. Goldschmidt-Clermont, S. Schneuwly, and L. Kauvar. Each was screened with nick-translated probes (1×10^8 to 3×10^8 cpm/μg) derived from cDNA or genomic DNA sequences of the *Antp* locus. The following lists indicate the stage of RNA used for the library from which these clones were isolated: embryonic, 903, 904, 950, 958, and 963; larval, 927, 964, and 965; pupal, 909.

RNase protection. Exon ends were mapped by the RNase protection method (62) with the following modifications. Annealing of 1×10^6 to 2×10^6 cpm of RNA probe (low specific activity) to either total RNA (15 to 30 μg) or poly(A)$^+$ RNA (1 to 5 μg) was followed by digestion at 30°C for 1 h with either 40 μg of RNase A per ml plus 2 μg of RNase T1 per ml or 2, 5, or 10 μg of RNase T1 per ml alone. RNase A cleaves 3′ to pyrimidine residues, whereas RNase T1 cleaves 3′ to guanine residues. After ethanol precipitation, the protected fragments were suspended directly in 80% formamide–0.25 mM EDTA–0.06% bromophenol blue–0.06% xylene cyanol, heated at 65°C for 30 min, and fractionated on a 6% polyacrylamide–8 M urea sequencing gel.

Estimating the size of protected probe fragments frequently leads to ambiguity in the exact site of RNase cleavage. The same ambiguity arises with S1 mapping or primer extension mapping of RNA molecule ends or splice junctions. We suggest that the use of a set of ribonucleases with different cleavage specificities, as exist for RNA sequencing (9), could be combined with knowledge of the DNA sequence to determine the position of RNA termini, splice sites, or single-base-pair differences between sequences.

Note that under some conditions RNase apparently cleaves between the two terminal nucleotides at the end of an RNA-RNA hybrid. As the concentration of nuclease increases, cleavage at the internal site is facilitated. We attribute this to partial strand melting at the ends of the hybrid molecules.

DNA sequencing. Dideoxynucleotide sequencing (45) of cDNA and genomic DNA fragments used site-directed subcloning into pGEM vectors. A series of nested deletions was generated by using exonuclease III plus S1 nuclease (17). Plasmid DNA was prepared to assay the extent of deletion (18), and supercoiled plasmid DNA was directly sequenced by using SP6 or T7 promoter primers and Promega Biotec conditions. In all cases, sequences of both strands of the cDNAs were determined. For handling DNA and protein sequence data, the GENEPRO programs written by J. Brown and J. Wallace (Riverside Scientific, Seattle, Wash.) and the Protein Coding Region Locator routine of the Pustell Sequence Analysis programs (International Biotechnologies, Inc.) were used.

RESULTS

Structures of *Antp* cDNA clones. Scott et al. (52) identified two major *Antp* poly(A)$^+$ RNA species, 5.0 and 3.5 kb in size, which are present throughout development. The 3.5-kb species predominates during early embryonic stages but is replaced by the 5.0-kb RNA by the end of this period. The 5.0-kb RNA then remains the major species through pupal development, although the total amounts in first-, second-, and early third-instar larvae are quite low. Some understanding of the structure of these RNAs came from the isolation and initial characterization of several cDNAs (10). cDNAs 903 and 909 were described as containing four and three exons, respectively. It has since been shown that one exon region common to the two cDNAs in fact contains four exons (11). To conform with these findings, the exon structure of the 903 cDNA will be diagrammed as 1-2-4-5-6-7-8 and that of 909 will be diagrammed as 3-4-5-6-7-8.

To better define the RNA structures, additional cDNA clones were isolated from several libraries constructed by

Michael Goldschmidt-Clermont, Stephan Schneuwly, and Larry Kauvar. These were screened with probes derived from previously isolated cDNAs or from genomic sequences. A diagram relating the *Antp* genomic DNA sequences and the cDNAs is shown in Fig. 1.

The complete sequences of the 903, 904, 958, 963, 964, and 965 clones, all derived from Oregon R *D. melanogaster* strains, are presented in Fig. 2. The sequences of the Canton S strain genomic regions, upstream and downstream of exons 1, 3, and 8, were also determined.

Two classes of large cDNAs were found. In cDNA 903, exons 1 and 2 are combined with exons 4 to 8, whereas in cDNA 909 and 950 exon 3 is combined with exons 4 to 8. Some of the cDNAs end with poly(A) stretches (indicated by * in Fig. 1). Together these define two polyadenylation sites; with 958, there are possibly three. The two sites in exon 8 are 1,424 bases apart. Each poly(A) addition site is preceded by a canonical AATAAA motif (41).

Alternative polyadenylation sites. While determining the size of each exon by S1 analysis and RNase protection, we found that sequences at the 3′ end of the last *Antp* exon contribute all or most of the 1.5-kb difference between the 5.0- and 3.5-kb RNAs (data not shown). To confirm that the 5.0- and 3.5-kb transcripts differ by this region, RNA blots were probed with the 3′ sequences of exon 8. Only the 5.0-kb *Antp* RNA band was detected (Fig. 3, lane g). These alternative 3′ ends for exon 8 are consistent with the position of poly(A) tracts found in the cDNAs.

Determining the 5′ ends of exons 1 and 3. The 5′ ends of exons 1 and 3 were mapped by RNase protection (62). Embryonic RNA was annealed to ^{32}P-labeled antisense RNA probes containing the 5′ end of each exon and flanking sequences. Hybrids were digested with either RNases A plus T1 (A+T1) or RNase T1 alone (T1). These RNases digest single-stranded RNAs at specific nucleotides and leave double-stranded molecules intact. The resulting protected fragments were resolved on a sequencing gel.

A+T1 mapping of exon 1 resulted in protection of a 116-nucleotide (nt) fragment, whereas T1 alone protected 128- and 130-nt fragments (Fig. 4). The 2-nt difference between the T1 products is attributed to cleavage at one of two neighboring T1 sites at the 3′ end of the hybrid and was dependent on enzyme concentration (see Materials and Methods). A+T1 cleaved at the starting nucleotide of cDNA 958, while T1 alone cleaved 12 nt upstream at the first accessible T1 site. This places the start of exon 1 at or within a few nucleotides of the 5′ end of the 958 cDNA.

The 5′ end of exon 3 was similarly determined. A+T1 digestion resulted in a 125-nt fragment, and T1 alone protected a 127-nt fragment (Fig. 4). These observations place the 5′ end of exon 3 in the region of nucleotide 23 (Fig. 2 and 4).

Comigrating RNA species. Taking together the exon composition of the cDNAs with the exon sizes from protection analyses and the cDNA sequences, transcript compositions and sizes were predicted. The first transcript class, represented by cDNA 903, would have exons 1-2-4-5-6-7-8. In this class, two RNAs are possible, 3.3 or 4.7 kb long, depending on the poly(A) site used. The second transcript class, of which cDNAs 909 and 950 are copies, would have exons 3-4-5-6-7-8. Two RNAs are again possible, 3.5 or 4.9 kb long, also dependent on the poly(A) site. To determine which of these four transcripts are actually made, blots of poly(A)$^+$ RNA from whole *D. melanogaster* embryos or pupae were probed with sequences from the five major exon regions. Each probe hybridized to both the 5.0- and 3.5-kb bands

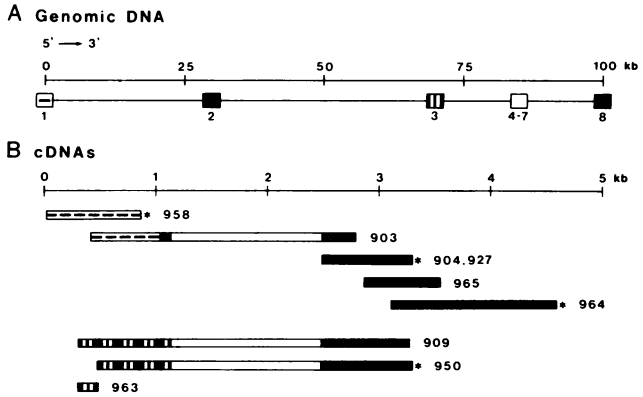


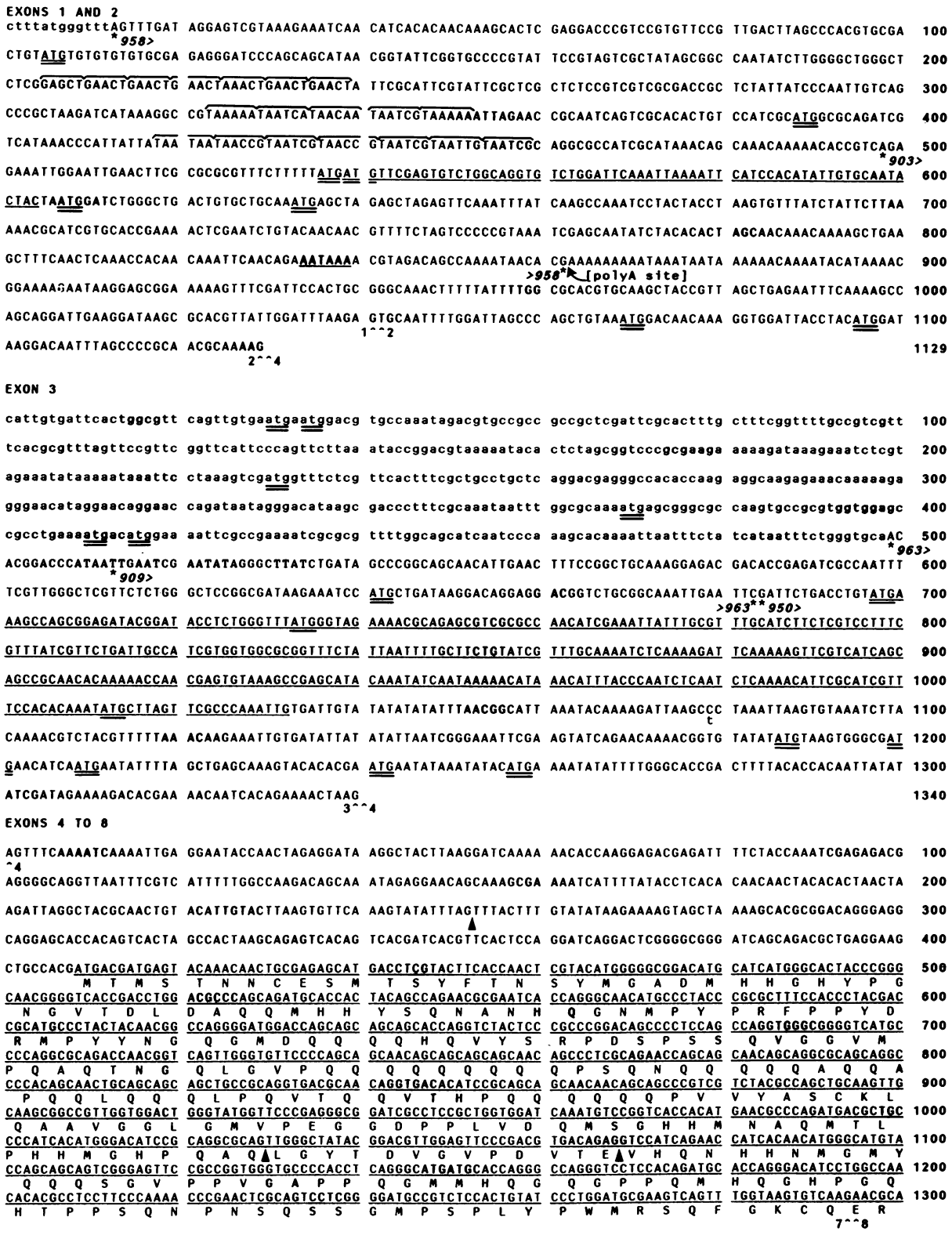
FIG. 1. Exon composition of *Antp* cDNAs. (A) The position of each of the eight *Antp* exons within the 100-kb locus is shown. (B) The cDNAs are aligned according to their sequences, each pattern corresponding to the exon origin of a sequence. None of the cDNAs are full length, and only a few include poly(A) regions (indicated by *).

(Fig. 3, lanes a to f). We conclude that both transcript classes make use of both poly(A) sites, resulting in production of all four RNAs predicted. Because of the small size difference between transcript classes, the 3.5- and 5.0-kb bands each consist of two comigrating RNAs.

***Antp* ORFs.** The defined RNA structures share a single, large open reading frame (ORF) of 1,134 nt. This codes for a 378-amino acid protein with a molecular weight of 42,780. The first AUG codon in this large ORF is positioned in a favorable context for initiation of translation, CCACGA TGA, compared with the consensus (CC)ACCATGG sequence (24). The putative *Antp* protein has two domains that have previously received attention. About one-third of the way into the protein, there is a stretch consisting mostly of glutamine residues. These contribute heavily to the total glutamine composition of over 18% of the peptide. The second domain in the *Antp* peptide is the region of about 60 basic amino acid residues termed the homeobox. It is located in the last quarter of the putative *Antp* peptide. Both of these regions are known to reside in *Antp* (11, 25, 31, 60).

In *Antp* transcripts of either exon composition class, the large ORF is preceded by more than 1,100 nt of leader sequence. These leaders are derived from either exons 1 and 2 or from exon 3. Although different in sequence, both have many AUG codons before the start of the large *Antp* protein (Fig. 2). For exon 1- and 2-containing RNAs, there are eight AUG codons before the *Antp* peptide start signal. Only the fourth of these AUGs lies in a reasonable consensus ribosome initiation site environment, and it could encode a 22-amino-acid peptide. The RNAs containing exon 3 have 15 upstream AUG codons and could generate peptides 2 to 112 amino acids long. Although it lacks the −3 purine residue of the consensus ribosome initiation sequence, the 112-codon ORF conforms approximately to *D. melanogaster* codon usage.

Tissue localization of *Antp* transcripts. We next asked whether the four major transcripts have distinct tissue specificities. RNA was prepared from hand-dissected larval tissue, divided into two samples, and electrophoresed on identical agarose gels. The blotted RNAs were hybridized with probes specific to either exon 1 or 3 (Fig. 5). No *Antp*



transcripts were observed in the eye-antennal disk. However, both probes hybridized to 5.0- and 3.5-kb transcripts in all thoracic imaginal disks, brain, and ventral nerve cord. Additionally, both *Antp* RNA sizes were detected in the larval gut by using a probe for exon 3, which demonstrates that *Antp* transcription is not limited to ectoderm (data not shown).

Although both polyadenylation sites are used in all tissues that transcribe *Antp*, there is a sevenfold preference for processing at the second poly(A) addition site in both larval

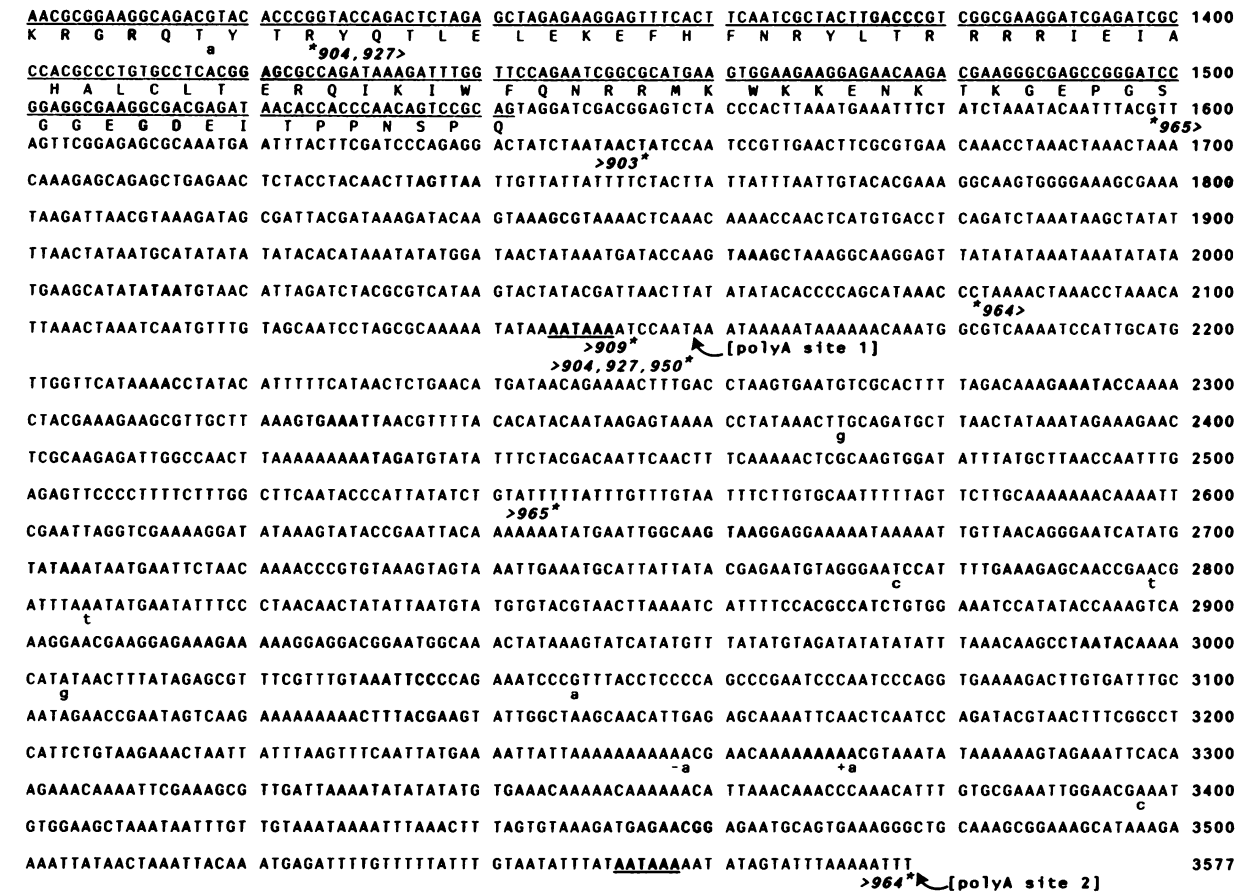


FIG. 2. Shown in capital letters are the sequences assembled from *Antp* cDNAs. Below these, * indicates the beginning (e.g., *958>) and end (e.g., >958*) positions of each clone. Three polyadenylation sites are marked in square brackets, and their associated upstream AATAAA motifs are underlined. The genomic DNA sequence was determined for exons 1, 3, and 8. Lowercase letters indicate sequences determined from genomic DNA. Letters below the cDNA sequence indicate internal sites where the genomic DNAs (strain Canton S) differ from the cDNAs (strain Oregon R) and are designated as nucleotide substitutions, deletions, or insertions (e.g., a, −a, +a, respectively). Major cDNAs contain exons 4 to 8 joined to either exons 1 and 2 or to exon 3, never to both. Splice junctions we determined are denoted by the linkage of the two exon numbers (e.g., 1^^2). In the exon 4 to 7 block, the exon 4-5, 5-6, and 6-7 splice junctions are denoted by filled carets (▲) at nucleotides 252, 1029, and 1068, respectively. These positions were taken from the published data of Schneuwly et al. (50). Exons 4 to 8 include the major ORF, which is underlined and below which lies the single-letter amino acid translation. Note the large number of glutamine (Q) residues. ATGs 5′ of the major ORF are indicated by a double underline. Also underlined are two potential ORFs, lying in the 5′ leader RNA sequences, one each in the exon 1-2 and exon 3 regions. These could be translated into 22- and 112-amino-acid peptides, respectively. Brackets above the sequence mark direct repeats.

brain and ventral nerve cord (Fig. 5). This preference is independent of whether the RNA contains exon 1 or exon 3.

Small *Antp* RNAs. Scott et al. (52) reported that exon 1 hybridizes to at least one small RNA. We show that this exon may hybridize to several poly(A)$^+$ RNAs about 1 kb long (Fig. 3, lanes a and b). We have isolated one cDNA, 958, which may be derived from one of these small RNAs. It is encoded by exon 1, is 849 base pairs long, and terminates in a 21-nt poly(A) stretch preceded by the canonical AATAAA signal.

DISCUSSION

Structure of the *Antp* transcripts. Transcription from the wild-type Antennapedia gene of *D. melanogaster* was studied by probe hybridization to RNA blots, RNA protection analysis, and sequencing of cDNAs. The major RNA species are estimated to be 3.5 and 5.0 kb long, including the poly(A) tails. From the results presented in this paper, together with other data, we conclude that the *Antp* gene is transcribed from two promoters located about 70 kb apart (see below). Synthesis of these transcripts begins at exon 1 or at exon 3, yielding processed RNAs of exon compositions 1-2-4-5-6-7-8 and 3-4-5-6-7-8. Once transcription extends into exon 8 from either promoter, two cleavage and polyadenylation sites are available. These are separated by 1.46 kb, which accounts for the size difference between the 3.5- and 5.0-kb RNAs. Because the size of exons 1 plus 2 equals that of exon 3, only two RNA bands instead of four are visible on gel blots. This leads to two pairs of comigrating RNAs.

5′ *Antp* mRNA sequences. Preceding the major ORF in each large *Antp* transcript lies approximately 1,100 nt of RNA sequence containing many AUGs. The unusual features of these 5′ sequences led us to consider roles for these in some aspect of translation (reviewed by T. Hunt, Nature [London] **316**:580–581, 1985). Several ways by which RNA regions could be used to regulate translation are discussed below with respect to *Antp*.

Several mechanisms for regulating translation work by restricting the access of mRNA to active ribosomes. In

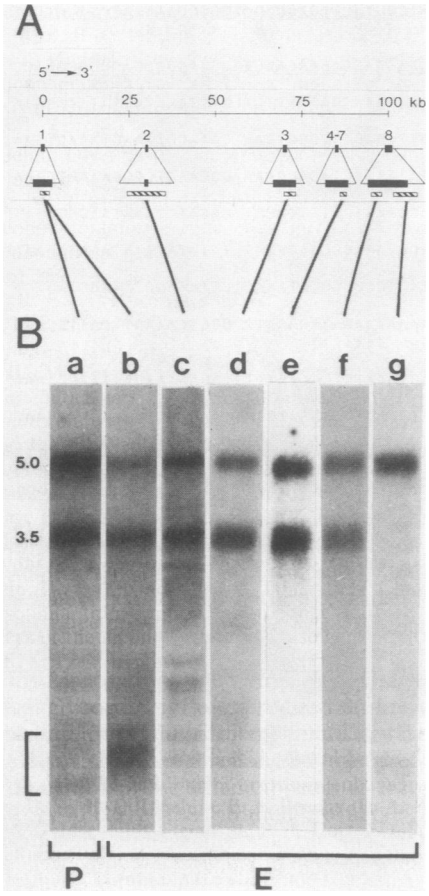


FIG. 3. Blot analysis of *Antp* transcripts. Poly(A)$^+$ RNA from either pupae (P) or 4- to 8-h-old embryos (E) was fractionated on formaldehyde-agarose gels, blotted, and hybridized with ^{32}P-labeled RNA probes of the sequences indicated in panel A. Probes contained exon 1 (a and b), exon 2 (c), exon 3 (d), exons 4 to 7 (e), the 5′ end of exon 8 (f), and the 3′ end of exon 8 (g). Each hybridized to both 5.0- and 3.5-kb RNAs, except the 3′ end of exon 8, which hybridized to only the 5.0-kb RNA. Several small RNAs were also detected (vertical bracket at left of panel B).

Xenopus laevis oogenesis, translation is prevented by a masking protein which attaches to mRNAs (43). Alternatively, for *D. melanogaster* heat shock protein 70 mRNA, access to active ribosomes under induction conditions depends on specific 5′ untranslated sequences (22, 30). Although we have no evidence suggesting that similar mechanisms are used with *Antp* transcripts, these mRNAs may have restricted access to active ribosomes because of retention in the nucleus. In situ hybridization to RNA in third-larval-instar imaginal disks indicates primarily nuclear localization, whereas in embryos the transcripts are found in both nuclei and cytoplasm (27).

The two classes of *Antp* RNAs contain 5′ regions with 8 or 15 upstream AUG codons. According to the ribosome scanning model of translation, upstream AUGs can affect the efficiency of translation of a downstream ORF (23). The model proposes that ribosomes bind the 5′ mRNA end and search in the 3′ direction until encountering a start AUG where it can initiate protein synthesis. It is postulated that AUGs between the 5′ end and the major ORF could slow or prevent the translation of the encoded protein. Site-directed mutagenesis of upstream ATGs in the yeast *GCN4* gene demonstrated that these AUGs in the mRNA can exert either a positive or a negative influence on translation of the downstream ORF (33).

RNA secondary structure also affects translation. The insertion of random secondary structure (synthetic hairpins) into a message reduced the efficiency of translation (39). Alternatively, increased efficiency of translation has been proposed for RNAs in which secondary structures could loop out 5′ sequences containing consensus AUGs (5, 14, 34). In *Antp* there are sequences in exon 3 which could form a weak hairpin structure spanning upstream AUG codons 8 and 9.

As an alternative to upstream start signals serving to regulate translation of the major ORF, any of these AUGs might initiate the translation of its own protein. Most of the 5′ ORFs terminate after less than 30 codons, and the AUGs vary greatly in their match to the consensus translational start sequence (24). The longest ORF, contained in exon 3, could encode a 112-amino-acid protein with characteristic *D.*

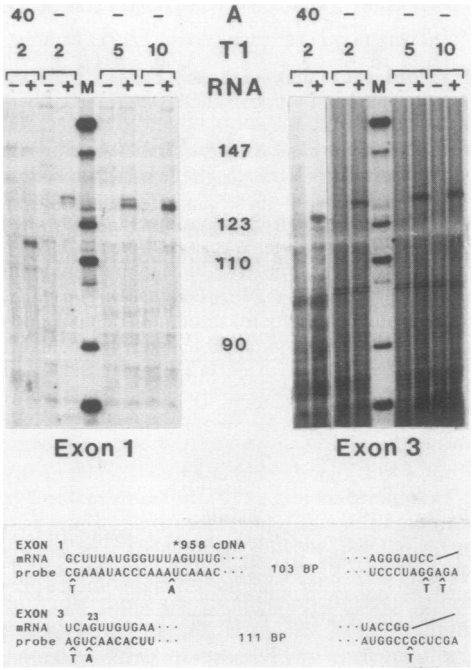


FIG. 4. 5′ End analysis of exon 1 and exon 3. RNase protection was used to determine the 5′ ends of exons 1 and 3. ^{32}P-labeled anti-sense RNA probes were annealed to embryonic RNA and digested with either RNase A (A) plus T1 or increasing concentrations of T1 alone. The amount of each enzyme used (μg/ml) is indicated. Protected fragments were fractionated on a sequencing gel along with labeled DNA size markers (M). For each enzyme condition, reactions were done with probe alone (−) and probe plus embryonic RNA (+). The 5′ and 3′ ends of the RNA-RNA hybrids are indicated. Possible mRNA sequences are drawn above the probe sequences. Sites of cleavage are indicated as T̂ and Â below the probe strand. At the right, a divergence occurs where the mRNA continues (—) and the probe extends into the vector. For the exon 1 probe, digestion with A+T1 resulted in the protection of a 116-nt fragment. Digestion with T1 alone resulted in fragments of 128 and 130 nt; their relative abundance is dependent on the concentration of T1. Sites of cleavage are indicated below the probe sequence. We conclude that the start of exon 1 is at or within a few nucleotides of the start of 958. Digestion of the exon 3 probe with A+T1 protects a 125-nt fragment, whereas T1 alone protects a 127-nt fragment. Based on the location of cleavage sites within this probe, the start of exon 3 is mapped at or within a few nucleotides of nucleotide 23.

melanogaster codon usage bias. However, the translational start does not contain the consensus purine at the −3 position. Although polycistronic messages are common in procaryotes, in eucaryotes they are known only in viral systems (3, 8, 15, 19).

We have given several examples of functional roles for 5′ mRNA sequences. One or several of these mechanisms may be used in the case of the *Antp* gene. Since there are two classes of *Antp* RNAs that differ in their 5′ regions, differences between the activities of the two *Antp* RNA classes may arise from these leader sequences and thereby play an important role in *Antp* gene functions.

Evidence for two *Antp* promoters. What data support the existence of two *Antp* promoters? (i) The structure of the cDNA clones and the RNA blotting data, taken together, indicate that there are two RNA classes, beginning either with exon 1 or with exon 3. These exons are separated in the genome by about 70 kb. (ii) Abbott and Kaufman have genetic evidence for two independent functional units within the *Antp* gene (M. Abbott, Ph.D. thesis, Indiana University, Bloomington, 1984). Their mutant analysis localizes unit one defects to the physical map region including exons 1 and 2, while unit two defects fall in the region including exon 3, in the same way each of our transcription units corresponds to one of these regions. (iii) We have analyzed transcription in *Antp* mutations, in which the two promoters have been physically separated by DNA rearrangement, and found that both halves of the locus retain promoter activity (E. M. Jorgensen and R. L. Garber, in preparation).

Biological functions arising from two promoters. In general terms, two promoters could be used to activate transcription at different times in the development of an organism, as demonstrated for the *D. melanogaster* alcohol dehydrogenase gene (47). However, blot analysis with *Antp* probes of RNA from embryonic, larval, and pupal stages demonstrated the activity of both promoters at each stage (Fig. 3 and 5).

Another function for two promoters could be to independently regulate gene activity in different tissues, as with the mouse α-amylase gene (48). Genetic analysis of *Antp* indicates two spatially defined functional units exist within the locus (Abbott, Ph.D. thesis). Somatic recombination clones homozygous for mutations in functional unit one were found to be defective in dorsal regions of thoracic segments one and two, tissues which are contributed by the humeral and wing imaginal disks. Defects were also seen in transplanted wing disks derived from homozygous *Antp* mutants (49). Unit two mutations, on the other hand, disrupt normal leg development. In whole animals, the transheterozygote (unit one mutation/unit two mutation) survives to adulthood and is phenotypically normal. On the basis of breakpoints which disrupt the *Antp* gene, Abbott has assigned the dorsal functions to exons 1 and 2 and the ventral ones to exon 3.

The two sites of *Antp* transcription initiation fall within the same regions as the functional units defined by Abbott. Since loss of functional unit one leads to a dorsal defect, we expected that transcripts originating from the first *Antp* promoter would be found in dorsal tissue and that RNAs from the second promoter would be localized ventrally. However, we found transcripts from each *Antp* promoter in both dorsal and ventral thoracic imaginal disks. We suggest two possible ways to reconcile the data. Although both transcripts are found in the same tissue, they may not be present in the same cells of that tissue. Alternatively, although both RNA classes are synthesized, only one may be translated in each tissue. The latter example makes possible a role for the 5′ upstream mRNA sequences in regulating the spatial pattern of *Antp* translation.

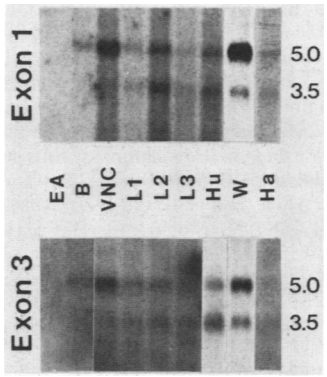


FIG. 5. Analysis of *Antp* transcripts in specific third-instar larval tissues. Total RNA isolated from individual larval tissues was probed with sequences from exon 1 or 3. The RNAs examined were extracted from eye-antennal imaginal disks (EA); brain lobes (B); ventral nerve cords (VNC); and first-leg (L1), second-leg (L2), third-leg (L3), humeral (Hu), wing (W), and haltere (Ha) imaginal disks. The four major transcripts are present in all these tissues except eye-antennal disks, although the second poly(A) site is preferred in neural tissue.

During final preparation of the present manuscript, similar results with *Antp* were published (50). The two studies complement in that each group isolated cDNA clones that the other lacked. In total, these clones confirmed the 5′ and 3′ ends of transcripts mapped approximately by S1 analysis, primer extension, and RNA protection. Several nucleotides in the two sets of sequences, which were determined independently, differ. We report these differences for the benefit of future studies in which they may be important. The format will be exon number (nucleotide number): our sequence/Schneuwly et al. sequence (see Fig. 2). Exon 1 (284):A/C. In this one case it is likely due to strain differences, since we sequenced only an Oregon R strain cDNA and the other researchers sequenced only a Canton S genomic clone. Exon 3 (751):G/nothing. Exon 7 (1113):G/A. Exon 8 (2768-2769):nothing-nothing/A-G, (2784-2785):GA/AG, (2788):nothing/GAA, (2791):nothing/A, (2799):C/G, (2831):nothing/A, and (2847):C/nothing. Only in the case of Exon 7 (1113) does a difference occur in the proposed protein-coding region, and in this case the difference does not alter the amino acid sequence.

ACKNOWLEDGMENTS

We thank Chris Gaiser and Paul Talbert for technical assistance with DNA sequencing and Lauren Frischer for the isolation of cDNA clone 958. We also thank Steven Henikoff, Gerold Schubiger, Thomas Kornberg, and the members of their laboratories for experimental procedure suggestions and Michael Abbott for the copy of his dissertation and discussions. We appreciate the comments on the manuscript given by Benjamin Hall and Steven Henikoff.

This research was supported by Public Health Service grants BRSG RR-07096 and GM33219 from the National Institutes of Health.

LITERATURE CITED

1. **Amasino, R. M.** 1986. Acceleration of nucleic acid hybridization rate by polyethylene glycol. Anal. Biochem. **152**:304–307.
2. **Beachy, P. A., S. L. Helfand, and D. S. Hogness.** 1985. Segmental distribution of bithorax complex proteins during *Drosophila* development. Nature (London) **313**:545–551.

3. **Bellini, W. J., G. Englund, S. Rozenblatt, H. Arnheiter, and C. D. Richardson.** 1985. Measles virus *P* gene codes for two proteins. J. Virol. **53**:908–919.
4. **Carroll, S. B., and M. P. Scott.** 1985. Localization of the fushi tarazu protein during *Drosophila* embryogenesis. Cell **43**:47–57.
5. **Darlix, J.-L., M. Zuker, and P.-F. Spahr.** 1982. Structure-function relationship of Rous sarcoma virus leader RNA. Nucleic Acids Res. **10**:5183–5196.
6. **Desplan, C., J. Theis, and P. H. O'Farrell.** 1985. The *Drosophila* developmental gene, engrailed, encodes a sequence-specific DNA-binding activity. Nature (London) **318**:630–635.
7. **DiNardo, S., J. M. Kuner, J. Theis, and P. H. O'Farrell.** 1985. Development of embryonic pattern in *D. melanogaster* as revealed by accumulation of the nuclear engrailed protein. Cell **43**:59–69.
8. **Dixon, L. K., and T. Hohn.** 1984. Initiation of translation of the cauliflower mosaic virus genome from a polycistronic mRNA: evidence from deletion mutagenesis. EMBO J. **12**:2731–2736.
9. **Donis-Keller, H., A. M. Maxam, and W. Gilbert.** 1977. Mapping adenines, guanines, and pyrimidines in RNA. Nucleic Acids Res. **4**:2527–2538.
10. **Garber, R. L., A. Kuroiwa, and W. J. Gehring.** 1983. Genomic and cDNA clones of the homeotic locus Antennapedia in *Drosophila*. EMBO J. **2**:2027–2036.
11. **Gehring, W. J.** 1985. Homeotic genes, the homeo box, and the genetic control of development. Cold Spring Harbor Symp. Quant. Biol. **50**:243–251.
12. **Hafen, E., M. Levine, and W. J. Gehring.** 1984. Regulation of Antennapedia transcript distribution by the bithorax complex in *Drosophila*. Nature (London) **307**:287–289.
13. **Hanahan, D., and M. Meselson.** 1980. Plasmid screening at high colony density. Gene **10**:63–67.
14. **Hay, N., H. Skolnik-David, and Y. Aloni.** 1982. Attenuation in the control of SV40 gene expression. Cell **29**:183–193.
15. **Heermann, K. H., U. Goldmann, W. Schwartz, T. Seyffarth, H. Baumgarten, and W. H. Gerlich.** 1984. Large surface proteins of hepatitis B virus containing the pre-*s* sequence. J. Virol. **52**:396–402.
16. **Henikoff, S.** 1983. Cloning exons for mapping of transcription: characterization of the *Drosophila melanogaster* alcohol dehydrogenase gene. Nucleic Acids Res. **11**:4735–4752.
17. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
18. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**:193–197.
19. **Jay, G., S. Nomura, C. W. Anderson, and G. Khoury.** 1981. Identification of the SV40 agnogene product: a DNA binding protein. Nature (London) **291**:346–349.
20. **Karr, T. L., Z. Ali, B. Drees, and T. Kornberg.** 1985. The engrailed locus of *D. melanogaster* provides an essential zygotic function in precellular embryos. Cell **43**:591–601.
21. **Kaufman, T. C., and M. K. Abbott.** 1984. Homeotic genes and the specification of segmental identity in the embryo and adult thorax of *Drosophila melanogaster*, p. 189–218. *In* G. M. Malacinski and W. H. Klein (ed.), Molecular aspects of early development. Plenum Publishing Corp., New York.
22. **Klemenz, R., D. Hultmark, and W. J. Gehring.** 1985. Selective translation of heat shock mRNA in *Drosophila melanogaster* depends on sequence information in the leader. EMBO J. **4**:2053–2060.
23. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.
24. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283–292.
25. **Laughon, A., S. B. Carroll, F. A. Storfer, P. D. Riley, and M. P. Scott.** 1985. Common properties of proteins encoded by the Antennapedia complex genes of *Drosophila melanogaster*. Cold Spring Harbor Symp. Quant. Biol. **50**:253–262.
26. **Laughon, A., and M. P. Scott.** 1984. Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. Nature (London) **310**:25–31.
27. **Levine, M., E. Hafen, R. L. Garber, and W. J. Gehring.** 1983. Spatial distribution of Antennapedia transcripts during *Drosophila* development. EMBO J. **2**:2037–2046.
28. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Martinez-Arias, A.** 1986. The Antennapedia gene is required and expressed in parasegments 4 and 5 of the *Drosophila* embryo. EMBO J. **5**:135–141.
30. **McGarry, T. J., and S. Lindquist.** 1985. The preferential translation of *Drosophila* hsp70 mRNA requires sequences in the untranslated leader. Cell **42**:903–911.
31. **McGinnis, W., M. S. Levine, E. Hafen, A. Kuroiwa, and W. J. Gehring.** 1984. A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes. Nature (London) **308**:428–433.
32. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
33. **Mueller, P. P., and A. G. Hinnebusch.** 1986. Multiple upstream AUG codons mediate translational control of *GCN4*. Cell **45**:201–207.
34. **Nomoto, A., T. Omata, H. Toyoda, S. Kuge, H. Horie, Y. Kataoka, Y. Genba, Y. Nakano, and N. Imura.** 1982. Complete nucleotide sequence of the attenuated poliovirus Sabin 1 strain genome. Proc. Natl. Acad. Sci. USA **79**:5793–5797.
35. **Nüsslein-Volhard, C.** 1979. Maternal effect mutations that alter the spatial coordinates of the embryo *Drosophila melanogaster*, p. 185–211. *In* S. Subtelny and I. R. Konigsberg (ed.), Determinants of spatial organization. Academic Press, Inc., New York.
36. **Nüsslein-Volhard, C., and E. Wieschaus.** 1980. Mutations affecting segment number and polarity in *Drosophila*. Nature (London) **287**:795–801.
37. **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.
38. **Pardue, M. L., and J. G. Gall.** 1975. Nucleic acid hybridization to the DNA of cytological preparations. Methods Cell Biol. **10**:1–16.
39. **Pelletier, J., and N. Sonenberg.** 1985. Insertional mutagenesis to increase secondary structure within the 5′ non-coding region of a eukaryotic mRNA reduces translational efficiency. Cell **40**:515–526.
40. **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**:37–43.
41. **Proudfoot, N. J., and G. G. Brownless.** 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
42. **Regulski, M., K. Harding, R. Kostriken, F. Karch, M. Levine, and W. McGinnis.** 1985. Homeo box genes of the Antennapedia and bithorax complexes of *Drosophila*. Cell **43**:71–80.
43. **Richter, J. D., and L. D. Smith.** 1984. Reversible inhibition of translation by *Xenopus* oocyte-specific proteins. Nature (London) **309**:378–380.
44. **Sanger, F., A. R. Coulson, B. G. 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.
45. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
46. **Sato, T., P. H. Hayes, and R. E. Denell.** 1985. Homoeosis in *Drosophila*: roles and spatial patterns of expression of the Antennapedia and Sex combs reduced loci in embryogenesis. Dev. Biol. **111**:171–192.
47. **Savakis, C., M. Ashburner, and J. H. Willis.** 1986. The expression of the gene coding for alcohol dehydrogenase during the development of *Drosophila melanogaster*. Dev. Biol.

114:194–207.

48. **Schibler, U., O. Hagenbüchle, P. K. Wellauer, and A. Pittet.** 1983. Two promoters of different strengths control the transcription of the mouse α-amylase gene *Amy-1*^{A} in the parotid gland and the liver. Cell **33**:501–508.
49. **Schneuwly, S., and W. J. Gehring.** 1985. Homeotic transformation of thorax into head: developmental analysis of a new Antennapedia allele in *Drosophila melanogaster*. Dev. Biol. **108**:377–386.
50. **Schneuwly, S., A. Kuroiwa, P. Baumgartner, and W. J. Gehring.** 1986. Structural organization and sequence of the homeotic gene Antennapedia of *Drosophila melanogaster*. EMBO J. **5**:733–739.
51. **Scott, M. P., and A. J. Weiner.** 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of *Drosophila*. Proc. Natl. Acad. Sci. USA **81**:4115–4119.
52. **Scott, M. P., A. J. Weiner, T. I. Hazelrigg, B. A. Polisky, V. Pirrotta, F. Scalenghe, and T. C. Kaufman.** 1983. The molecular organization of the Antennapedia locus of *Drosophila*. Cell **35**:763–776.
53. **Shepherd, J. C. W., W. McGinnis, A. E. Carrasco, E. M. De Robertis, and W. J. Gehring.** 1984. Fly and frog homoeo domains show homologies with yeast mating type regulatory proteins. Nature (London) **310**:70–71.
54. **Struhl, G.** 1981. A homeotic mutation transforming leg to antenna in *Drosophila*. Nature (London) **292**:635–638.
55. **Struhl, G.** 1982. Genes controlling segmental specification in the *Drosophila* thorax. Proc. Natl. Acad. Sci. USA **79**:7380–7384.
56. **Struhl, G.** 1983. Role of the *esc*$^{+}$ product in ensuring the selective expression of segment-specific homeotic genes in *Drosophila*. J. Embryol. Exp. Morphol. **76**:297–331.
57. **Thomas, P. S.** 1983. Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods Enzymol. **100**:255–266.
58. **Wakimoto, B. T., and T. C. Kaufman.** 1981. Analysis of larval segmentation in lethal genotypes associated with the Antennapedia gene complex. Dev. Biol. **81**:51–64.
59. **Wedeen, C., K. Harding, and M. Levine.** 1986. Spatial regulation of Antennapedia and Bithorax gene expression by the Polycomb locus in *Drosophila*. Cell **44**:739–748.
60. **Wharton, K. A., B. Yedvobnick, V. G. Finnerty, and S. Artavanis-Tsakonas.** 1985. Opa: a novel family of transcribed repeats shared by the Notch locus and other develomentally regulated loci in *D. melanogaster*. Cell **40**:55–62.
61. **White, R. A. H., and M. Wilcox.** 1984. Protein products of the bithorax complex in *Drosophila*. Cell **39**:163–171.
62. **Zinn, K., D. DiMaio, and T. Maniatis.** 1983. Identification of two distinct regulatory regions adjacent to the human β-interferon gene. Cell **34**:865–879.