The EMBO Journal vol.5 no.4 pp.733–739, 1986

Structural organization and sequence of the homeotic gene *Antennapedia* of *Drosophila melanogaster*

Stephan Schneuwly, Atsushi Kuroiwa[1], Paul Baumgartner and Walter J.Gehring

Department of Cell Biology, Biozentrum, University of Basel, Klingelbergstr. 70, CH-4056 Basel, Switzerland

[1]Present address: Metropolitan Institute for Neurosciences, Fuchu-City, Tokyo, Japan

Communicated by W.J.Gehring

The structure of the *Drosophila melanogaster Antennapedia (Antp)* gene has been investigated by the isolation and sequencing of different cDNAs and genomic clones. Northern analysis, S1 mapping and primer extension experiments reveal a complex and unusual gene structure. The gene is composed of two promoters, eight exons spanning >100 kb, and two termination processing regions. Four major polyadenylated transcripts were found, two of them starting at a second internal promoter in front of exon 3. All four transcripts have extremely long untranslated leader and trailer sequences in the range of 1–2 kb. Despite the complex transcriptional organization, the open reading frame is the same in all transcripts, and starts in exon 5 giving rise to a protein of mol. wt. 42 800. The putative protein is rich in glutamine (18%) and proline (10%). The homeobox, a region which previously has been shown to be highly conserved among homeotic genes, is contained in the open reading frame and located in the last exon. Functional implications of the complex structure with respect to development and its relation to the mutant phenotypes are discussed.

Key words: Drosophila melanogaster/homeotic genes/gene structure/DNA sequence/internal promoter

Introduction

Different groups of genes have been shown to control the spatial organization of the *Drosophila* embryos. First, there are maternally active genes, which specify the antero–posterior and dorso–ventral axes of the embryo (Gehring, 1973; Nüsslein-Volhard, 1979). A second group of genes, the segmentation genes, determine the number and polarity of the segments (Nüsslein-Volhard and Wieschaus, 1980) and thirdly, the homeotic genes which appear to specify segment identity. Most of the homeotic genes are clustered in two gene complexes on the third chromosome of *D. melanogaster*, the Bithorax complex (BX-C) (Lewis, 1978) and the Antennapedia complex (ANT-C) (for a review see Kaufman, 1983) which contains the *Antennapedia* (*Antp*) gene. The most striking feature of mutants at the *Antp* locus is a transformation of the antennae into mesothoracic legs, which represents a dominant gain of function. The loss of function due to the deletion of the *Antp* locus leads to a recessive lethal phenotype, which involves the transformations of thoracic structures into more anterior structures including parts of the head (Schneuwly and Gehring, 1985; Struhl, 1981; Wakimoto and Kaufman, 1981; Sato *et al.*, 1985). This is a transformation in the anterior direction whereas the dominant phenotype involves a posterior transformation.

To obtain insights to the molecular basis of such complex mutants, it was important to isolate the relevant genes. Recent developments in gene cloning made it possible to isolate several homeotic genes by 'chromosomal walking' (Bender *et al.*, 1983) including *Antp* (Garber *et al.*, 1983; Scott *et al.*, 1983). *In situ* hybridization to tissue sections allowed us to show that *Antp* gene expression during development coincides with the mutant defects observed in recessive *Antp* alleles (Hafen *et al.*, 1983; Levine *et al.*, 1983; Schneuwly and Gehring, 1985). In this study we present the gene organization, the sequence of different cDNAs and the corresponding genomic exons. The 5′ and 3′ ends were mapped by primer extension and S1 analysis and the transcriptional activity of the locus was investigated. A second, internal promoter was found in front of the third exon and at least two termination processing regions were detected.

Results

Isolation and sequencing of different cDNAs and the corresponding genomic subclones

The chromosomal region covering the *Antp* locus and two cDNAs (909 and 903) were isolated previously in our laboratory (Garber *et al.*, 1983). Since these two cDNAs are incomplete and differ substantially in structure, we isolated additional cDNA clones (303 and 311) from a different cDNA library, kindly provided by L.Kauvar (Poole *et al.*, 1985). A map of the different cDNAs analyzed in this study and the corresponding genomic subclones is shown in Figure 1.

The complete sequence of cDNAs 909 (3 kb) and 903 (2.3 kb) was determined, whereas in cDNAs 303 and 311 only the 3′ and 5′ ends which do not overlap with cDNA 909 were sequenced. The comparison of cDNA sequences with those obtained from the corresponding genomic subclones is shown in Figure 2. The gene is composed of eight exons and the first three exons are interrupted by extremely large introns (Figure 1). The largest intron spans ~60 kb and includes the third exon (see below).

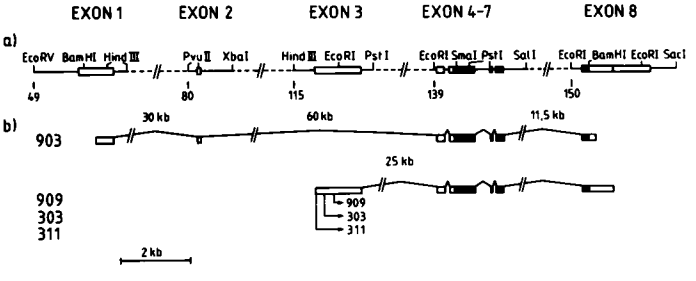
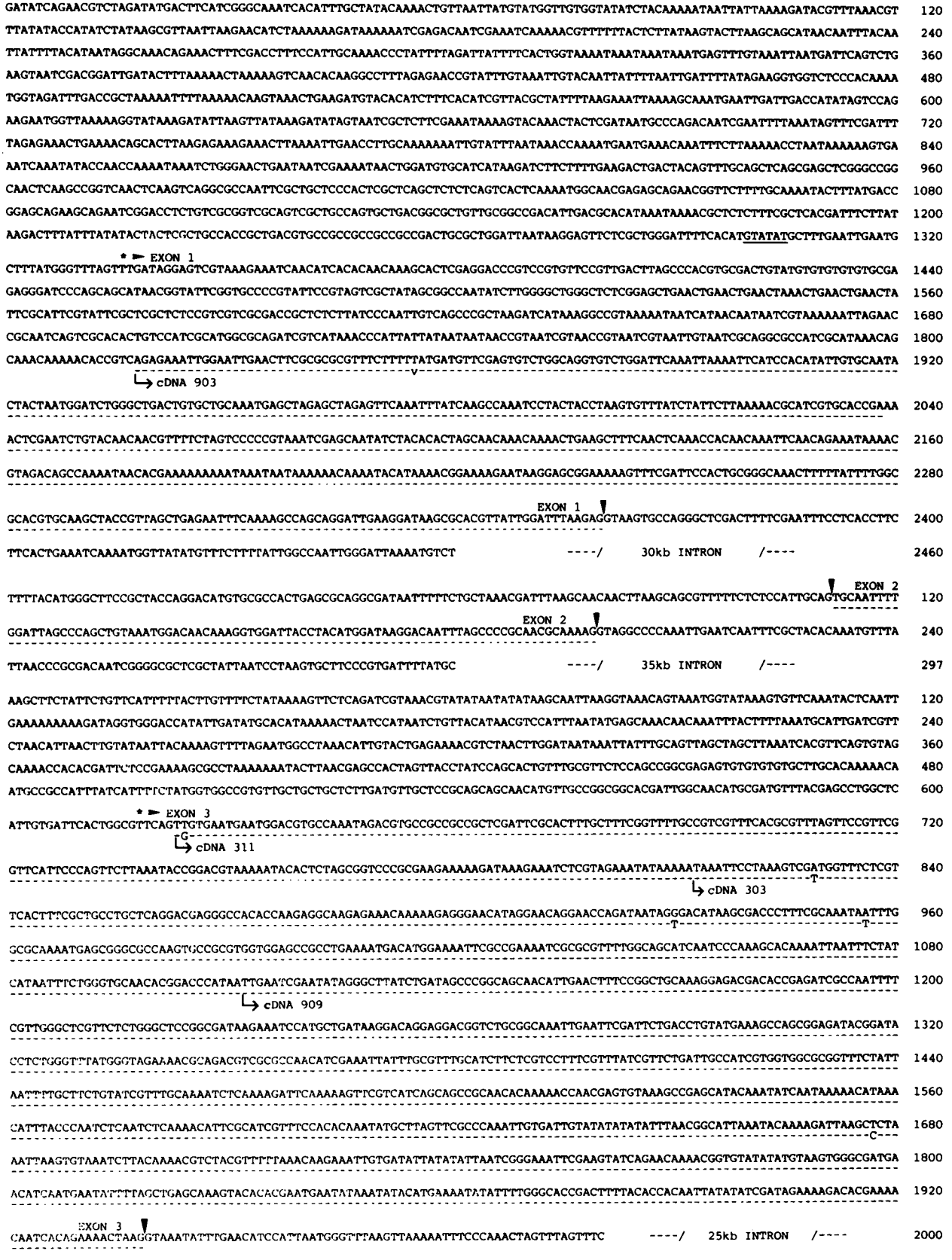


Fig. 1. Organization and restriction map of genomic subclones and cDNAs used for sequencing. (**a**) Genomic subclones of the *Antp* gene. The numbers below the genomic subclones correspond to the map position according to Garber *et al.* (1983). The open boxes show the limits of the different exons whereas the filled boxes correspond to the open reading frame. The large introns are not drawn to scale. Their size is indicated on the figure. (**b**) Different cDNAs used in this study.


EXON 1
cDNA 903
EXON 1
EXON 2
30kb INTRON
EXON 2
cDNA 311
EXON 3
cDNA 303
cDNA 909
EXON 3
25kb INTRON

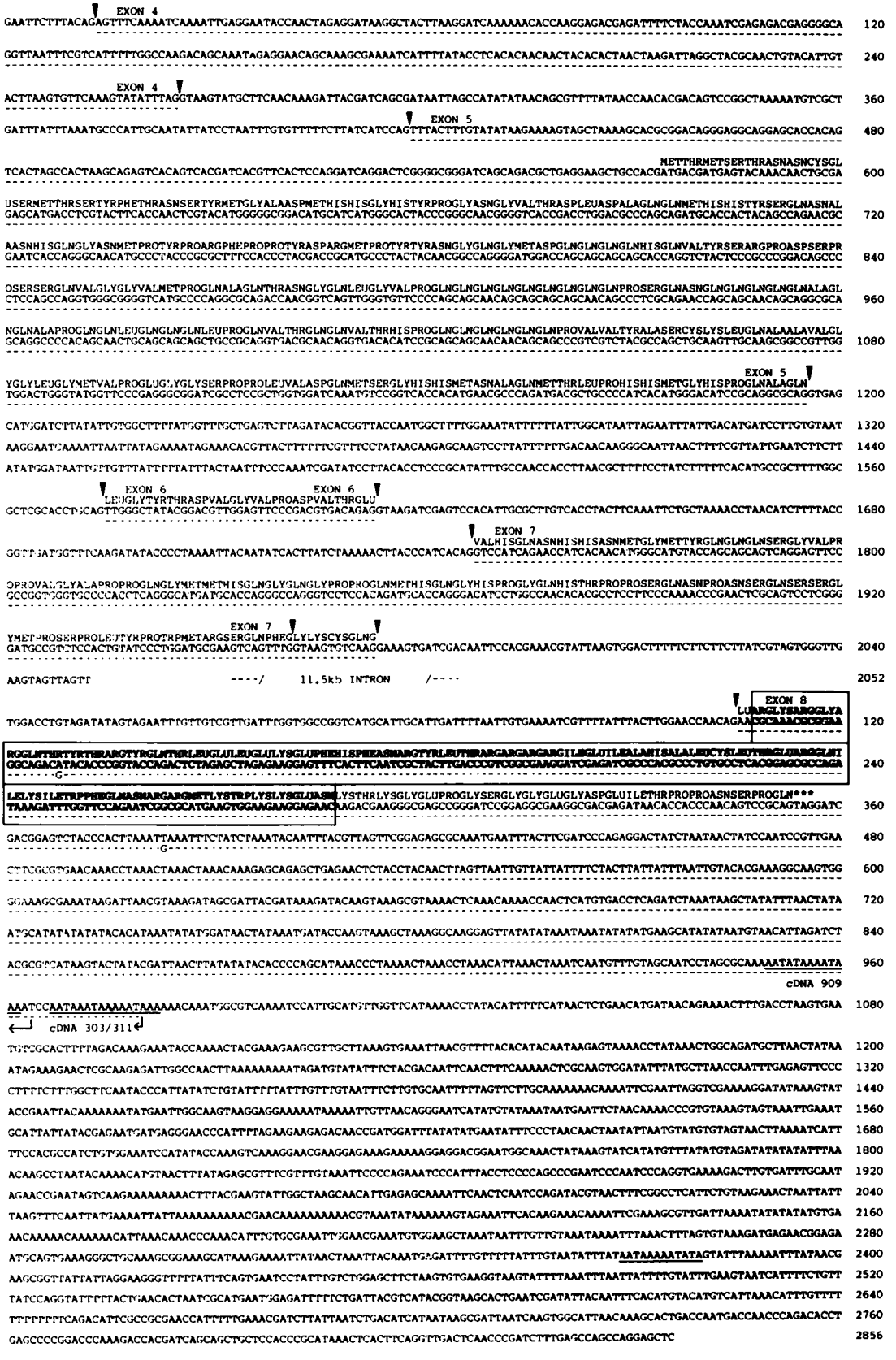


Fig. 2. DNA sequence and deduced amino acid sequence of different *Antp* cDNAs and the corresponding genomic sequences. TATA-box and polyadenylation signals are underlined. The two transcription starts and the translation termination are indicated by asterisks. The limits of all exons and cDNAs are marked by arrowheads. The boxed area corresponds to the homeobox. The four gaps in the sequence correspond to the large introns, which were not sequenced. The splicing difference in cDNA 909 and cDNA 303 in exon 7 is indicated by two different arrowheads.

All exon/intron boundaries (Table I) are in agreement with the proposed concensus sequence (Mount, 1982; Breathnach and Chambon, 1981) except the donor splice site of intron 7, where there is a GA at the beginning of the intron instead of the highly conserved GT. This sequence arrangement has been found in both 903 and 909. However, cDNA 303 has exactly at this position a 12-bp deletion leading to a second splice site which corresponds perfectly to the consensus sequence. We assume that this difference is due to polymorphism in the fly strains that were used in this analysis.

Table I. Intron/exon boundaries

		Donor site		Acceptor site
Consensus		$\overset{C}{A}AG/GT\overset{A}{G}AGT$		$\left(\overset{T}{C}\right)_nN\overset{C}{T}AG/G$
Intron 1		GAG/GTAAGT		TTGCAG/T
Intron 2		AAG/GTAGGC		———
Intron 3		AAG/GTAAAT		TTACAG/A
Intron 4		TAG/GTAAGT		ATCCAG/T
Intron 5		CAG/GTGAGC		CTGCAG/T
Intron 6		GAG/GTAAGA		TCACAG/G
Intron 7	303	TTG/GTAAGT	or	
	903, 909	AAG/GAAAGT		CAACAG/A

Comparison of the donor and acceptor splice sites of the *Antp* gene with the consensus sequences (Mount, 1982). Homologous nucleotides are underlined.

The different cDNAs fall into two classes: cDNA 903 which contains exons 1 and 2 but lacks the third exon, and the second class (including 909, 303 and 311) which lacks the first two exons but contains exon 3. Exons 4–7 and parts of 8 are shared by all cDNAs that we have isolated. This stuctural organization can be explained either by alternative splicing or a second internal promoter in front of exon 3 (see below). Interestingly, despite this difference, all cDNAs share the same large open reading frame encoding a protein with 378 amino acids and a mol. wt. of 42 800. It is possible, that, in addition to the protein encoded by the large open reading frame, small peptides encoded by the two leader sequences are also produced. Comparison of all potential initiation codons (ATG) with the published consensus sequence for eukaryotic initiation codons ($CC\overset{A}{G}CCATG$, Kozak, 1984) shows that the ATG of the long open reading frame is the first one which closely fits the consensus sequence.

Transcriptional activity at the Antp locus

To see how the different cDNAs relate to the actual transcripts at the *Antp* locus, we decided to use exon specific probes for Northern analysis. Figure 3a shows a developmental Northern

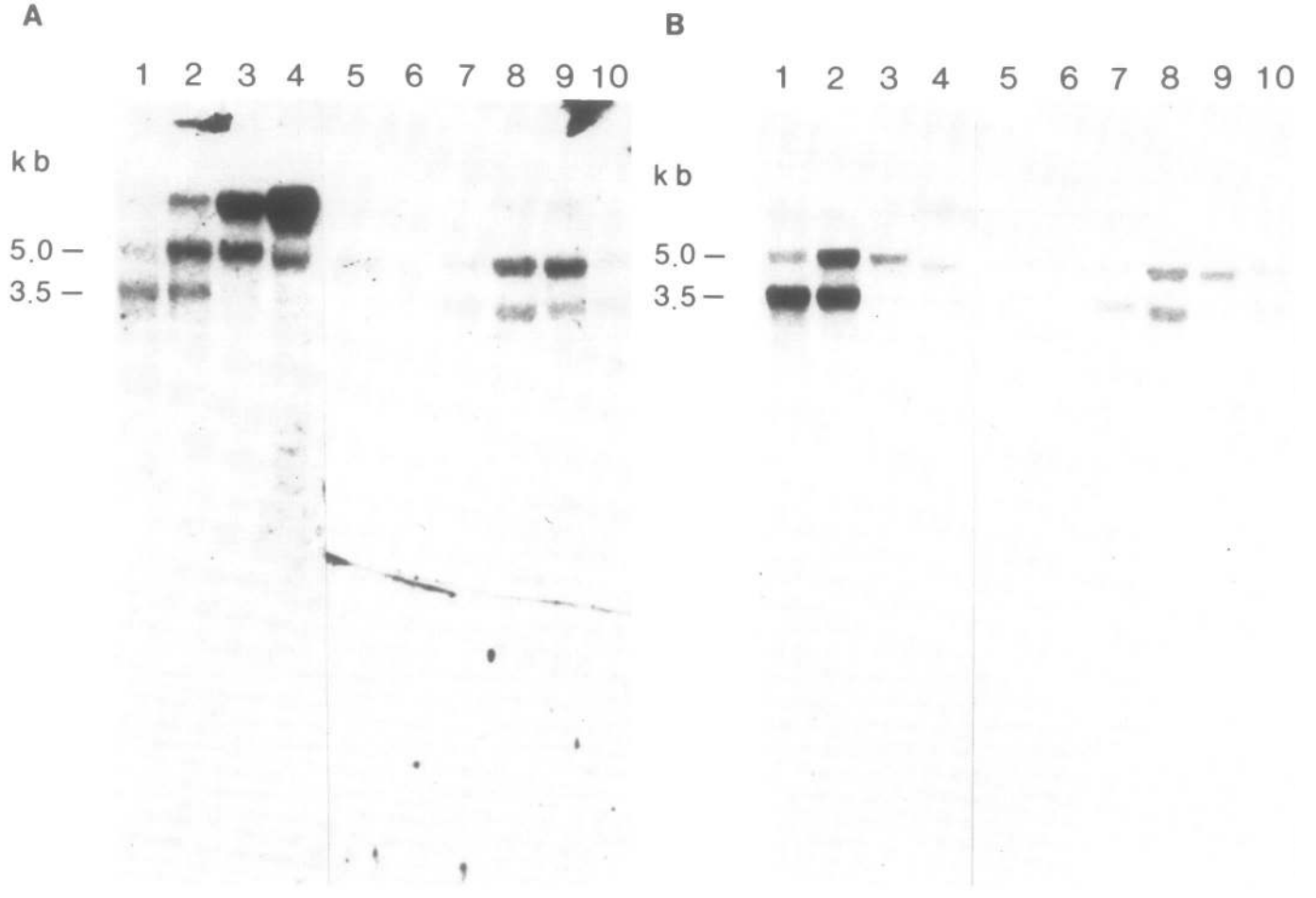


Fig. 3. Northern blot analysis of *Antp* transcripts. Each lane contains 40 μg of poly(A)$^+$ RNA from the following stages: 0–6 h (**lane 1**), 6–12 h (**lane 2**), 12–18 h (**lane 3**), 18–24 h (**lane 4**) of embryonic development and 1st (**lane 5**), 2nd (**lane 6**), 3rd early (**lane 7**), 3rd late (**lane 8**) larval stages and early (**lane 9**) and late (**lane 10**) pupae. Specific transcripts of 3.5 kb and 5.0 kb are marked. The high mol. wt band in some of the lanes is due to DNA contamination. (**A**) Developmental profile using a nick-translated probe derived from exon one. (**B**) Developmental profile using a nick-translated probe derived from exon three.

profile of transcripts homologous to the first exon, whereas in Figure 3b we used a probe which is specific for exon 3. Surprisingly, both developmental profiles and the relative abundance of the different transcripts are very similar. In both cases, there is a predominant 3.5-kb and a 5.0-kb transcript. The 3.5-kb RNA is more abundant in early embryos and in late larval and early pupal stages, whereas the longer transcript is present throughout development although drastically decreased in early larval stages. Using a probe which covers exons 4–8, the developmental profile is exactly the same (data not shown). This might suggest that the transcripts containing the first and the third exon are the same but a more detailed transcriptional analysis (see below) shows that two pairs of transcripts co-migrate just accidently. The whole *Antp* region was further screened for additional transcripts using intron-specific probes, but we detected no additional transcripts in any of the large introns (data not shown).

5′ End analysis of the Antp transcripts

To define the 5′ ends of the different *Antp* mRNAs we performed primer extension analysis using different subclones derived from the first and the third exon. The results of the primer extension reactions are shown in Figure 4A. The first primer, which is 248 nucleotides long and ends at the *Bam*HI site at position 1449 (Figure 2), is extended for 113 nucleotides and therefore places the 5′ end at position 1336 as indicated by an asterisk. Using a second primer derived from the third exon, which is 210 nucleotides long and ends at the *Hin*fI site at position 671, we obtained an extension of 53 nucleotides placing a second transcription start at position 618 in the exon 3 sequence. Both 5′ ends were confirmed by S1 mapping using genomic clones (data not shown). cDNA 311, which is the longest cDNA containing the third exon, lacks only the first five nucleotides and therefore is almost of full length. The upstream region of both transcription initiation sites was analyzed for common promoter sequences (Breathnach and Chambon, 1981) but only in the case of the first promoter did we find a reasonable TATA-box at the appropriate distance. Although there is no TATA-box in front of the second transcription start, we found that the putative CAP site fits well with a *Drosophila* CAP-consensus sequence $ATCA^{G}_{T}T^{T}_{C}$ (Hultmark *et al.*, 1986).

3′ End analysis of the Antp transcripts

cDNA 303 as well as cDNA 311 are polyadenylated and sequence analysis therefore places the 3′ end at position 979 in exon 8 (Figure 2). Indeed, there are polyadenylation signals (AATAAA) (Proudfoot and Brownlee, 1976) just preceding the end of these cDNAs. Summing up the sizes of all exons using either exons 1, 2 and 4–8 or exons 3–8 could account for transcripts with a length of either 3.2 or 3.4 kb [poly(A) tail not included] but not for the large (5 kb) transcripts. For this reason, we performed S1 analysis hybridizing the *Bam*HI–*Sac*I fragment from exon 8 with poly(A)$^+$ RNA. After S1 treatment, the protected fragments were separated on alkaline agarose gels, transferred to nitrocellulose filters and subsequently hybridized with nick-translated fragments. Figure 4B shows the results of the S1 analysis. Using a large fragment (*Bam*HI–*Sac*I) from exon 8, two different fragments are protected. The smaller fragment coincides with the first 3′ end already found in cDNAs 303 and 311. However, there is also a larger fragment that is 1.4 kb longer and therefore places a second 3′ end close to the polyadenylation signals at position 2370 of exon 8. Hybridizing a smaller fragment (*Eco*RI–*Sac*I, see Figure 1) results in only the large fragment appearing on the filter. Northern analysis using the same fragment similarly lights up only the 5-kb transcripts. This clearly

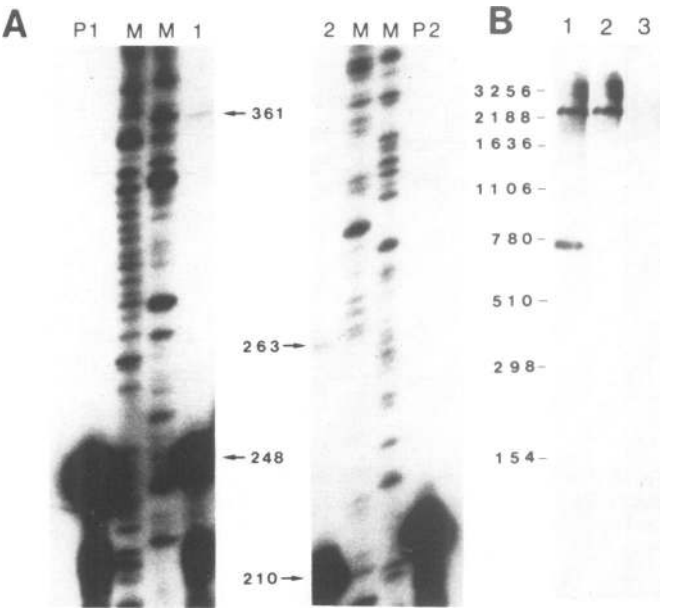


Fig. 4. **(A)** Mapping the sites of transcription initiation by primer extension. 40 μg of poly(A)$^+$ RNA derived from 3rd instar larvae was hybridized in each lane. **Lane 1**: primer extension reaction using a single-stranded primer derived from M13 subclones of the first exon. The primer is 248 nucleotides long, ends at position 1449 (Figure 2) and is extended for 113 nucleotides. **Lane 2**: primer extension reaction using a primer derived from the third exon. The primer is 210 nucleotides long, ends at the *Hin*fI site at position 671 (Figure 2) and is extended for 53 nucleotides. **Lanes P1** and **P2**: appropriate primers before extension reaction. **Lane M**: sequencing reactions as size markers. **(B)** Defining the transcription termination sites by S1 nuclease mapping. The chromosomal DNA fragment *Bam*HI–*Sac*I from exon eight (Figure 1) was hybridized to 40 μg of poly(A)$^+$ RNA derived from 3rd instar larvae, treated wtih S1 nuclease, separated on alkaline agarose gels and transferred to nitrocellulose filters. The filter was hybridized with either the *Bam*HI–*Sac*I fragment (**lane 1**) or the *Eco*RI–*Sac*I fragment (**lane 2**) (see Figure 1). **Lane 3** is a control hybridization using total yeast RNA. Mol. wt. markers are various pBR322 digests.

indicates that the 5-kb transcripts arise by readthrough at the first polyadenylation sites generating a 1.9-kb long trailer sequence which does not show new significant open reading frames.

Discussion

The complex transcription pattern of the *Antp* gene is summarized in Figure 5. There are four major transcripts which differ at the 5′ end as well as at the 3′ end. The difference at the 5′ end is due to a second promoter which is located in front of exon 3. The existence of a second promoter was first inferred from the sizes of the exons, the corresponding transcripts and cDNAs. Primer extension experiments allowed us to map the second site of transcription initiation to be in front of the third exon. The possibility that the different transcripts arise by differential splicing is excluded because there is no acceptor splice site in front of exon 3 and so exon 2 can be directly spliced to exon 4, as it is in cDNA 903. The difference at the 3′ end is due to readthrough of the first polyadenylation signals leading to a much longer trailer sequence. Interestingly, the large open reading frame is not affected by this complex gene organization. Cloning different regions of the predicted open reading frame in *Escherichia coli* expression vectors results in fusion proteins of the expected sizes and antibodies raised against one of the fusion proteins recognize a *Drosophila* protein, which is spatially localized in embryos exactly in those areas where *Antp* transcripts

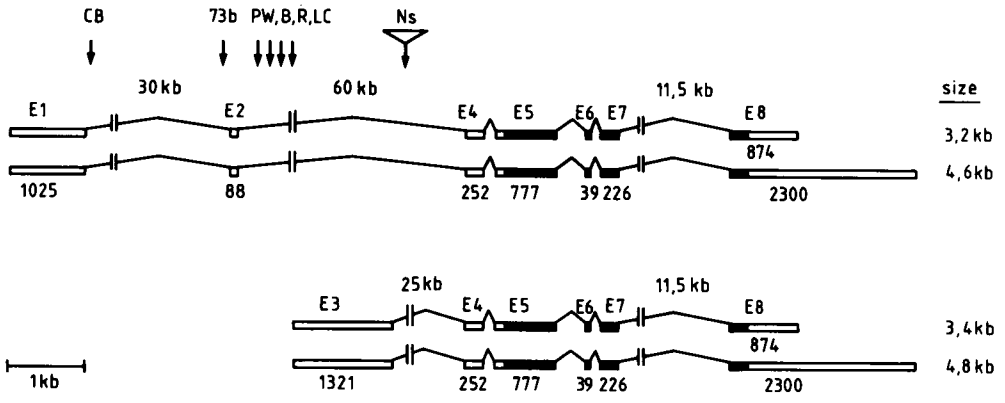


Fig. 5. The transcriptional organization of the *Antp* gene. The open boxes show the different exons. The size is indicated below in base pairs. The open reading frame is indicated by filled boxes. The homeobox is located in exon 8. The large introns are not drawn to scale; their size is indicated in kb. Mutant breakpoints and insertions are indicated by arrows.

have been localized by *in situ* hybridization to tissue sections (J.Wirz, U.Weber and W.Gehring, unpublished results). The amino acid composition of this protein is rather unusual having 18% glutamine and 10% proline. The glutamines are clustered in repeats and flanked by prolines. DNA sequencing homology to this repeat (called M-repeat, McGinnis *et al.*, 1984b) has been found to be present in different developmentally regulated genes (Poole *et al.*, 1985; Wharton *et al.*, 1985; Kuroiwa *et al.*, 1985) and is present in several hundred copies per genome (McGinnis *et al.*, 1984b). It is probably identical to the opa-repeat (Wharton *et al.*, 1985). The most striking feature of the *Antp* protein is certainly the homeobox, a domain which previously has been shown to be highly conserved among different homeotic genes in *Drosophila* and other metazoa including man (McGinnis *et al.*, 1984a). The fact that antibodies raised against different homeotic genes containing the homeobox recognize nuclear proteins (White and Wilcox, 1984; Beachy *et al.*, 1985; Carroll and Scott, 1985; DiNardo *et al.*, 1985), and the homology to the yeast mating-type genes (Shepherd *et al.*, 1984), which in the case of $\alpha 2$ have been shown to encode a DNA-binding protein (Johnson and Herskowitz, 1985), strongly suggest that these proteins control development by directly regulating gene expression.

In Figure 5, we also show the location of previously mapped lesions in dominant mutant alleles of *Antp* (Garber *et al.*, 1983; Scott *et al.*, 1983). Most of the dominant mutants exhibiting an antenna to leg transformation have large cytologically visible inversions where one breakpoint disrupts the *Antp* gene. It is interesting to note that none of these alleles affect the open reading frame but they all separate either the first, or the first and the second exon from the protein coding part. This shows that in the dominant gain of function mutants the protein coding region is not altered but it is separated from one of its controlling regions. Therefore, we assume that the dominant phenotype is not due to an altered gene product but is rather due to altered expression of the same protein. The *Antp* gene product might be expressed from new promoter elements, which have been juxtaposed to the *Antp* gene by an inversion event. There are two exceptional mutants, *Nasobemia* and $Antp^{72j}$, which have the same dominant phenotype but are homozygous viable (Gehring, 1966; Baker, 1974). Both mutants are cytologically normal and are reported to have an insertion of a repetitive element at the *Antp* locus (Scott *et al.*, 1983; confirmed by our unpublished results). In *Nasobemia* it is very likely that this insertion is the cause of the mutant phenotype. Since the insertion is in an intron (see Figure 5), splicing might still produce correct transcripts permitting normal embryonic development, but transcripts originating from a promoter within the repetitive element might be sufficient to cause the dominant phenotype. The second allele, $Antp^{72j}$, was reported to have an insertion beyond the 3′ end of the *Antp* gene (Scott *et al.*, 1983) but the same insertion is also present in Oregon-R, a wild-type strain (unpublished results). Therefore, we consider it unlikely that this insertion causes the *Antp* phenotype.

What is the function of the two promoters? In a simple model, one might propose that each of the two promoters is active during a distinct period of development, as is the case for the *Drosophila* alcohol dehydrogenase gene where there is a larval and an adult promoter (Benyajati *et al.*, 1983). Our Northern experiments show that this is not true for the *Antp* gene because transcripts coming from both promoters show approximately the same developmental profile. This leaves us with the interesting possibility of having two promoters with spatial specificity of expression. Recent experiments involving the fusion of sequences from both promoters to the protein-coding part of the *E.coli lacZ* gene and testing of these constructs in transformed flies provide direct functional evidence for the existence of both promoters (Schneuwly and Gehring, unpublished results). The spatial distribution of the β-galactosidase in these different constructs is currently being investigated. Besides the level of gene activation there might be post-transcriptional mechanisms controlling the spatial distribution of the *Antp* gene product. Recently, it has been shown that several oncogenes are post-transcriptionally regulated by controlling RNA turnover (Piechaczyk *et al.*, 1985). Furthermore, it was shown that alterations of the structure of oncogenes influence messenger stability (Triesman, 1985). The unusual size of the *Antp* leader and trailer sequences leaves plenty of space for signals regulating RNA stability. The fact that transcripts having the longer 3′ trailer are present throughout development could indicate that at least this part of the trailer somehow stabilizes the RNA. In addition, the trailer might also contain signals for control at the level of transcription termination. Our current research efforts are directed towards the understanding of the various control levels of *Antp* gene expression.

Materials and methods

The isolation of genomic lambda phages of the *Antp* region, cDNA 903 and 909 has been previously described (Garber *et al.*, 1983). cDNAs 303 and 311 were isolated from an embryonic cDNA library (E8) kindly provided by L.Kauvar (Poole *et al.*, 1985) using standard screening techniques (Maniatis *et al.*, 1982). Two of the cDNAs, 909 and 303, contain cloning artefacts. As shown by *in situ* hybridization to polytene chromosomes, they contain sequences which do not belong to the *Antp* gene. The structure of these cDNAs clearly suggest that two DNA fragments have been artificially ligated together by a blunt-end ligation during the construction of the cDNA library.

RNA isolation, Northern blotting and S1 mapping were done as described by Kuroiwa *et al.* (1984). ^{32}P-Labeled primers for primer extension were generated using the procedure of Burke (1984) and hybridized overnight (Maniatis *et al.*, 1982). Reverse transcription reaction was carried out using the recommended conditions supplied by the manufacturer (Anglian Biotechnology). Before loading on a 6% sequencing gel, the primer extension reaction was RNAse I treated.

DNA sequencing was performed using the dideoxy chain termination procedure (Sanger *et al.*, 1977). cDNA 909 was sequenced on both strands using a set of overlapping deletions generated by DNase I cuts in the presence of manganese (Lin *et al.*, 1985). Genomic subclones, cDNA 903, and the 5′ and 3′ ends of cDNA 303 and 311 were sequenced using known restriction sites and subcloning in M13 vectors (Messing, 1983). DNA sequencing analysis was performed using the programs of Staden (1977).

Acknowledgements

We thank L. Kauvar and M.Goldschmidt-Clermont for providing their excellent cDNA libraries; R.Staden, C.Nager and T.Bürglin for the use of their computer programs, and Daniel Wey for his help in preparing photographs. We are grateful to Erika Wenger-Marquardt for the efficient typing of the manuscript and Y.Hiromi, P.LeMotte, C.O'Kane and J.Wirz for critically reading it. This work was supported by the Kanton Basel-Stadt and a grant from the Swiss National Science Foundation.

References

Baker,W.K. (1974) *Dros. Inf. Serv.*, **51**, 21.
Beachy,P.A., Helfand,S.L. and Hogness,D.S. (1985) *Nature*, **313**, 545–551.
Benyajati,C., Spoerel,N., Haymerle,H. and Ashburner,M. (1983) *Cell*, **33**, 125–133.
Bender,W., Akam,M.A., Beachy,P.A., Karch,F., Pfeifer,M., Lewis,E.B. and Hogness,D.S. (1983) *Science*, **221**, 23–29.
Breathnach,R. and Chambon,P. (1981) *Ann. Rev. Biochem.*, **50**, 349–383.
Burke,J.F. (1984) *Gene*, **30**, 63–68.
Carroll,S.B. and Scott,M.P. (1985) *Cell*, **43**, 47–57.
DiNardo,S., Kuner,J.M., Theis,J. and O'Farrell,P.H. (1985) *Cell*, **43**, 59–69.
Garber,R.L., Kuroiwa,A. and Gehring,W.J. (1983) *EMBO J.*, **2**, 2027–2036.
Gehring,W.J. (1966) *Arch. Jul. Kl.-Stfg. Vererb. Forsch.*, **41**, 44-54.
Gehring,W.J. (1973) In Ruddle,F.H. (ed.), *Genetic Mechanisms of Development*. Academic Press, NY, pp. 103–128.
Hafen,E., Levine,M., Garber,R.L. and Gehring,W.J. (1983) *EMBO J.*, **2**, 617–623.
Hultmark,D., Klemenz,R. and Gehring,W.J. (1986) *Cell*, **44**, 429–438.
Johnson,A.D. and Herskowitz,I. (1985) *Cell*, **42**, 237–247.
Kaufman,T.C. (1983) In Jefferey,W.R. and Raff,R.A. (eds.) *Time, Space, and Pattern in Embryonic Development*. Alan R.Liss Inc., NY, pp. 365–383.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Kuroiwa,A., Hafen,E. and Gehring,W.J. (1984) *Cell*, **37**, 825–831.
Kuroiwa,A., Kloter,U., Baumgartner,P. and Gehring,W.J. (1985) *EMBO J.*, **4**, 3757–3764.
Levine,M., Hafen,E., Garber,R.L. and Gehring,W.J. (1983) *EMBO J.*, **2**, 2037–2046.
Lewis,E.B. (1978) *Nature*, **276**, 565–570.
Lin,H., Lei,S. and Wilcox,G. (1985) *Anal. Biochem.*, **147**, 114–119.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, NY.
McGinnis,W., Garber,R.L., Wirz,J., Kuroiwa,A. and Gehring,W.J. (1984a) *Cell*, **37**, 403–408.
McGinnis,W., Levine,M.S., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984b) *Nature*, **308**, 428–433.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.
Mount,S. (1982) *Nucleic Acids Res.*, **10**, 459–472.
Nüsslein-Volhard,C. (1979) In Subtelny,S. and Konigsberg,I.R. (eds), *Determinants of Spatial Organization*. Academic Press, NY, pp. 185–211.
Nüsslein-Volhard,C. and Wieschaus,E. (1980) *Nature*, **287**, 795–801.
Piechaczyk,M., Yang,J., Blanchard,J.-M., Jeanteur,P. and Marcu,K.B. (1985) *Cell*, **42**, 589–597.
Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acac. Sci. USA*, **74**, 5463-5467.
Sato,T., Hayes,P.H. and Denell,R.E. (1985) *Dev. Biol.*, **111**, 171–192.
Schneuwly,S. and Gehring,W.J. (1985) *Dev. Biol.*, **108**, 377–386.
Scott,M.P., Weiner,A.J., Hazelrigg,T.I., Polisky,B.A., Pirrotta,V., Scalenghe,F. and Kaufman,T.C. (1983) *Cell*, **35**, 763–776.
Shepherd,J.C.W., McGinnis,W., Carrasco,A.E., De Robertis,E.M. and Gehring,W.J. (1984) *Nature*, **310**, 70–71.
Staden,R. (1977) *Nucleic Acids Res.*, **4**, 4037–4051.
Struhl,G. (1981) *Nature*, **292**, 635–638.
Treisman,R. (1985) *Cell*, **48**, 889–902.
Wakimoto,B.T. and Kaufman,T.C. (1981) *Dev. Biol.*, **81**, 51–64.
Wharton,K.A., Yedvobnick,B., Finnerty,V.G. and Artavanis-Tsakonas,S. (1985) *Cell*, **40**, 55–62.
White,R.A.H. and Wilcox,M. (1984) *Cell*, **39**, 163–171.

Received on 13 January 1986