The EMBO Journal vol.4 no.13B pp.3757–3764, 1985

Cloning of the homeotic *Sex combs reduced* gene in *Drosophila* and *in situ* localization of its transcripts

Atsushi Kuroiwa[1], Urs Kloter, Paul Baumgartner and Walter J.Gehring

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

[1]Present address: Metropolitan Institute of Neurosciences, 2–6 Musashidai, Fuchu-City, Tokyo 113, Japan

Communicated by W.J.Gehring

We have extended our 'chromosomal walk' in the Antennapedia-complex (ANT-C) by isolating overlapping DNA sequences spanning the chromosomal segment between *Antennapedia (Antp)* and *Deformed (Dfd)*. The transcription units, homeoboxes and M-repeats were mapped within this region. Four transcription units *Antp, fushi tarazu (ftz), Sex combs reduced* (*Scr*) and *Dfd* contain both a homeobox and an M repeat, whereas at least two additional transcription units, x and z, were found to lack these elements. The *Scr* locus was identified by deletion mapping. It consists of at least two exonic regions separated by a large intron. The homeobox is located in the 3′ exon and is 82% homologous to the one in *Antp*. *Scr* encodes a major 3.9-kb RNA. A corresponding cDNA clone was used as a probe for *in situ* hybridization to sections of various embryonic stages. At gastrula stages Scr^+ transcripts accumulate in the posterior head and the anterior thoracic region of the germ band. At later stages a strong accumulation of transcripts is observed in the suboesophageal and the prothoracic ganglion of the ventral nervous system.

Key words: Drosophila/homeotic genes/*Sex combs reduced*/cloning/homeobox

Introduction

The basic principle of organization in insect embryogenesis involves the formation of a series of body segments along the anterior-posterior axis. Three major classes of genes have been identified which control this process: maternal-effect genes, expressed during oogenesis, which specify the anterior-posterior and the dorso-ventral axes (Gehring, 1973; Nüsslein-Volhard, 1979; Anderson and Nüsslein-Volhard, 1984); the segmentation genes, which determine the number and the polarity of the segments (Nüsslein-Volhard and Wieschaus, 1980; Wieschaus *et al.*, 1984; Jürgens *et al.*, 1984; Nüsslein-Volhard *et al.*, 1984); and the homeotic genes which control segment identity (see Ouweneel, 1976). Homeotic mutants generally transform parts of a segment (or an entire segment) into the corresponding parts of another segment.

There are two major clusters of homeotic genes in *Drosophila melanogaster* the Bithorax-complex (BX-C) which specifies the identity of the thoracic and abdominal segments (Lewis, 1978; Sanchez-Herrero *et al.*, 1985) and the Antennapedia-complex (ANT-C) which determines the head and the anterior thoracic segments (Kaufman *et al.*, 1980; Wakimoto *et al.*, 1984; Struhl, 1981, 1983). The ANT-C also includes some segmentation genes (Wakimoto *et al.*, 1984).

The homeotic gene *Sex combs reduced* (*Scr*) is a member of the ANT-C. Scr^- mutants show a reduction in the number of sex comb teeth on the male foreleg, which represents a partial transformation of the first to the second thoracic segment, and a partial transformation of the labial palps into maxillary palps (Lewis *et al.*, 1980a, 1980b). The same segmental transformations, prothorax to mesothorax and labial to maxillary, are also found in homozygous Scr^- embryos (Struhl, 1983; Wakimoto and Kaufman, 1981). Genetic and molecular mapping of *Scr* has given ambiguous results since some of the mutants appear to map to the right of *ftz* and some to the left, towards the centromere (Hazelrigg and Kaufman, 1983; Scott *et al.*, 1983). *Multiple sex combs* ($In(3)Msc^1$) appears to be a dominant allele of *Scr*, since hemizygous $In(3)Msc^1/Df(3R)Scr$ animals have the same phenotype as Scr^- homozygotes. However, in addition to having a reduced number of sex comb teeth on the foreleg, $In(3R)Msc^1/+$ males also carry a small number of sex comb teeth on the second legs, indicating a partial transformation in the opposite direction, i.e., from meso- to prothorax.

In earlier reports (Garber *et al.*, 1983; Kuroiwa *et al.*, 1984) we have described the cloning of the distal part of the ANT-C including the *Humeral (Hu), Antp* and *ftz* loci. Here we report the cloning and molecular analysis of the region between *Antp* and *Deformed (Dfd)* with particular emphasis on the *Scr* locus. As expected, *Scr* contains a homeobox which is a small highly conserved DNA segment characteristic for homeotic genes and other controlling genes involved in the spatial organization of the embryo (McGinnis *et al.*, 1984a, 1984b; Scott and Weiner, 1984; Gehring, 1985). The Scr^+ transcripts are localized in those regions of the embryo which are affected in Scr^- mutants.

Results

'Chromosomal walk' from Antennapedia to Deformed

Previously we reported the isolation of overlapping chromosomal DNA segments from the distal part of the ANT-C (Garber *et al.*, 1983) which includes the region from −40 to +194 kb on the map shown in Figure 1. The zero coordinate is defined as the breakpoint of $In(3R)Hu$ which was used to jump into the ANT-C. We continued our 'walk' towards the centromere until we reached lambda 99 at +254, the *Dfd* locus, which had previously been cloned by means of cross-homology to the *Antp* homeobox (McGinnis *et al.*, 1984a). By whole genome Southern blotting, the breakpoints of the chromosomal rearrangements $Df(3R)9A99$, $Df(3R)4Scb$ and $In(3R)Msc^1$ (see Table I) were determined within the cloned chromosome segment. The results are summarized in Figure 1. The distal breakpoint of $Df(3R)9A99$ has been mapped earlier between +167 and +182 (Garber *et al.*, 1983) and now more precisely in the 1.7-kb *Hind*III-*Bam*HI fragment at +174. This deletion removes *proboscipedia (pb), Scr* and *ftz*, leaving *Antp* intact (Jürgens *et al.*, 1984). The distal breakpoint of $Df(3R)4Scb$ was mapped precisely to the 1.6 kb *Hind*III fragment at +215. The proximal breakpoint of $In(3R)Msc^1$ maps at +190 which confirms the data obtained previously by Scott *et al.* (1983).

© IRL Press Limited, Oxford, England

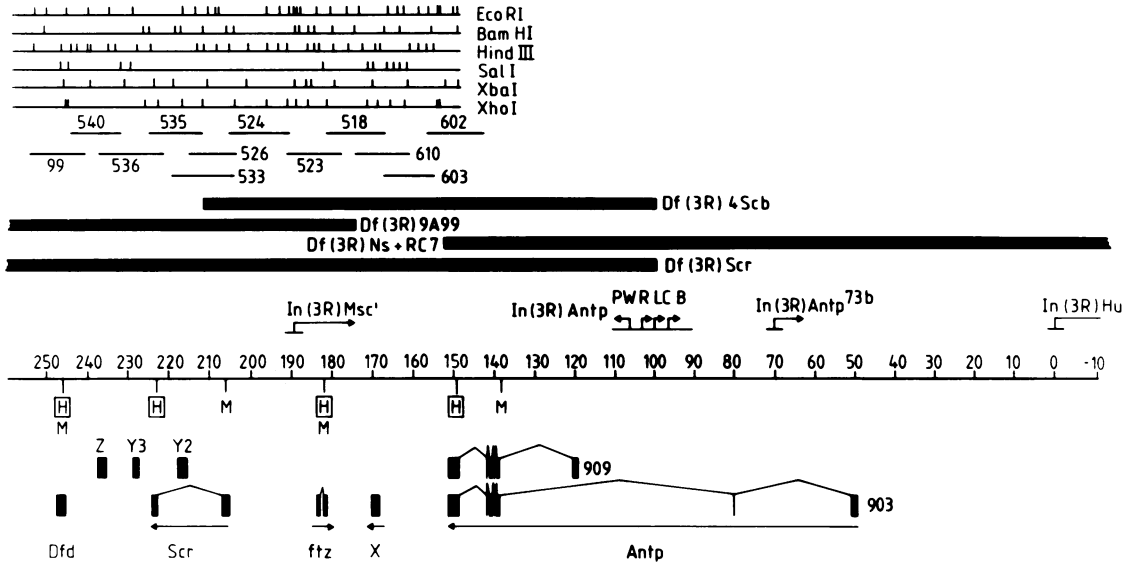

Fig. 1. Map of 'chromosome walk' in the ANT-C. Top: restriction map of the proximal region of the ANT-C. The numbers and extent of the hybrid phage clones used in this study are shown under the restriction maps. The black bars indicate the deleted regions of four deficiencies. The chromosomal breakpoint and directions at various inversions are indicated by arrow. The centromere is to the left, the telomere to the right. The coordinates are given in kb. The breakpoint of *In(3R)Hu* is defined as the 0 coordinate, H = homeobox, M = M repeat. The boxes in the lower part of the figure represent the localization of the exons (not drawn to scale), the connecting lines represent the introns of the respective transcription units. The horizontal arrows indicate the direction of transcription. The map includes unpublished data of S.Schneuwly *et al.* and W.J.McGinnis *et al.*.

Table I. Mutations analyzed

Allele	Cytology	Genes affected in ANT-C	Ref.[a]
Msc[1]	*In(3R)84B1;84F1,2*	*Scr*	A
Scr	*Df(3R)84A1,2;84B1,2*	*lab pb zen Scr ftz Antp*	B
4Scb	*Df(3R)84B1;84B2*	*Scr ftz Antp*	C
9A99	*Df(3R)84A1;84B1,2*	*pb Scr ftz*	C
Ns^{+RC7}	*Df(3R)84B1,2;84D*	*Antp*	D

[a]A: Lindsley and Grell (1968); BB: Wakimoto *et al.* (1984); C: Jürgens *et al.*, (1984); D: Struhl (1981) and Garber *et al.* (1983).

Localization of the homeoboxes and M repeats

The *Antp* cDNA clone (p903) contains a 180-bp DNA sequence, the homeobox, which cross-hybridizes with homologous sequences in *ftz* and *Ubx* (McGinnis *et al.*, 1984a, 1984b; Scott and Weiner, 1984). Using this DNA segment as a probe we have cloned various other homeotic genes of *Drosophila* and the corresponding genes from other metazoa, including vertebrates (for review, see Gehring, 1985). Using the *Antp* homeobox (p903G) as a probe, we identified four homologous regions along the 'chromosomal walk'. These correspond to the three previously known homeoboxes in *Antp*, *ftz* and *Dfd* (see Figure 1), and a new one at +222 kb (Figure 2a, lanes e – h) which corresponds to *Scr*, as will be shown below.

Antp cDNA carries another repetitive element, the M repeat (McGinnis *et al.*, 1984a) which is present >100 times per genome and codes for stretches of polyamino acids within the homeotic proteins. It also occurs in several other homeotic genes and in the *Notch* locus where it has been designated *opa* (Wharton *et al.*, 1985) but preliminary sequencing data indicate that it may not code for the same polyamino acid in all cases. Using the *Pvu*II-*Pvu*II subclone of the *Antp* cDNA clone p903 (McGinnis *et al.*, 1984a), we detected three areas of cross-homology to the left of *Antp* under reduced stringency conditions of hybridization: the 5-kb *Eco*RI fragment of lambda 99 (*Dfd*), a 0.4-kb fragment at +205 and more weakly the 3.6-kb *Eco*RI fragment containing *ftz* (data not shown). The location of these repetitive elements is summarized in Figure 1.

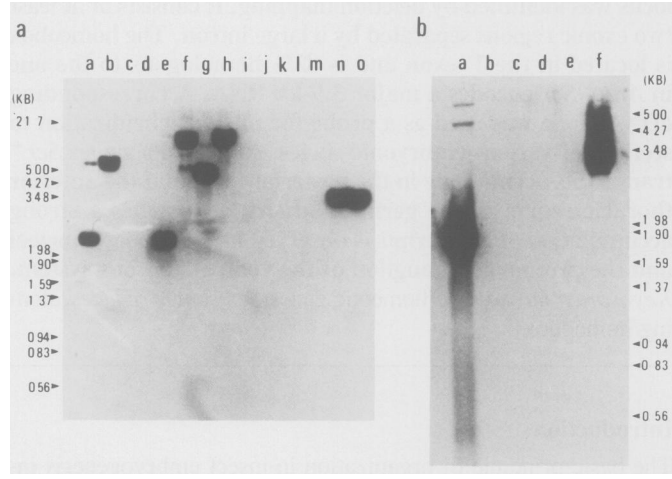

Fig. 2. Localization of the homeoboxes. Southern blots of cloned DNA segments digested with the respective restriction enzymes were probed with the *Antp* homeobox probe (p903G) under low stringency hybridization conditions (see McGinnis *et al.*, 1984a). The clones displayed on the blots are arranged according to their order along the chromosome (see Figure 1). The hybridization signals in **a, lanes a** and **b** correspond to the homeobox in *Dfd*, in **lanes e – h** to *Scr*, **lanes n** and **o** to *ftz*, and **lane f** to *Antp*. The clones and enzymes used are the following: **a: a,** 99 (*Kpn*I and *Eco*RI); **b,** 99 (*Eco*RI); **c,** 540 (*Eco*RI and *Xba*I); **d,** 540 (*Eco*RI); **e,** 536 (*Eco*RI and *Xba*I); **f,** 536 (*Eco*RI); **g,** 535 (*Eco*RI and *Xba*I); **h,** 535 (*Eco*RI); **i,** 533 (*Eco*RI); **j,** 526 (*Hind*III and *Eco*RI); **k,** 526 (*Eco*RI); **l,** 524 (*Hind*III and *Eco*RI); **m,** 524 (*Eco*RI); **n,** 523 (*Eco*RI and *Xba*I); **o,** 523 (*Eco*RI). **b: a,** 523 (*Eco*RI); **b,** 518 (*Eco*RI); **c,** 518 (*Eco*RI and *Xba*I); **d,** 610 (*Bam*HI); **e,** 603 (*Bam*HI and *Sal*I). An *Eco*RI and *Hind*III digest of lambda DNA was used as a standard mol. wt. marker.

Northern analysis of the transcripts

The transcription units localized in this region were analyzed by Northern blotting using poly(A)$^+$ RNA from various developmental stages and probing with cloned DNA fragments from

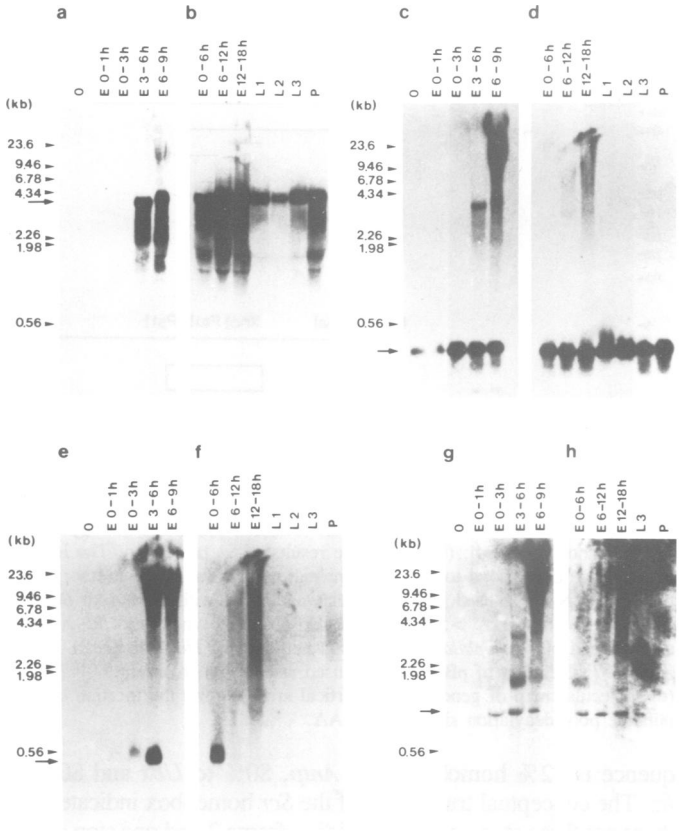

Fig. 3. Northern analysis of the transcripts. Poly(A)$^+$ RNA from various stages of the *Drosophila* life cycle was fractionated on formaldehyde-containing 0.8% agarose gels. After transferring the RNA to nitrocellulose filters, the transcripts were detected by hybridization with nick-translated DNA from various parts of the 'chromosome walk'. A *Hind*III digest of lambda DNA was used as a mol. wt. marker. The same pair of filters was reused with four different probes after removal of the previous probe. The various developmental stages are labelled in the figure. O = unfertilized eggs, E = embryos, L1, L2, L3 = 1st, 2nd and 3rd instar larvae, P = pupae (1 day old). (**a** and **b**) Hybridization profile of the DNA fragment +222.2 to +230.0. A major 3.9-kb transcript is detected (arrow). (**c** and **d**) Hybridization profile of the DNA fragment +217.7 to +220.0. The high mol. wt. smear in the 6−9 h embryo lane is due to contaminating DNA and the 3.9-kb band in the same and the adjacent lane is an artefact caused by incomplete removal of the probe used in **a**. The size of the Y2 transcript was estimated by extrapolation to be ~380 nucleotides (arrow). (**e** and **f**) Hybridization profile of the 1.3-kb *Hind*III fragment from +168.5 to +169.8. The high mol. wt. smear is again due to contaminating DNA in these samples. An early transcript of ~520 nucleotides is detected (arrow). (**g** and **h**) Hybridization profile of the 4.8-kb *Eco*RI fragment from +236.0 to +240.8. A 1.2-kb transcipt z is detected (arrow). The band at 1.8 kb appears to be an artefact caused by the contaminating rRNA and the high mol. wt. smearing is due to DNA contamination.

+150 to +246. By this method we found at least three transcription units in addition to *ftz* in this region. The fragments from +204.0 to +209.0 and from +222.2 to +230.0 which carry an M repeat and a homeobox, respectively, hybridize to a 3.9-kb RNA (Figure 3a and b, and our unpublished results). This RNA can be detected as early as 3−4 h after oviposition and continues to be expressed throughout development at least until the pupal stage. The intensity of the hybridization signal on Northern blots is about equal to that of *Antp*. On the basis of the analysis of the two deletions *Df(3R)9A99* and *Df(3R)4Scb* presented above we assign this 3.9-kb transcript to the *Scr* locus. These two deletions overlap in the area from +215 to +174 and remove both *ftz* and *Scr*. Since *ftz* has been identified definitely by germ line transformation (Hiromi *et al.*, in preparation), we can assign the second transcript in this section to *Scr*. This conclusion is consistent with the presence of a homeobox and strongly supported by our *in situ* hybridization data on the localization of the transcripts in the embryo.

The DNA fragments from +217.7 to +220.0 and +227.0 to +230.0 (Figure 1) hybridize on Northern blots to a small RNA of ~380 nucleotides, the transcripts Y2 and Y3 (Figure 3c and d; data for Y3 not shown). This RNA is not only present in embryonic, larval and pupal stages, but also in the unfertilized egg. The amount of this RNA present in 6−12 h embryos is ~10 times higher than that of the 3.9-kb RNA.

The 1.3-kb *Hind*III fragment from +168.5 to 169.8 hybridizes to a 0.52-kb RNA (Figure 3e and f). This transcript is only present at early embryonic stages, with a maximum between 3 and 6 h after oviposition. We designate the gene coding for this transcript as gene X, until it can be assigned to a mutant phenotype. The hybridization signal is several fold lower than that of *ftz*.

Finally, a 1.2-kb transcript *z* was detected by using the 4.8-kb or 5.7-kb *Eco*RI fragments from +236.0 to +246.5 as probes (Figure 3a and h).

Mapping of the Scr transcript

Those chromosomal DNA fragments carrying the *Scr* exons as determined by Northern analysis were examined further by S1 mapping. Poly(A)$^+$ RNA isolated from 6−12 h embryos was hybridized to chromosome DNA fragments, digested with S1 nuclease and the protected fragments were separated on agarose gels. The results are shown in Figure 4a and b. First, a 1.0-kb fragment is protected by the 5.1-kb chromosomal DNA fragment from coordinates +204.0 to +209.1 (lane a in Figure 4). Cleavage of this chromosomal fragment with *Hind*III yields fragment e which protects a 0.76-kb fragment (lane e in Figure 4b). Second, a 2.6-kb fragment was protected by the DNA fragment from coordinate +222.2 to +230.0 (lane g in Figure 4b). Cleavage of this chromosomal DNA fragment with *Xba*I yields fragment i which protects a shorter 1.9-kb fragment (lane i in Figure 4b). Thus, *Scr* has at least two exons, a 5′ exon with an M repeat and a 3′ exon carrying the homeobox.

To determine the structural organization of the *Scr* gene more precisely, we screened a cDNA library constructed from poly(A)$^+$ RNA isolated from 3−12 h *Drosophila* embryos (kindly provided by Kauvar, Drees, Poole and Kornberg) for *Scr* cDNA clones. From ~2 × 10^6 plaques, six *Scr* cDNA clones were isolated. Clones cY20 and cY22 are identical and carry the largest insert of 3.2 kb including both exons. The other four clones are shorter and carry only the 3′ exon. The structure of these cDNA clones and a summary of the S1 mapping experiments are shown in Figure 5.

Mapping of the transcript of gene X

DNA fragments around coordinate +170 were used to determine the exonic regions of gene X. Fragments a, e and f (Figure 6) protect a fragment of 370 nucleotides, whereas fragment c, which overlaps with e, protects a fragment of only 240 nucleotides. Thus gene X was mapped as indicated in Figure 6c. Since the protected fragment of 370 bases is shorter than the RNA transcript by ~150 bases, there may still be another short exon, assuming that the poly(A) tail does not exceed 100 bases. Preliminary sequencing data show that there is a polyadenylation signal (AATAAA) 39 bp downstream of the *Xho*I site. Therefore, we may be missing a small exon at the 5′ end of the gene.

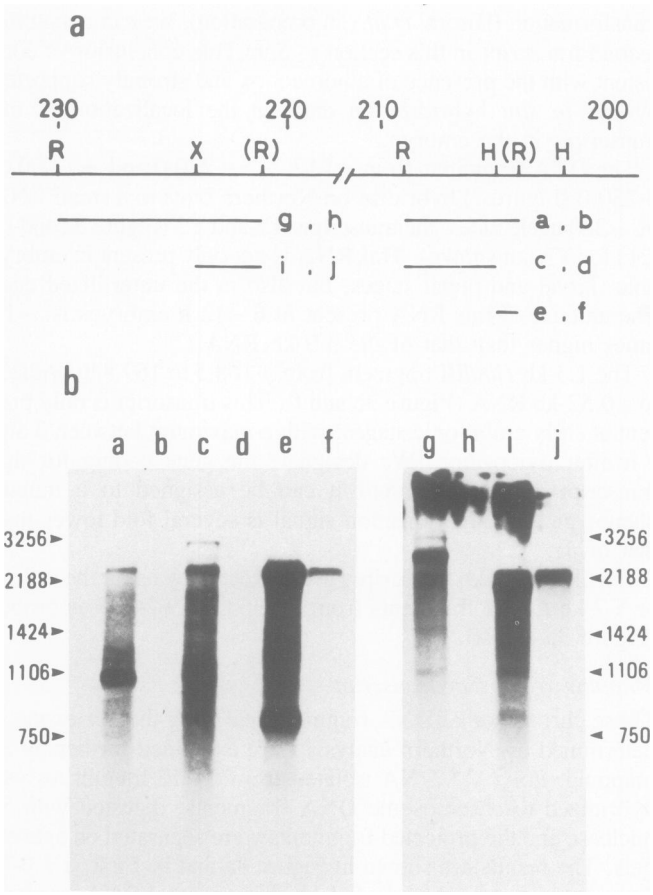

Fig. 4. S1 mapping of the 3.9-kb *Scr* transcript. (**a**) Cloned chromosomal DNA fragments used for the S1 protection experiments. R = *Eco*RI, (R) = artificial *Eco*RI site, X = *Xba*I, H = *Hind*III. (**b**) Results of S1 protection. The letters above each lane correspond to the fragment indicated in (**a**). **Lanes a, c, e, g** and **i** show the results of S1 protection by poly(A)$^+$ RNA from 6–12 h *Drosophila* embryos. **Lane b, d, f, g** and **j** show the protection by yeast RNA as a control. **Lanes a–f** are probed with the fragment used for **a** and **b** reaction. **Lanes g–j** are probed with the fragment used for the **g** and **h** reaction. *Hinc*II and *Eco*RI + *Ava*I + *Pst*I digests of pBR322 were used as standard mol. wt. markers.

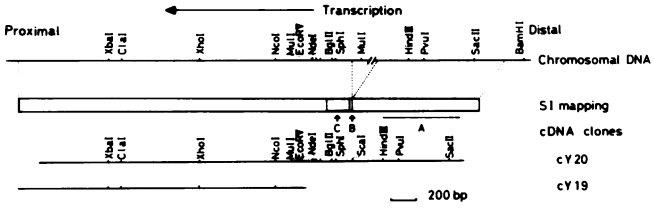

Fig. 5. Molecular map of the *Scr* locus and the 3.9-kb transcript. Summary of the S1 analysis and restriction map of the chromosomal and cDNA. A indicates the location of the M repeat, B shows the exon junction, and C represents the homeobox. The direction of transcription is indicated by an arrow.

The sequence of the Scr homeobox

To determine the nucleotide sequence of the *Scr* homeobox, the smallest DNA fragment which cross-hybridizes with the *Antp* homeobox was subcloned into the M13 vector and sequenced by the dideoxy method (Sanger *et al.*, 1977). The corresponding *Scr* cDNA fragment was also sequenced. The sequence comparisons between the homeoboxes of *Scr, Antp, Ubx* and *ftz* are shown in Figure 7a and Table II. At the DNA level the *Scr* sequence is 82% homologous to *Antp*, 80% to *Ubx* and 80% to *ftz*. The conceptual translation of the *Scr* homeobox indicates that there are three stop codons in reading frame 2 and one stop codon in reading frame 3 upstream of the homeobox in the cDNA, we conclude that frame 1 is used, which is in agreement with the sequencing data obtained for other homeoboxes. Since most of the base changes occur in the third position of the codons, the homology to other homeoboxes is greater at the protein level. It is highest for *Antp* (92%) and lower for *Ubx* and *ftz* (83% and 80%, respectively). The only striking difference between *Scr* and the other homeoboxes is found in amino acid 6 and 7 (Thr and Ser), which are Gln and Thr in *Antp, Ubx* and *ftz*. Amino acids 8–35 and 37–59 are identical in *Scr* and *Antp*. The sequences on either side of the homeobox in the cDNAs diverge strongly from one another.

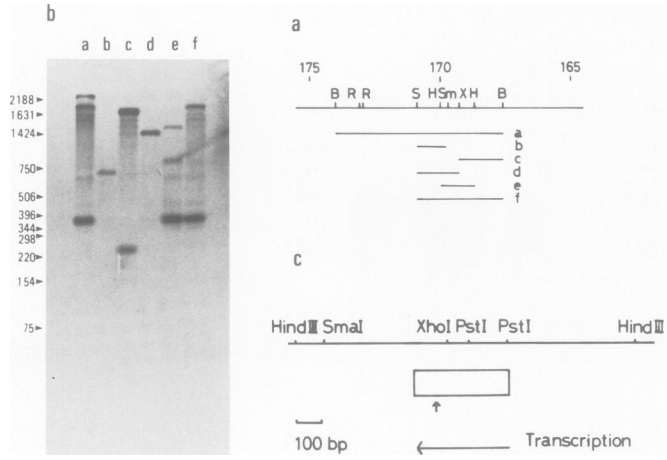

Fig. 6. S1 mapping of gene X. (**a**) Shows the DNA fragments used for S1 protection analysis. R = *Eco*RI, B = *Bam*HI, H = *Hind*III, S = *Sal*I, Sm = *Sma*I and X = *Xho*I. (**b**) Shows the results of S1 protection. The letters in each lane correspond to the fragment indicated in (**a**). 0–6 h embryonic poly(A)$^+$ RNA was used, and the filter was probed with the 6.4-kb *Bam*HI fragment a. The bands >500 bp were also observed when yeast RNA was used instead of *Drosophila* RNA (results not shown). *Hinf*I or *Eco*RI + *Ava*I + *Pst*I digests of pBR322 were used as mol. wt. markers.
(**c**) Molecular map of gene X. The vertical arrow shows the location of the putative polyadenylation signal AATAAA.

Localization of the Scr$^+$ and X$^+$ transcripts by in situ hybridization

The *Scr* cDNA clones (pcY20) was used as a probe for *in situ* hybridization to the *Scr*$^+$ transcripts in wild-type embryos. In sections of embryos during a germband extension (~5 h after oviposition) the transcripts accumulate primarily in one region of the germband on the anterior ventral side of the embryo (Figure 8a and b). According to the fate map (Hartenstein and Campos-Ortega, 1985) this region corresponds to the labial and prothoracic segments of the developing embryo. After germband retraction, when segmentation becomes clearly visible in the nervous system, the highest density of silver grains is found over the suboesophageal ganglion and the anterior part of the prothoracic ganglion (Figure 8c–f). Weaker but significant labeling is found over the entire nervous system, including the brain. The observed segmental localization of the transcripts corresponds to the localization of the defects observed in *Scr*$^-$ mutants (see Introduction). A second site of strong labeling can be detected in the pharynx region. Since there are no good morphological markers at this stage, we cannot assign this site to a certain segment, and later development stages will have to be examined.

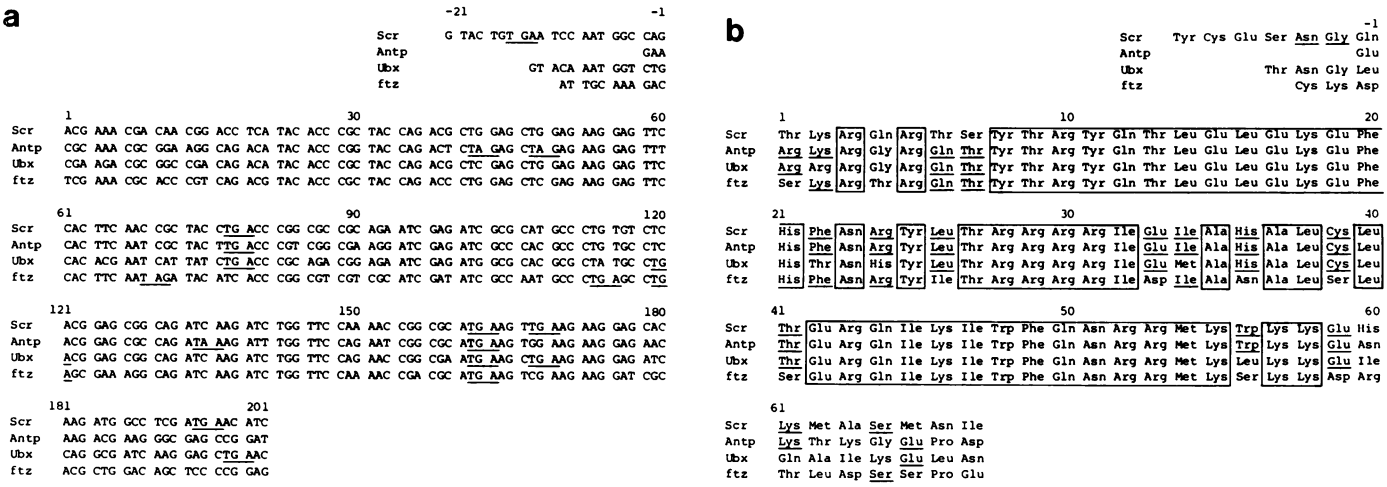

Fig. 7. Sequence analysis of the *Scr* homeobox. (**a**) DNA sequence of the homeoboxes in *Scr, Antp, Ubx* and *ftz* (5′ to 3′). All sequences start with the first nucleotide of the exon carrying the homeobox. The codons are aligned according to the common open reading frame. Stop codons in the other reading frames are underlined. The sequences for *Antp, Ubx* and *ftz* are from McGinnis *et al.* (1984a, 1984b) and Scott and Weiner (1984). (**b**) Conceptual translation of the common open reading frame of *Scr, Antp, Ubx* and *ftz*. Amino acids which are invariant in all four homeo domains are boxed in. Amino acids shared by at least two genes are underlined.

Table II. Comparison of homeobox sequences

	Scr Antp	*Scr Ubx*	*Scr ftz*	*Antp Ubx*	*Antp ftz*	*Ubx ftz*
Homologous nucleotides (out of 180)	148	144	144	145	138	134
% Homology	82	80	80	81	77	74
Position in codon	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
Differences	7 4 21	8 7 21	10 5 21	5 5 25	10 5 27	12 8 26
Homologous amino acid (out of 60)	55	50	48	54	50	46
% Homology	92	83	80	90	83	77

As indicated by Northern experiments, gene X is most strongly expressed in 3−6 h embryos. The 1.3-kb *Hind*III fragment (+168.7 to 170.0) which contains most of the RNA coding region of gene X was used for *in situ* hybridization to sections across early embryos. At the blastoderm stage, a broad band of hybridization signals is observed from ~50 to 65% egg length (0% = posterior pole), Figure 9a and b. Tangential sections indicate that this band forms a belt around the embryo. The width of this belt indicates that the transcripts are preferentially located within two or three segments. At the early gastrula stage, the silver grains are located over the cephalic furrow and both anterior and posterior to it. The localization corresponds to the posterior cephalic and anterior thoracic segments in the fate map. The signal intensity is considerably weaker than that for *Scr*.

Discussion

The structure of ANT-C was dissected further by extending the 'chromosomal walk' from *Antp* to *Dfd* and mapping the transcripts in this region. A small 0.5-kb transcript encoded by gene X was mapped between *Antp* and *ftz*. *Df(3R)9A99* deletes most of the genes of the proximal ANT-C, including *pb, Dfd, Scr* and *ftz*, but leaves gene X and *Antp* intact. Therefore, gene X cannot correspond to *Scr*. We also cloned the chromosomal DNA from the *Nasobemia (Ns)* mutant which maps to the left of $Antp^{73b}$ (Tottoli, 1977) and found no major change in this region of the DNA which consists of single copy sequences (data not shown). At present we cannot assign any particular mutant to this region.

Genetic complementation tests indicate that *Df(3R)4Scb* is Scr^-, ftz^-, $Antp^-$. We have mapped its distal breakpoint within *Antp* and its proximal breakpoint in the intron of the 3.9-kb RNA encoding region. In the area of overlap between *Df(3R)9A99* and *Df(3R)4Scb*, which genetically includes *Scr* and *ftz*, there are only two transcripts, the 1.9-kb transcript of *ftz* (Kuroiwa *et al.*, 1984; Weiner *et al.*, 1984) and the 3.9-kb transcript. The *ftz* gene has been identified unequivocally by P factor-mediated germ line transformation (Hiromi *et al.*, in preparation). Therefore, we assign the 3.9-kb transcript to the *Scr* locus. *Scr* is a homeotic gene and we have found both a homeobox and an M repeat in this gene. Northern analysis, S1 mapping and cDNA clones have identified at least two exons adding up to a total length of 3.6 kb which is slightly smaller than the mRNA size (3.9 kb) as determined by Northern analysis. An exon smaller than 100 bp may have been missed by the methods that we have used. The dominant Msc^1 inversion which also has an *Scr* phenotype, as well as extra sex combs on the second and third legs, has its proximal breakpoint outside the 3.9-kb RNA encoding region. This inversion may either exert a position effect on *Scr* or the proximal breakpoint may be located in the controlling region of this gene which would explain the *Scr* phenotype. Hazelrigg and Kaufman (1983) and Scott *et al.* (1983) have found three chromosomal rearrangements with breakpoints to the right of *ftz* which are also associated with an *Scr* phenotype. There are several possibilities to explain these observations. One is that gene X may have an *Scr* like function which is supported by our *in situ* hybridization data indicating that the X transcripts also accumulate in the anterior thoracic and posterior head segments. Alternatively, gene X, which is expressed early in embryogenesis, may exert a positive regulatory function on *Scr*. In this case, inactivation of X would result in an *Scr* phenotype. Further work is needed to clarify this point.

In the *Scr* region two more transcribed segments, Y2 and Y3, were found both encoding a small RNA of ~380 nucleotides. This RNA is >10 times more abundant than the 3.9-kb transcript. Y2 is located within the *Scr* intron. This is similar to the case of the troponin T gene, where a small, abundant RNA is encoded within an intron of the troponin transcription unit (Medford

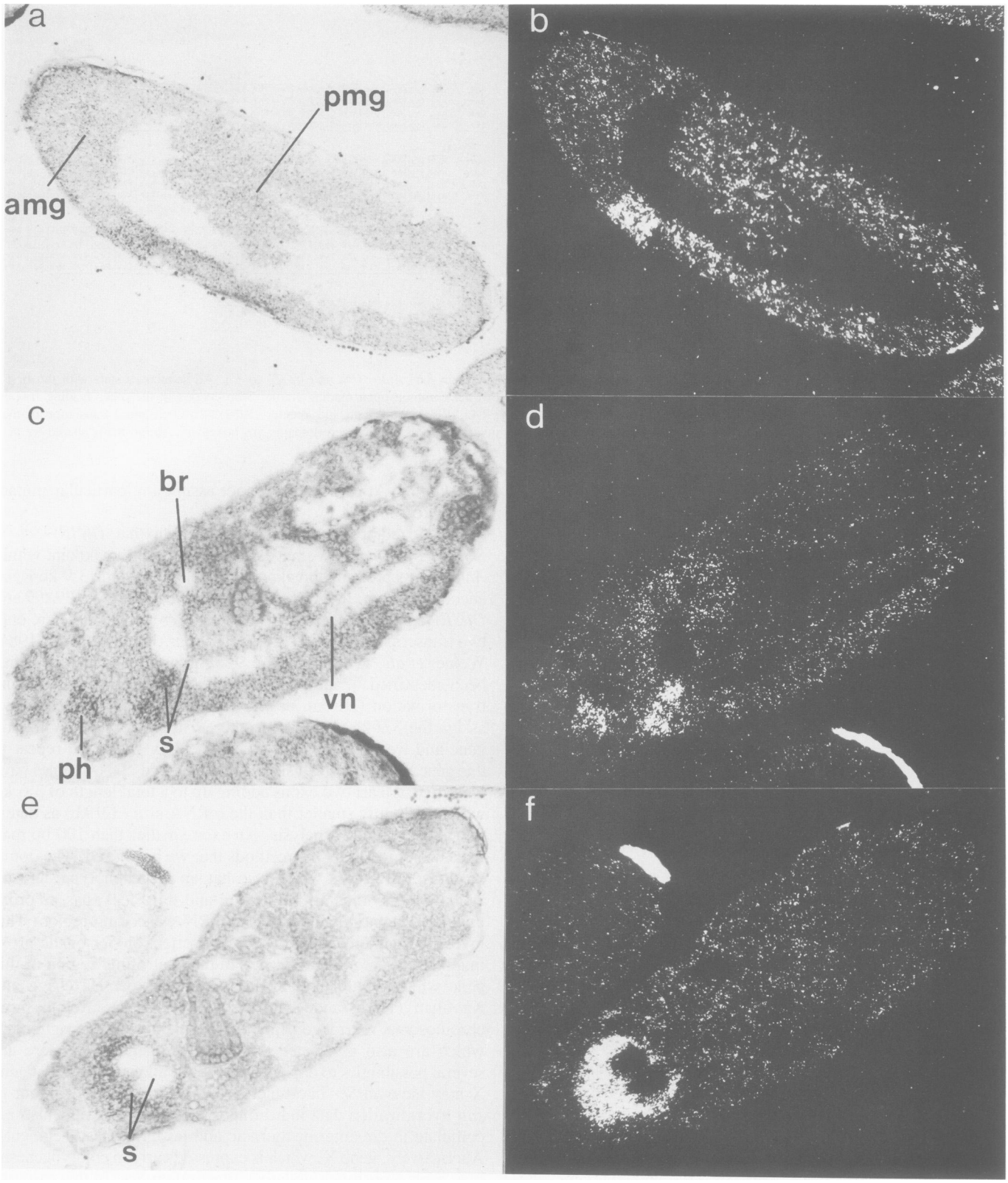

Fig. 8. Localization of Scr^+ transcripts in tissue sections of developing embryos. For each section a brightfield and the corresponding darkfield micrograph are shown. The anterior end of the embryo is on the left, dorsal is on the top. The frozen sections were hybridized with the pcY20 Scr^+ cDNA probe and autoradiographed for 10 days. (**a** and **b**) Sagittal section through an embryo during germ band extension (~5 h). (**c** and **d**) Sagittal section through an 18 h embyro. (**e** and **f**) Frontal section through an 18 h embryo. Abbreviations: amg = anterior midgut invagination, br = brain, ph = pharynx region, pmg = posterior midgut invagination, vn = ventral nerve cord, S = subesophageal ganglion.

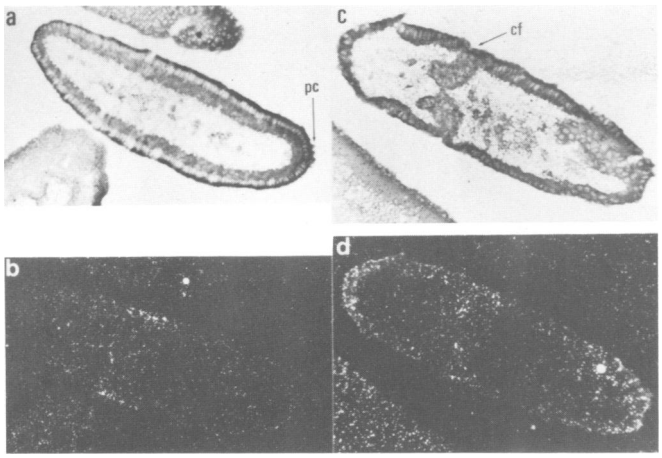

Fig. 9. Localization of gene X transcripts in tissue sections of early embryos. The 1.3-kb chromosomal DNA fragment (+168.7 to +170.0) was used as a probe. The autoradiographs were exposed for 30 days. The brightfield and the corresponding darkfield micrographs are shown. The anterior end is on the left. (**a** and **b**) Sagittal section through an embryo at the cellular blastoderm stage (3 h). (**c** and **d**) Frontal section through an embryo at the beginning of gastrulation (3½ h). cf = cephalic furrow, pc = pole cells.

et al., 1984) and also reminiscent of the rat brain ID sequences which are found in introns and code for small, abundant RNAs (Sutcliff *et al.*, 1984). Finally, the transcription unit z, which encodes a 1.2-kb transcript, has been mapped between *Dfd* and *Scr*. Scott *et al.* (1984) have reported a repetitive element within this area in the Canton-S stock, which may correspond to z.

The analysis of the homeoboxes and M repeats in the proximal ANT-C has confirmed the strong correlation between these elements and homeotic genes or other genes involved in the spatial organization of the embryo. *Antp, ftz, Scr* and *Dfd* carry both a homeobox in the most distal 3′ exon and an M repeat. The *Scr* homeobox sequence is most closely related to *Antp* which represents the prototype sequence (W.Gehring, in preparation). The homology between the homeoboxes in these two genes is 82% at the DNA, and 92% at the protein level. The M repeat appears to code for stretches of polyamino acids like polyglutamine or polyasparagine, although in some cases it is not yet clear which of the open reading frames are used for translation. The M repeat (also designated *opa*) has also been found in *Notch* and several other genes that do not have a homeobox (Wharton *et al.*, 1985). It is present in several hundred copies per genome. At the present time the significance of these polyamino acids stretches is not known. In *Ubx* a stretch of polyglycine has been found which is thought to form a hinge region within the protein (Beachy *et al.*, 1985).

Our *in situ* hybridization experiments indicate that the Scr^+ transcripts accumulate preferentially in the suboesophageal ganglion and in the anterior part of the prothoracic ganglion. Fewer transcripts are found in the other parts of the nervous system including the brain and the posterior ventral nerve cord. This localization clearly corresponds to the Scr^- mutant phenotype which is due to the loss of gene function, since in Scr^- mutant embryos the labial and prothoracic segments are homeotically transformed.

Materials and methods

Isolation of nucleic acids

Charon phage DNA and plasmid DNA was isolated as described elsewhere (Garber *et al.*, 1983; Kuroiwa *et al.*, 1984). Embryonic RNA was extracted by the hot phenol method (Lepesant *et al.*, 1978). Larval and pupal RNA was extracted by the guanidium isothiocyanate-hot phenol method (Maniatis *et al.*, 1982) after pulverizing frozen larvae or pupae with a pestle and mortar in liquid nitrogen. Poly(A)$^+$ RNA was selected by oligo(dT)-cellulose chromatography (Collaborative Research Inc. Type T3).

Cloning

Phages containing wild-type *D. melanogaster* (Canton S) DNA were isolated from the library of Maniatis *et al.* (1978). Restriction fragments were subcloned into pAT153 (Twigg and Sherratt, 1982), or M13mp8 or M13mp9 replicative form DNA that had been cleaved with an appropriate restriction enzyme. Recombinant pAT153 was used to transform competent *E. coli* HB101, EM1035 or DH-1 cells prepared by the procedure of Morrison (1979). M13 recombinants were used to transform JM101 host cells. The wild-type embryonic cDNA library was constructed by Kauvar *et al.* and screened with nick-translated chromosomal DNA (a mixture of +204.0 to +209.1 and +222.2 to +230.0) subcloned into pAT153.

Electrophoresis of nucleic acids and transfer to filters

40 μg poly(A)$^+$ RNA from each developmental stage was fractionated by electrophoresis on a 0.8% agarose gel containing 2.2 M formaldehyde (Lehrach *et al.*, 1977). After the electrophoresis, the fractionated RNA was directly transferred by nitrocellulose filters (Schleicher and Schüll BA85) overnight in 20 × SSC by the method of Southern (1977). Hybridization and washing conditions were described previously (Kuroiwa *et al.*, 1984).

Transcriptional mapping with nuclease S1

The conditions of hybridization and S1 treatment were described previously (Kuroiwa *et al.*, 1984). After precipitation of the S1 digest with isopropanol, fragments were fractionated by alkaline agarose gel electrophoresis. After neutralizing the gel with a solution containing 0.5 M Tris-HCl pH 7.5 and 1.5 M NaCl, fragments were transferred to nitrocellulose filters by the Southern method. The protected fragments were detected by hybridizing with a nick-translated probe as described previously (Kuroiwa *et al.*, 1984).

Determination of the direction of transcription

To determine the transcription direction, the *Nco*I-*Eco*RI 780-bp fragment from the second *Scr* exon or the *Sma*I-*Hind*III 1120-bp fragment were cloned in M13mp8 and mp9 separately. The single-stranded phage DNA was isolated and hybridized with 20 μg of poly(A)$^+$ RNA under the S1 mapping condition. After incubation, the hybridization mixture was treated with S1 nuclease and the protected fragment was separated on an alkaline agarose gel. After transferring to a nitrocellulose filter, the protected fragment was detected by hybridization with a nick-translated DNA probe.

In situ hybridization to sections

A detailed description of the *in situ* hybridization method used for detecting specific RNAs in tissue sections has been published previously (Hafen *et al.*, 1983). The hybridization probes used in this analysis had a specific activity of $\sim 5-7 \times 10^8$ d.p.m./μg.

Acknowledgements

We are grateful to Tom Maniatis for the *Drosophila* DNA lambda library, Lawrence Kauvar, Barry Drees, Stephan Poole and Thomas Kornberg for aliquots of their excellent cDNA library and Gerd Jürgens for mutant stocks. Our progress benefitted greatly from discussions with Iain Mattaj, Christiane Nüsslein-Volhard and Gerd Jürgens. This work was supported by the Swiss National Science Foundation and the Kanton Basel-Stadt.

References

Anderson,K.V. and Nüsslein-Volhard,C. (1984) *Nature*, **311**, 223-227.
Beachy,P.A., Helfand,S.L. and Hogness,D.S. (1985) *Nature*, **313**, 545-551.
Garber,R.L., Kuroiwa,A. and Gehring,W.J. (1983) *EMBO J.*, **2**, 2027-2036.
Gehring,W. (1973) in Ruddle,F. (ed.), *Genetic Mechanisms of Development*, Academic Press, NY, pp. 103-128.
Gehring,W.J. (1985) *Cell*, **40**, 3-5.
Hafen,E., Levine,M., Garber,R.L. and Gehring,W.J. (1983) *EMBO J.*, **2**, 617-623.
Hartenstein,V. and Campos-Ortega,J.A. (1985) *Wilhelm Roux's Arch. Dev. Biol.*, **194**, 181-195.
Hazelrigg,T. and Kaufman,T.C. (1983) *Genetics*, **105**, 581-600.
Jürgens,G., Wieschaus,E., Nüsslein-Volhard,C. and Kluding,H. (1984) *Wilhelm Roux's Arch. Dev. Biol.*, **193**, 283-295.
Kaufman,T., Lewis,R. and Wakimoto,B. (1980) *Genetics*, **94**, 115-133.
Kuroiwa,A., Hafen,E. and Gehring,W.J (1984) *Cell*, **37**, 825-831.
Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry (Wash.)*, **16**, 4743-4751.

Lepesant,J.-A., Kejzlarova-Lepesant,J. and Garen,A. (1978) *Proc. Natl. Acad. Sci USA*, **75**, 5570-5574.
Lewis,E.B. (1978) *Nature*, **276**, 565-570.
Lewis,R.A., Kaufman,T.C., Denell,R.E. and Tallerico,P. (1980a) *Genetics*, **95**, 367-381.
Lewis,R.A., Wakimoto,B.T., Denell,R.E. and Kaufman,T.C. (1980b) *Genetics*, **95**, 383-397.
Lindsley,D.L. and Grell,E.M. (1968) Carnegie Institute of Washington Publ. No. 627.
Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,G.K. and Efstratiadis,A. (1978) *Cell*, **15**, 687-701.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
McGinnis,W., Levine,M.S., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984a) *Nature*, **308**, 428-433.
McGinnis,W., Garber,R.L., Wirz,J., Kuroiwa,A. and Gehring,W.J. (1984b) *Cell*, **37**, 403-408.
Medford,R.M., Nguyen,H.T., Destree,A.T., Summers,E. and Nadel-Girard,B. (1984) *Cell*, **38**, 409-421.
Morrison,D.A. (1979) *Methods Enzymol.*, **68**, 326-331.
Nüsslein-Volhard,C. (1979) in Konigsberg,I. and Subteiny,S. (eds.), *Determinants of Spatial Organization*, Academic Press, NY, pp. 185-211.
Nüsslein-Volhard,C. and Wieschaus,E. (1980) *Nature*, **287**, 795-801.
Nüsslein-Volhard,C., Wieschaus,E. and Kluding,H. (1984) *Wilhelm Roux's Arch. Dev. Biol.*, **193**, 267-282.
Ouweneel,W.J. (1976) *Adv. Genet*, **18**, 179-248.
Sanchez-Herrero,E., Vernos,I., Rand,M. and Morata,G. (1985) *Nature*, **313**, 108-113.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Scott,M.P. and Weiner,A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4115-4119.
Scott,M.P., Hazelrigg,T.I., Polisky,B.A., Pirrotta,V., Scalenghe,F. and Kaufman,T.C. (1983) *Cell*, **35**, 763-776.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Struhl,G. (1981) *Nature*, **292**, 635-638.
Struhl,G. (1983) *J. Embryol. Exp. Morphol.*, **76**, 297-331.
Sutcliff,J.G., Milner,R.J., Gottesfeld,J.M. and Reynolds,W. (1984) *Science (Wash.)*, **225**, 1308-1315.
Tottoli,M. (1977) Diploma Thesis, University of Basel, Switzerland
Twigg,A.J. and Sherratt,D. (1980) *Nature*, **283**, 216-218.
Wakimoto,B.T. and Kaufman,T.C. (1981) *Dev. Biol.*, **81**, 51-64.
Wakimoto,B.T., Turner,F.R. and Kaufman,T.C. (1984) *Dev. Biol.*, **102**, 147-172.
Weiner,A.J., Scott,M.P. and Kaufman,T.C. (1984) *Cell*, **37**, 843-851.
Wharton,K.A., Yedvobnick,B., Finnerty,V.G. and Artavanis-Tsakanos,S. (1985) *Cell*, **40**, 55-62.
Wieschaus,E., Nüsslein-Volhard,C. and Jürgens,G. (1984) *Wilhelm Roux's Arch. Dev. Biol.*, **193**, 296-307.

Received on 30 September 1985; revised on 22 October 1985