The EMBO Journal vol.4 no.11 pp.2961–2969, 1985

Isolation of *caudal*, a *Drosophila* homeo box-containing gene with maternal expression, whose transcripts form a concentration gradient at the pre-blastoderm stage

Marek Mlodzik, Anders Fjose[1] and Walter J.Gehring

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

[1]Present address: Laboratory of Biotechnology, University of Bergen, P.Box 3152 Arstad, 5001 Bergen, Norway

Communicated by W.J.Gehring

We report the isolation and characterization of *caudal (cad)*, a previously unknown *Drosophila* homeo box-containing gene from the 38E region on the left arm of the second chromosome. This homeo box has diverged from the prototype sequence in *Antennapedia*, but contains subregions which are highly homologous. By Northern analysis and *in situ* hybridization experiments two transcripts of ~2.4 kb were found to accumulate in nurse cells and in the oocyte during oogenesis. These transcripts generate a transient concentration gradient along the antero-posterior axis at the syncytial blastoderm stage. At the cellular blastoderm stage transcripts accumulate in a single band from 13–19% egg length at the posterior end. One zygotic transcript of 2.6 kb is detected. At later stages this transcript is localized in ectodermally and endodermally derived tissues such as the proctodeum, the Malpighian tubules and the posterior midgut. The 2.6-kb transcript is detectable until the onset of metamorphosis.

Key words: homeo box/homeotic genes/maternal expression/RNA localization/endoderm

Introduction

The molecular processes generating a highly organized complex organism from a single cell, the egg, are under genetic control. Two types of genes affecting embryonic development may be distinguished: genes whose activity is required during oogenesis (maternal effect genes) and genes whose activity is required after fertilization when the genome of both parents is expressed (zygotic genes). The phenotypes of the few *Drosophila* maternal effect mutations which have been studied in detail suggest that the maternal contribution is of a rather general nature, resulting in the definition of spatial coordinates in the developing embryo (Nüsslein-Volhard, 1979). Maternal effect genes can affect either the antero-posterior or the dorso-ventral fate of the embryonic cells. Dramatic alterations in the embryonic pattern along the antero-posterior axis are observed for example in the mutants *bicaudal* and *dicephalic* (Bull, 1966; Lohs-Schardin, 1982) which lead to the formation of double abdomina and double heads, respectively. Other mutants showing a maternal effect, such as *extra sex combs (esc)* (Struhl, 1981, 1983) and *Polycomb (Pc)* (Denell, 1978; Duncan and Lewis, 1982) suggest that these genes may act as negative regulators of homeotic selector genes of both the *Bithorax (BX-C)* and the *Antennapedia* complex *(ANT-C)*. Additional genes belonging to this homeotic regulator class have been identified, e.g., *Polycomblike (Pcl)* (Duncan, 1982) and *super sex combs* (*sxc*, Ingham, 1984). Pc^+ and esc^+, which assure the proper spatial expression of the homeotic selector genes, are acting during oogenesis in the maternal germ line, but the absence of their products can be largely compensated by the wild-type allele introduced by the sperm during fertilization (Duncan and Lewis, 1982; Struhl, 1983). These regulator genes seem to be involved in the interpretation or processing of the positional information laid down during oogenesis. Some maternal effect genes that seem to be directly involved in the establishment of positional information have been identified, but very few have been isolated (*dorsal*, Steward *et al.*, 1984). The zygotic genes, important during early embryogenesis, can be subdivided into two major groups in *Drosophila*: the segmentation genes, which are responsible for the formation of the correct number and polarity of the segments, and the homeotic selector genes, which give each segment its own identity. The latter group acts after the segmentation genes but still as early as the cellular blastoderm, and from this time on the cells are determined (Chan and Gehring, 1971) and have specific segmental identities (Wieschaus and Gehring, 1976; Lawrence and Morata, 1977).

The discovery of the homeo box, a conserved DNA element shared by several homeotic genes (McGinnis *et al.*, 1984a, 1984b; Scott and Weiner, 1984) opens up new possibilities for the isolation of developmentally important genes in *Drosophila* (McGinnis *et al.*, 1984a, 1984b; Fjose *et al.*, 1985) as well as in higher metazoans including vertebrates and man (McGinnis *et al.*, 1984b, 1984c; Carrasco *et al.*, 1984; Levine *et al.*, 1984). In *D. melanogaster* the homeo box has been found so far only in zygotic genes that belong to the homeotic selector group (McGinnis *et al.*, 1984a, 1984b; Fjose *et al.*, 1985) or in the segmentation gene *fushi tarazu (ftz)* (Kuroiwa *et al.*, 1984; Hafen *et al.*, 1984; Laughon and Scott, 1984) and *engrailed (en)* (Fjose *et al.*, 1985; Poole *et al.*, 1985). In those maternal effect genes that have been cloned recently, no homeo box sequences have been found (Steward *et al.*, 1984; Frei *et al.*, 1985). However, in *Xenopus laevis* a maternally expressed gene sharing homeo box homology has been isolated (Müller *et al.*, 1984).

Using the homeo box homology, we have isolated a gene that shows maternal as well as zygotic expression during *Drosophila* development. *In situ* localization of its zygotic transcripts at the cellular blastoderm stage reveals that it is expressed only in tissue derived from the posterior 20% of the embryo. During later embryogenesis transcripts are localized in structures like the hindgut, the Malpighian tubules and the posterior midgut. Therefore, we name this gene *caudal (cad)*. *In situ* hybridization to the maternal transcripts of *cad* at the pre-blastoderm stage reveals that the transcript distribution follows a concentration gradient.

Results

Isolation of clones

The *D. melanogaster* genomic library of Maniatis *et al.* (1978) was screened under reduced hybridization stringency as previously described by McGinnis *et al.* (1984a). Using a small 450-bp homeo box-containing fragment from the *Ultrabithorax (Ubx)*

© IRL Press Limited, Oxford, England.

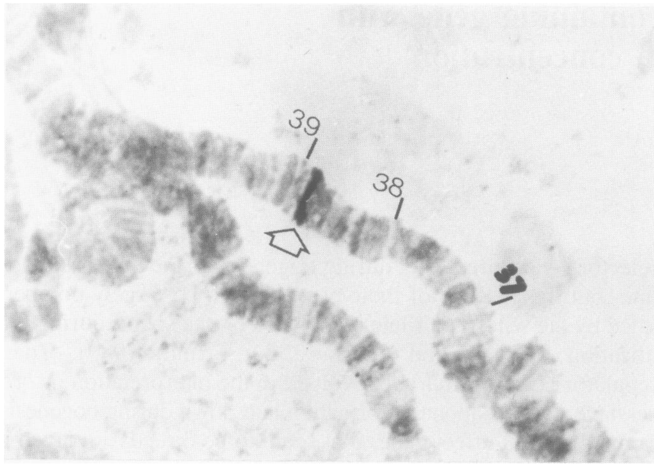

Fig. 1. *In situ* hybridization of biotinylated F33 DNA to salivary gland polytene chromosomes. A hybridization signal (arrow) is observed on the left arm of the second chromosome at position 38E.

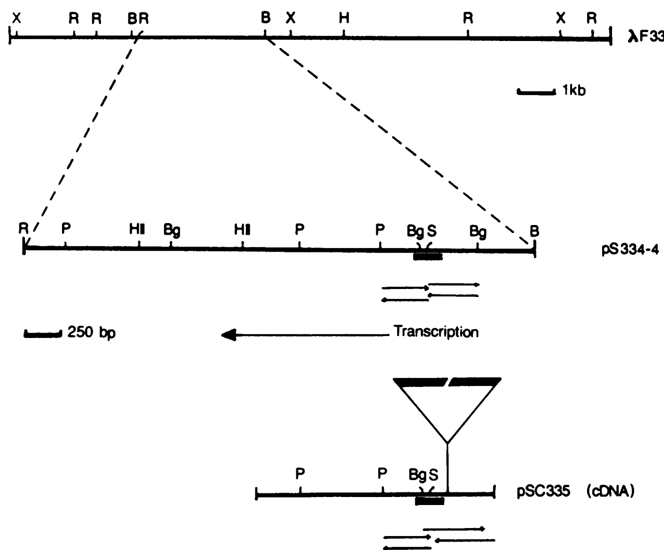

Fig. 2. Restriction maps of clones originating from 38E. Black boxes identify the regions containing homeo box homology. Restriction enzymes: B=*Bam*HI, Bg=*Bgl*II, H=*Hind*III, HII=*Hinc*II, p=*Pst*I, R=*Eco*RI, S=*Sac*I, X=*Xba*I. The direction of transcription is indicated. Arrows below the map represent sequence determination strategies. The position of the intron is indicated by a triangle.

gene as a probe, clones mapping to several different cytological positions were obtained (Fjose *et al.*, 1985). Among these, the clone λF33 was found to originate from 38E on the left arm of the second chromosome (see Figure 1). Mutants which affect pattern formation were so far not known to map in this chromosomal region.

Although F33 was isolated by screening with a *Ubx* homeo box, it also cross-hybridizes to the same extent with homeo box probes of $Antp^+$ and ftz^+. The homeo box homology in the clone was localized by hybridizing Southern blots of restricted DNA with Ubx^+ and $Antp^+$ homeo box probes under conditions of reduced stringency. Both probes hybridized to a 3.6-kb *Eco*RI/*Bam*HI fragment as indicated in Figure 2. A subclone of this fragment in the pSP64 vector (Green *et al.*, 1983), pS334-4, was used for further defining the region of cross-homology by Southern blot analyses. A 1-kb *Pst*I/*Bam*HI fragment was shown to contain homeo box homology (Figure 2). The 38E homeo box was also characterized by probing whole genome Southern blots under conditions of reduced hybridization stringency. The number and the pattern of bands detected with the 38E probe is similar to those detected with $Antp^+$ and Ubx^+ homeo boxes (McGinnis *et al.*, 1984a, 1984b), however the intensity of the cross-hybridizing bands with 38E is lower than with $Antp^+$ and Ubx^+ (data not shown).

The subcloned fragment pS334-4 was further used as a probe to isolate cDNA clones from a cDNA pool prepared from poly(A)$^+$ RNA from 3 – 12 h embryos by L.Kauvar. One cDNA clone (pSC335) of 1.65 kb containing the homeo box was isolated (see Figure 2) indicating that these sequences from the 38E region are transcribed during embryogenesis. Furthermore several smaller cDNA clones of 0.54 – 0.8 kb were obtained from a 0 – 3 h cDNA pool. All of them are included in the 1.65-kb cDNA as judged from restriction mapping and Southern blotting data (not shown).

DNA sequences in the homeo box region

To analyse the fine structure of the 38E region sharing homeo box homology, DNA from genomic and cDNA clones was sequenced. Segments of ~700 bp from pS334-4 and the cDNA clone pSC335 were sequenced according to the strategies shown in Figure 2. Comparison of the genomic and cDNA sequences revealed that an intron is located just 15 bp upstream from the 5′ end of the theoretical start of the homeo box (Figure 3). This feature is shared with several other homeo box-containing genes (Kuroiwa *et al.*, 1984 and in preparation; Scott and Weiner, 1984). When the 700-bp sequence from the cDNA (only 591 bp are shown in Figure 3) is read in-frame with the homeo box, no stop codons are found further downstream within the sequenced region. One interesting feature derived from the cDNA sequence is a long AAC/AAT stretch 30 codons upstream of the homeo box (see Figure 3), which is interrupted twice by different codons. This sequence could code for a poly-asparagine stretch of 20 amino acids. Similar monotonous stretches coding for polyglutamine have been found in other homeotic genes and in genes involved in neurogenesis and designated as M- or opa-repeats (McGinnis *et al.*, 1984a, and personal communication; Wharton *et al.*, 1985; Poole *et al.*, 1985).

The homeo box of 38E is less homologous to *Antp*, *ftz* and *Ubx* than these three are to one another (Figure 4). At the nucleotide level the cross-homology to *Antp*, *ftz* and *Ubx* is 62%, 60% and 60%, respectively, which is considerably lower than the 75 – 79% found for the three other *Drosophila* homeo boxes (McGinnis *et al.*, 1984b). It is, however, less diverged than the *engrailed (en)* homeo box (Fjose *et al.*, 1985; Poole *et al.*, 1985). The level of cross-homology is not evenly distributed throughout the 180-bp sequence. Both, the 5′ and the 3′ end of the homeo box show relatively little homology to the prototype sequence found in *Antp* (see Figure 4). In contrast, the major part (position 22 – 165) of the 38E homeo box shares a considerable amount of homology (~70%) with both *Antp* and *Ubx*. Comparing the region of the homeo box that is homologous to the yeast mating type genes the homology is even more striking: from nucleotide 118 to 165 the 38E homeo box matches *Antp* at 38 of 48 positions (79% homology) and *Ubx* at 39 of 48 (81%). In this region the amino acids homology is 81% with both *Antp* and *Ubx*. A similar distribution of cross-homology with *Antp* and *Ubx* has also been observed within the *en* homeo box (Fjose *et al.*, 1985; Poole *et al.*, 1985).

Unlike most other homeo domains in *Drosophila* (McGinnis

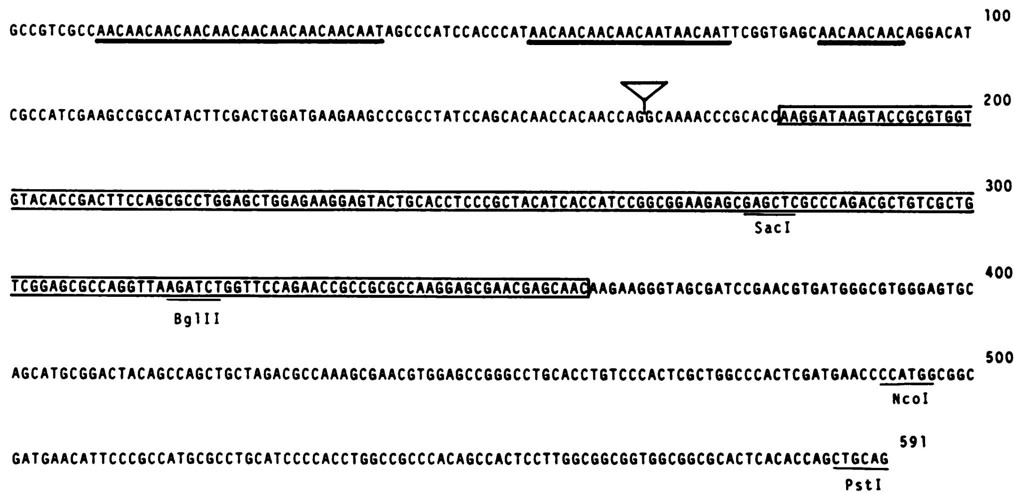

Fig. 3. DNA sequence in the homeo box homology region. cDNA sequence (5′-3′) of a 591-bp region including the homeo box (boxed) of 38E. The sequence of the AAC/AAT stretch is underlined. The position of the intron is indicated by a triangle.

et al., 1984b; Laughon and Scott, 1984) and vertebrates (Carrasco *et al.*, 1984; McGinnis *et al.*, 1984c; Levine *et al.*, 1984; Müller *et al.*, 1984) the 38E homeo domain shares less homology, 53%, 58% and 52% (Figure 5) with $Antp^+$, ftz^+ and Ubx^+, respectively, at the protein than at the DNA level (~60%; Figure 4). Also in this respect it is similar to en^+ (Fjose *et al.*, 1985; Poole *et al.*, 1985). On the basis of these features it seems as if the *engrailed* and the 38E homeo boxes have been subjected to some of the same constraints when they diverged. In spite of this, their cross-homologies at the DNA and protein level are only 50% and 38%, respectively. The only region in which their homeo boxes have not diverged extensively is the 48-nucleotide segment at position 118–165, where both protein and DNA homologies are ~75%. The most highly conserved region of the homeo domain corresponds to amino acids 42–50, which are thought to correspond to the α-helix, which is making the contact to the DNA in the major groove in various regulatory proteins of procaryotic genes with sequence specificity (Pabo and Sauer, 1984; Shepherd *et al.*, 1984; Laughon and Scott, 1984).

Transcription during development

The cDNA clone pSC335 was used to determine the temporal expression pattern of the 38E gene by Northern blotting analysis. Using single-stranded RNA probes, generated by *in vitro* transcription with the SP6 system (Green *et al.*, 1983; Melton *et al.*, 1984), we also examined the direction of transcription. The isolated cDNA of 1.65 kb was inserted into pSP65 in both orientations and transcribed in both directions separately, followed by hybridizations to Northern blots. When pSC335 was transcribed from right to left in Figure 2 no complementary RNA was detected (data not shown). RNA transcribed from left to right, hybridized to a developmental Northern blot, is shown in Figure 6, indicating that the 38E region is transcribed *in vivo* in the orientation indicated in Figure 2, as expected from the DNA sequence of the homeo box.

Electrophoretically fractionated poly(A)$^+$ RNA from egg follicles, different embryonic stages, first, second and late third instar larvae, early pupae and adult males and females was blotted on nitrocellulose filters and subsequently hybridized to pSC335 (see above). As shown in Figure 6 pSC335 strongly hybridizes to poly(A)$^+$ RNA from egg follicles and 0–2 h embryos. The observed transcripts form a double band at 2.4 kb

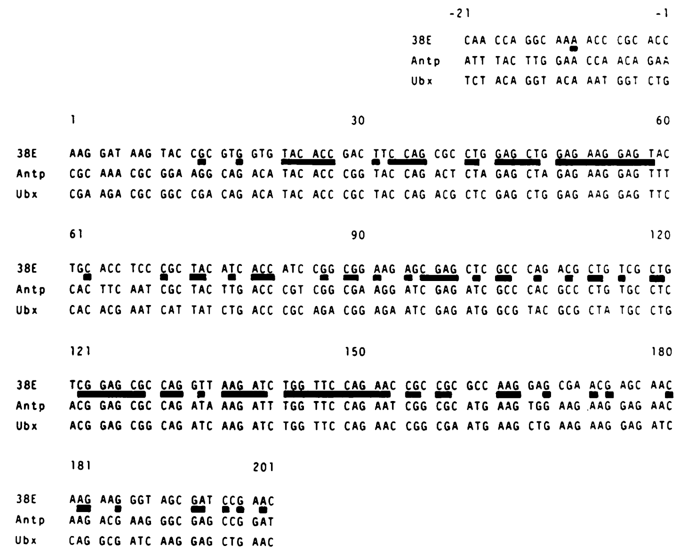

Fig. 4. DNA sequence of the 38E homeo box (5′-3′) aligned with homologous regions from *Antp* and *Ubx*. The sequence is shown from −21 (21 bp upstream of the homeo box) to +201 (21 bp downstream of the homeo box). Codons of the common open reading frame are aligned (beginning with nucleotide 1). The black bar underlines the nucleotide homology between *Antp*, *Ubx* and 38E.

(by comparison with single-stranded DNA fragments used as a standard). In 2–4 h embryos, the abundance of these two transcripts decreases and a new transcript 2.6 kb in length, is observed. In older embryos, 4 h or more, the maternally derived transcripts are no longer detectable, but the zygotic 2.6-kb transcript shows approximately the same abundance during later stages of development up to early pupae. The use of actin as an internal standard (see Materials and methods) showed that the differences in the intensities of the 2.6-kb bands are artefactual, due to variations in the amounts of poly(A)$^+$ RNA loaded in each lane or to degradation of the RNA (smear in 12–20 h embryo lane). In adult females the two transcripts at 2.4 kb are abundant. In males only a weak signal around 2.4 kb is detected. This is most likely due to a slight contamination with female RNA, because in another separate male RNA extraction no signal was observed.

	-7						-1
38E	Gln	Pro	Gly	Lys	Thr	Arg	Thr
Antp	Ile	Tyr	Leu	Glu	Pro	Thr	Glu
ftz	Met	Leu	Thr	Asp	Cys	Lys	Asp
yeast a1	His	His	Ile	Lys	Lys	Glu	Lys

	1									10										20
38E	Lys	Asp	Lys	Tyr	Arg	Val	Val	Tyr	Thr	Asp	Phe	Gln	Arg	Leu	Glu	Leu	Glu	Lys	Glu	Tyr
Antp	Arg	Lys	Arg	Gly	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln	Thr	Leu	Glu	Leu	Glu	Lys	Glu	Phe
ftz	Ser	Lys	Arg	Thr	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln	Thr	Leu	Glu	Leu	Glu	Lys	Glu	Phe
yeast a1	Ser	Pro	Lys	Gly	Lys	Ser	Ser	Ile	Ser	Pro	Gln	Ala	Arg	Ala	Phe	Leu	Glu	Gln	Val	Phe

	21									30										40
38E	Cys	Thr	Ser	Arg	Tyr	Ile	Thr	Ile	Arg	Arg	Lys	Ser	Glu	Leu	Ala	Gln	Thr	Leu	Ser	Leu
Antp	His	Phe	Asn	Arg	Tyr	Leu	Thr	Arg	Arg	Arg	Arg	Ile	Glu	Ile	Ala	His	Ala	Leu	Cys	Leu
ftz	His	Phe	Asn	Arg	Tyr	Ile	Thr	Arg	Arg	Arg	Arg	Ile	Asp	Ile	Ala	Asn	Ala	Leu	Ser	Leu
yeast a1	Arg	Arg	Lys	Gln	Ser	Leu	Asn	Ser	Lys	Glu	Lys	Glu	Glu	Val	Ala	Lys	Lys	Cys	Gly	Ile

	41									50										60
38E	Ser	Glu	Arg	Gln	Val	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Ala	Lys	Glu	Arg	Thr	Ser	Asn
Antp	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Trp	Lys	Lys	Glu	Asn
ftz	Ser	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Ser	Lys	Lys	Asp	Arg
yeast a1 spliced	Thr	Pro	Leu	Gln	Val	Arg	Val	Trp	Phe	Ile	Asn	Lys	Arg	Met	Arg	Ser	Lys	Stop		

	61						67
38E	Lys	Lys	Gly	Ser	Asp	Pro	Asn
Antp	Lys	Thr	Lys	Gly	Glu	Pro	Asp
ftz	Thr	Leu	Asp	Ser	Ser	Pro	Glu

Fig. 5. Conceptual translation of the common homologous open reading frame in the 38E homeo box. The 38E sequence is compared with *Antp*, *ftz* and the corresponding part of the *a1* mating-type protein. A black bar underlines the amino acid residues homologous to both *Antp* and *ftz*. A broken bar underlines residues that show conservative changes.

In situ localization of transcripts

The homeotic selector genes of *Drosophila* seem to be involved in the spatial organization of the embryo and the specification of different anatomical segments (Lewis, 1978; Kaufman *et al.*, 1980). Recently, *in situ* hybridization to tissue sections has shown that homeotic transcripts are restricted spatially to regions where the functions of the respective genes are presumably required for proper development (Hafen *et al.*, 1983, 1984; Levine *et al.*, 1983; Akam, 1983; McGinnis *et al.*, 1984a; Fjose *et al.*, 1985; Kornberg *et al.*, 1985). Here we describe a homeo box-containing gene which has not been identified so far by mutation. Localization of its transcripts in tissue sections using pSC335 cDNA as hybridization probe is therefore a direct molecular approach to obtain information about the site(s) of action of the 38E homeo box gene.

In contrast to other homeo box genes in *Drosophila*, the 38E gene is expressed during oogenesis as well as during the developmental stages after blastoderm formation (Figure 6). During early oogenesis transcripts are localized in the cytoplasm of nurse cells and the oocyte (stage 7–8 oocytes, Figure 7a and b). Hybridization signals are only observed within the egg chamber (germline) but not over the surrounding follicle cells and connective tissue (somatic cells). The distribution and intensity of the hybridization signals in the oocyte at the end of oogenesis remain the same during the first hour of embryogenesis. A young embryo (~1 h after fertilization) is shown in Figure 7c and d, where hybridization signals are fairly evenly distributed throughout the section. The transcripts appear to be associated with the nuclei and the surrounding cytoplasm but excluded from the yolk. Following the 8th nuclear division when the nuclei start to migrate towards the periphery, the transcripts disappear in the

Fig. 6. Transcription from 38E during *Drosophila* development. Poly(A)$^+$ RNA isolated from the indicated stages was subjected to Northern blot analysis and probed with pSC335. The smear in the 12–20 h embryo lane is due to degradation of RNA. 2 μg of RNA were loaded from egg follicles and 0–2 h embryos, all other lanes contain 4 μg of RNA. foll = egg follicles, embryonic stages in hours (h), 1L = 1st instar larvae, 2L = 2nd instar larvae, 3L = late 3rd instar larvae, eP = early pupae, m = males, f = females.

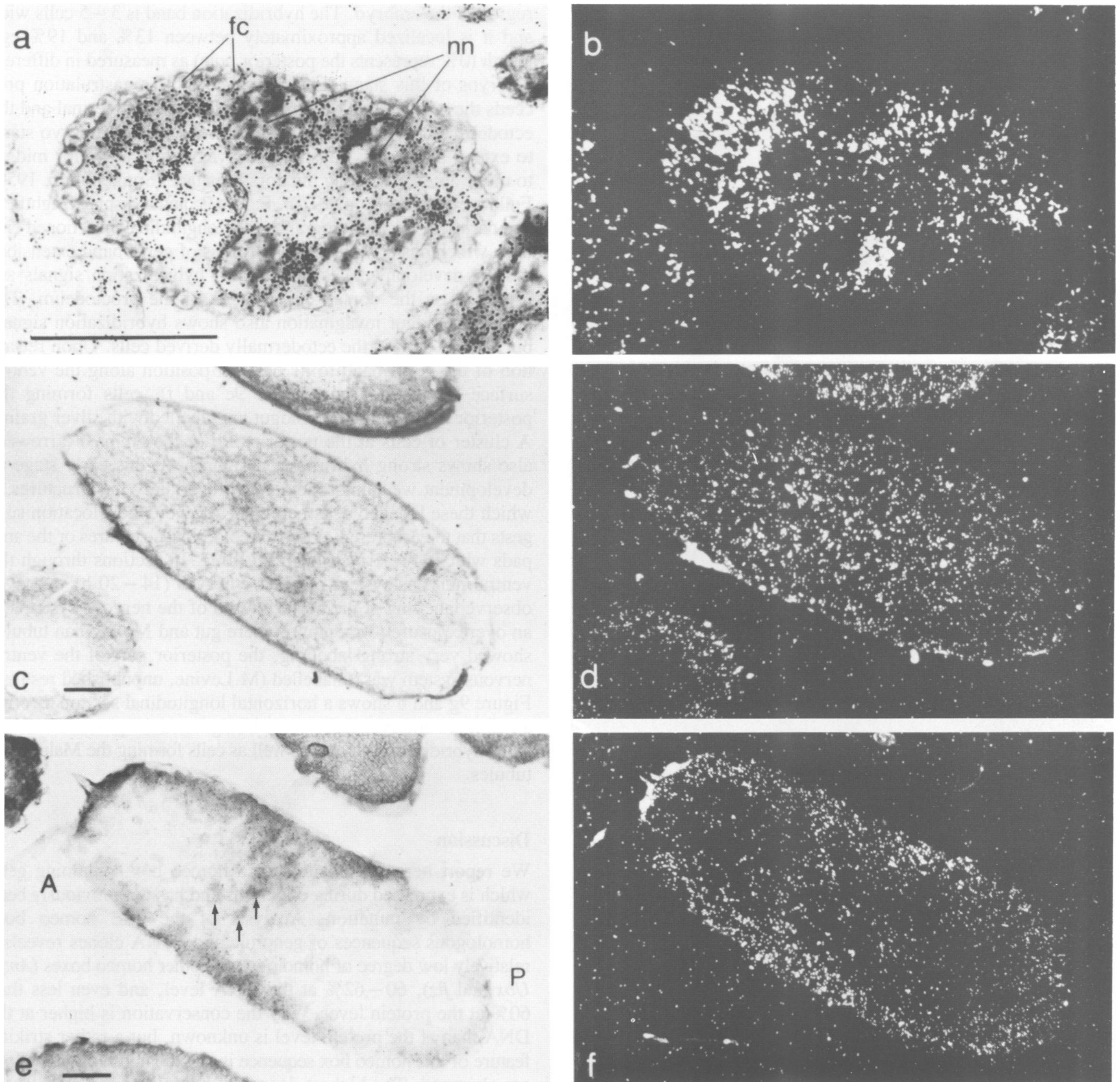

Fig. 7. Localization of 38E transcripts during oogenesis up to nuclear migration (~ 2 h after fertilization). Tissue sections from wild-type ovaries were hybridized with tritiated pSC335 probe, washed and autoradiographed for 15 days. **(a)** Section through early stage follicles (stage 7 – 8), the corresponding dark-field photomicrograph is shown in **(b)**. **(c)** Longitudinal section through an embryo ~ 1 h after fertilization. **(d)** A dark-field photomicrograph of the section shown in **(c)**. **(e)** A longitudinal section through an embryo during the process of nuclear migration; the arrows indicate the labelled cytoplasmic islands surrounding the nuclei. The corresponding dark-field photomicrograph is shown in **(e)**. fc, follicle cells; nn, nucleus of nurse cell; A, anterior; P, posterior. The horizontal bars represent a length of 0.05 mm.

central, yolk-containing region of the embryo (Figure 7e and 7f, arrows indicate cytoplasmic islands that are still located in the yolk-containing region). This probably reflects the displacement of cytoplasm which is known to co-migrate outwards with the nuclei (Foe and Alberts, 1983). After the process of nuclear migration is completed, a striking change in the distribution of silver grains is observed at the syncytial blastoderm stage. The transcripts disappear from the anterior end of the embryo, resulting in an anterio-posterior gradient pattern observed in all embryos at this stage (Figure 8). During further nuclear cleavages the gradient is shifted more towards the posterior end. The progressive loss of label from the anterior end does not seem to lead to an increased number of silver grains at the posterior end as judged by grain counting. The mechanism of formation of the gradient remains enigmatic, but from the above data we favour the hypothesis that localized transcript degradation is involved, although the possibility that the transcripts are transported cannot be excluded.

When cellularization is completed at the beginning of gastrulation a single band of labelled cells is observed in the posterior

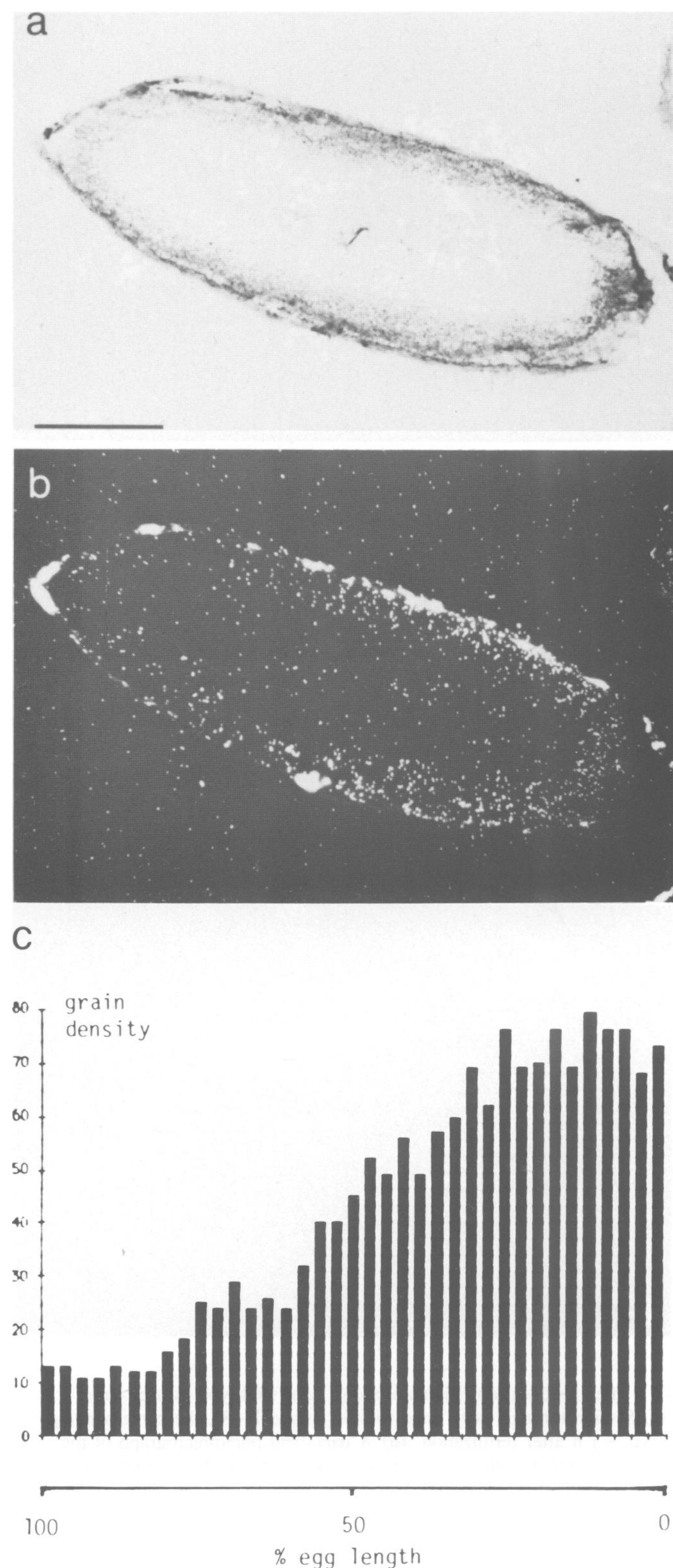

Fig. 8. Localization of 38E transcripts at the syncytial blastoderm stage (~2.5 h). **(a)** A longitudinal section through an embryo at the 11th to 12th nuclear division stage. The tissue section was hybridized with tritiated pSC335 probe, washed and autoradiographed for 21 days. **(b)** Darkfield photomicrograph corresponding to **(a)**. The anterior end of the section points to the left and the posterior to the right. The horizontal bar represents a length of 0.1 mm. Panel **(c)** shows a grain density diagram along the antero-posterior axis. Grains were counted in the cortical cytoplasm at a 400-fold magnification of the section shown in **(a)** and **(b)**. The scale in the vertical axis indicates (grains/cm^2)×2. In the horizontal axis 0% represents the posterior and 100% the anterior pole of the embryo.

region of the embryo. The hybridization band is 3–5 cells wide and it is localized approximately between 13% and 19% egg length (0% represents the posterior pole) as measured in different embryos of this stage (Figure 9a and b). As gastrulation proceeds the germ band, which consists of the mesodermal and the ectodermal cell layers, on the ventral side of the embryo starts to extend backwards causing the invaginating posterior midgut to move anteriorly along the dorsal surface (Sonnenblick, 1950; Fullilove and Jacobson, 1978). Figure 9c and d display brightfield and darkfield photomicrographs of a longitudinally sectioned embryo which has completed this process of germ band extension. At this developmental stage (~6 h) hybridization signals are observed on the dorsal side in cells of the proctodeum. The posterior midgut invagination also shows hybridization signals but weaker than in the ectodermally derived cells. Upon retraction of the germ band to its original position along the ventral surface of the embryo (Figure 9e and f), cells forming the posterior midgut and the hindgut are labelled with silver grains. A cluster of cells at the posterior tip of the embryo (arrowed) also shows strong hybridization signals. At this early stage of development we cannot identify with certainty the structures to which these labelled cells belong. However, their location suggests that they represent posterior epidermal structures or the anal pads which are derived from this area. In sections through the ventral nervous system of older embryos (14–20 h) we never observe labelling at the posterior end of the nervous system. In an overexposure (3 months), where gut and Malpighian tubules showed very strong labelling, the posterior part of the ventral nervous system was unlabelled (M.Levine, unpublished results). Figure 9g and h shows a horizontal longitudinal section through an 18-h old embryo. Cells of the hindgut and the posterior midgut show hybridization signals as well as cells forming the Malpighian tubules.

Discussion

We report here the isolation of a homeo box-containing gene which is expressed during oogenesis and has not previously been identified by mutation. Analysis of the 38E homeo box-homologous sequences of genomic and cDNA clones reveals a relatively low degree of homology with other homeo boxes (*Antp*, *Ubx* and *ftz*), 60–62% at the DNA level, and even less than 60% at the protein level. Why the conservation is higher at the DNA than at the protein level is unknown, but a rather striking feature of this homeo box sequence is that the homologous codons are clustered. The highest degree of homology can be detected in the region which shows significant homologies with the yeast mating type genes a1 and $\alpha 2$ (Shepherd *et al.*, 1984) and also with procaryotic DNA binding proteins (Laughon and Scott, 1984). This region includes the putative α-helix which is thought to bind to the DNA in the major groove and to confer sequence specificity. This part of the homeo box can be designated as the core element since it is most highly conserved in evolution. The core element is invariant in *Antp*, *Ubx*, *ftz* (McGinnis *et al.*, 1984b) and *Sex combs reduced (Scr)* (A.Kuriowa *et al.*, in preparation) and shows only one amino acid substitution in 38E, in *engrailed (en)* and in *engrailed-related (er)* (Fjose *et al.*, 1985; Poole *et al.*, 1985) which have diverged to an even larger extent than 38E. The core element is also invariant in two homeo box-containing genes from *Xenopus* (Carrasco *et al.*, 1984; Müller *et al.*, 1985), two from the mouse (McGinnis *et al.*, 1984c; Colberg-Poley *et al.*, 1985) and two human genes (Levine *et al.*, 1984).

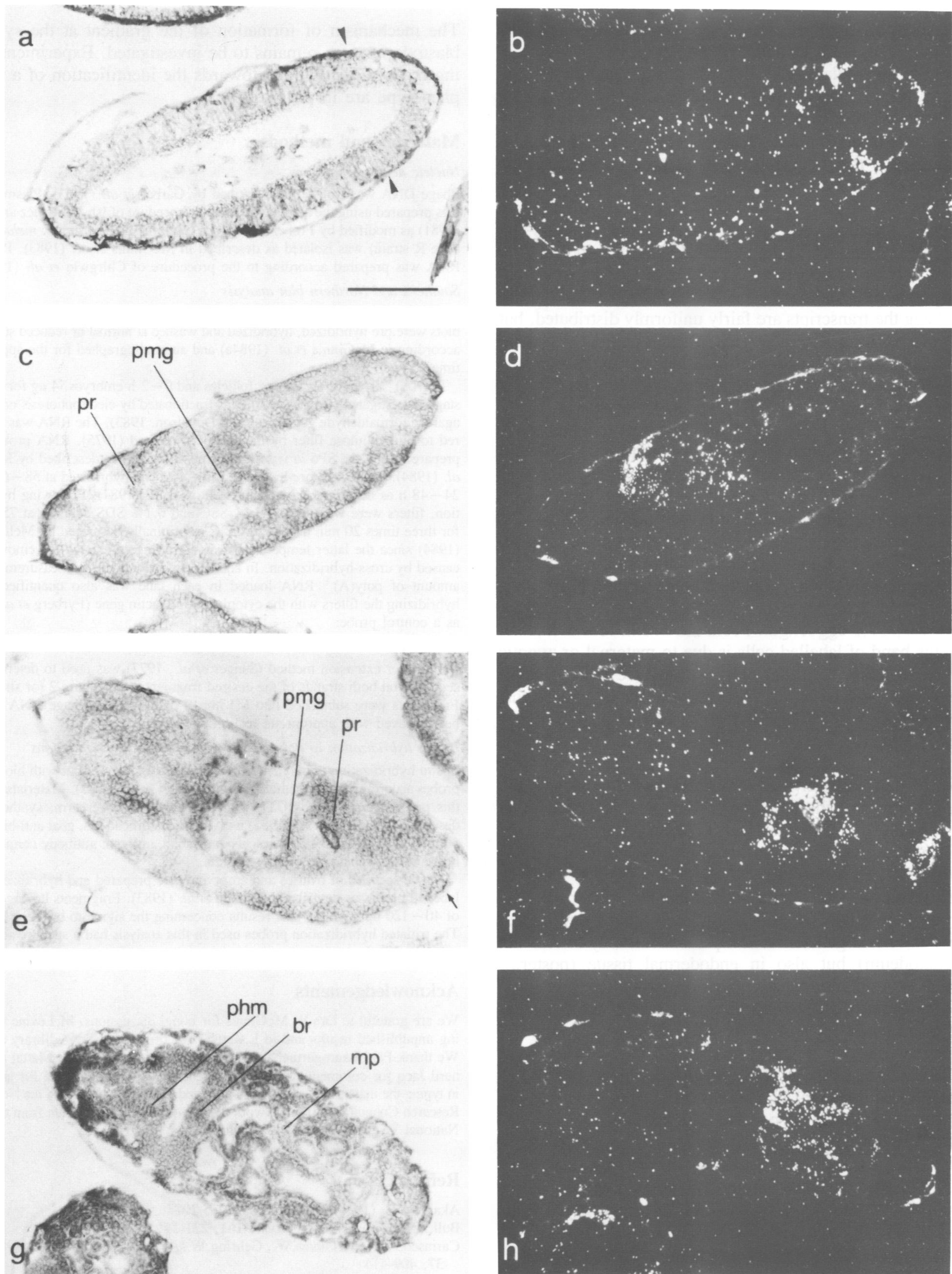

Fig. 9. Distribution of 38E transcripts at the cellular blastoderm and during later developmental stages. The anterior end of the sections always points to the left and the posterior to the right. The sections were hybridized to tritiated pSC335 probe, washed and autoradiographed for 21 days. The bright-field and the corresponding dark-field photomicrographs are shown. **(a)** and **(b)** A longitudinal section through an embryo of the late cellular blastoderm stage (~3 h). The arrows indicate cells that show hybridization signals. **(c)** and **(d)** A longitudinal section through an embryo at the extended germ band stage (6 h). A sagittal tissue section through an embryo which has completed germ band retraction (10 h) is shown in **(e)** and **(f)**. The arrows indicate labelled cells at the end of the germ band. **(g)** and **(h)** A longitudinal section through an 18 h old embryo. pr, proctodeum; pmg, posterior midgut invagination; mp, Malpighian tubules; phm, pharyngeal muscles; br, brain. The horizontal bar represent a length of 0.1 mm.

By Northern blotting we detected two maternal transcripts forming a doublet of ~2.4 kb and a zygotic transcript of ~2.6 kb that begins to accumulate in 2–4 h embryos, around the cellular blastoderm stage. The different sized transcripts could be due to the alternative use of promoters, different termination processing signals or alternative splicing. In the case of the maternal transcripts, the occurrence of a specific degradation intermediate cannot be excluded.

By *in situ* hybridization to tissue sections we can first detect transcripts in the nurse cells and the oocyte which suggests that they may be synthesized in the nurse cells and transported into the oocyte, since the oocyte nucleus is largely inactive during most of oogenesis (Mahowald and Kambysellis, 1980). In the fertilized egg the transcripts are fairly uniformly distributed, but appear to be excluded from the yolk. During nuclear migration the transcripts are associated with the islands of cytoplasm surrounding the nuclei which leads to a temporary accumulation in the cortical cytoplasm. During the syncytial blastoderm (9th to 13th nuclear division) a striking change is observed which leads to the formation of an antero-posterior concentration gradient of the transcripts. The mechanism of gradient formation cannot be deduced from our data, but may involve the specific degradation of the transcripts at the anterior pole, since there is no corresponding increase in the number of grains at the posterior pole. However, we cannot exclude the possibility that the transcripts may be transported during this phase. At the cellular blastoderm stage transcripts accumulate in a band of cells at the posterior end from 13–19% egg length. From our *in situ* results it is not clear if this band of labelled cells is due to maternal or zygotic transcripts since on Northern blots we observe both kinds of transcripts in 2–4 h embryos, but we favour the hypothesis that after cellularization new transcription starts from the 38E homeo box gene, so that this band of labelled cells would originate from zygotic transcription. Comparing this band of expression to the fate map of the blastoderm as described by Hartenstein *et al.* (1985), it seems that the 38E homeo box gene is expressed at this stage in the posterior-most part of the ectoderm in precursor cells of the proctodeum and possibly also in the posterior most segments A9 and A10.

During further development cells forming the proctodeum, the Malpighian tubules and the posterior midgut show expression of the 38E homeo box gene, so it is expressed not only in ectodermal (proctodeum) but also in endodermal tissue (posterior midgut). It is not certain which posterior parts of the epidermis express the described gene. However, it seems not to be expressed in the posterior part of the ventral nervous system. A more detailed analysis is necessary to answer these open questions. On the basis of the *in situ* hybridization data we call this gene *caudal (cad)*. The localization of transcripts at the cellular blastoderm stage seems to overlap at least partially with the posterior-most hybridization band seen in ftz^+ hybridization experiments (Hafen *et al.*, 1984) and is posterior to *infra-abdominal 7 (iab-7)*, the posterior-most homeo box-containing gene in the *bithorax* complex *(BX-C)* (Levine *et al.*, in preparation). This raises the possibility that *cad* originally belonged to the *BX-C*. It is interesting to note that the chromosomal deficiency *Df(3R)P9*, which removes the entire *BX-C* (Lewis, 1978), transforms all segments from T3 to A8 into thoracic segments, but leaves the most posterior segment, A9 with the anal pads, the tuft and the terminal sensory organs intact, which indicates that a gene outside the *BX-C* specifies the most posterior parts of the animal.

We have not yet identified a mutant phenotype for this gene, so that the function of the cad^+ products remains an open question. The mechanism of formation of the gradient at the syncytial blastoderm stage remains to be investigated. Experiments aiming in this direction and towards the identification of a mutant phenotype are in progress.

Materials and methods

Nucleic acids preparation

Phage DNA was isolated as described by Garber *et al.* (1983). Plasmid DNA was prepared using the alkaline extraction procedure of Ish-Horowicz and Burke (1981) as modified by Frei *et al.* (1985). Genomic DNA from *D. melanogaster* (Ore-R strain) was isolated as described in McGinnis *et al.* (1983). Poly(A)$^+$ RNA was prepared according to the procedure of Chirgwin *et al.* (1979).

Southern and Northern blot analysis

Southern blots were prepared essentially as described by Southern (1975). The blots were pre-hybridized, hybridized and washed at normal or reduced stringency according to McGinnis *et al.* (1984a) and autoradiographed for the appropriate time.

Poly(A)$^+$ RNA, 2 μg for egg follicles and 0–2 h embryos, 4 μg for all other stages investigated (see Figure 6), was fractionated by electrophoresis on a 1.1% agarose formaldehyde gel (Rozek and Davidson, 1983). The RNA was transferred to nitrocellulose filter by the Southern method (1975). RNA probes were prepared using the SP6 *in vitro* transcription method as described by Melton *et al.* (1984). Blots were pre-hybridized (2–4 h) and hybridized at 58–60°C for 24–48 h as described previously (Melton *et al.*, 1984). Following hybridization, filters were washed in 0.1 × SSC and 0.1% SDS solution at 75–80°C for three times 20 min instead of 65°C as originally described by Melton *et al.* (1984) since the latter temperature was not stringent enough to remove bands caused by cross-hybridization. In addition to optical density measurements, the amount of poly(A)$^+$ RNA loaded in each lane was also quantified by rehybridizing the filters with the cytoplasmic 5C actin gene (Fyrberg *et al.*, 1983) as a control probe.

DNA sequence analysis

The primer extension method (Sanger *et al.*, 1977) was used to determine the sequence on both strands of the desired fragments (see Figure 2 for strategies). Fragments were subcloned into M13mp10 or M13mp11 phage DNA that had been cleaved with appropriate restriction enzymes.

In situ hybridization to polytene chromosomes and tissue sections

In situ hybridization to polytene chromosomes was carried out with biotinylated probes according to the method of Langer-Safer *et al.* (1982). Materials used for this procedure were bio-dUTP with an 11-carbon linker arm, synthesised as described by Langer *et al.* (1981) with minor modifications, goat anti-biotin IgG purchased from Enzo-Biochemicals and rabbit anti-goat antibody conjugated to horseradish peroxidase from Dakopatts.

Tissue sections of ovaries and embryos were prepared and hybridized to ^{3}H-labelled probes as described by Hafen *et al.* (1983). Fragments having a length of 40–120 bp gave the best results concerning the signal-to-background ratio. The tritiated hybridization probes used in this analysis had a specific activity of $1-2 \times 10^8$ d.p.m./μg.

Acknowledgements

We are grateful to Drs W.McGinnis for initial discussions, M.Levine for sharing unpublished results and to L.Kauvar for providing cDNA library aliquots. We thank Paul Baumgartner for assistance in sequencing; Iain Mattaj and Bernard Jacq for comments and Ursula Weber and Christel Riedel for assistance in typing the manuscript. A.F. was supported by a fellowship from the Norwegian Research Council NAVF. The work was made possible by a grant from the Swiss National Science Foundation and the Kanton Basel-Stadt.

References

Akam,M.E. (1983) *EMBO J.*, **2**, 2075-2084.
Bull,A.L. (1966) *J. Exp. Zool.*, **161**, 221-241.
Carrasco,A.E., McGinnis,W., Gehring,W.J. and De Robertis,E.M. (1984) *Cell*, **37**, 409-414.
Chan,L.-N. and Gehring,W.J. (1971) *Proc. Natl. Acad. Sci. USA*, **68**, 2217-2221.
Chirgwin,J.M., Przybyla,A.E., McDonald,R.J. and Rutter,W.J. (1979) *Biochemistry (Wash.)*, **18**, 5294-5299.
Colberg-Poley,A.M., Voss,S.D., Chowdhury,K. and Gruss,P. (1985) *Nature*, **314**, 713-718.
Dennell,R.E. (1978) *Genetics*, **90**, 277-289.
Duncan,I. (1982) *Genetics*, **102**, 49-70.
Duncan,I. and Lewis,E.B. (1982) in Subtelny,S. (ed.), *Developmental Order:*

Its Origin and Regulation, Liss, NY, pp. 533-554.
Fjose,A., McGinnis,W.J. and Gehring,W.J. (1985) *Nature*, **313**, 284-289.
Foe,V.E. and Alberts,B. (1983) *J. Cell Sci.*, **61**, 31-70.
Frei,E., Baumgartner,S., Edström,J.E. and Noll,M. (1985) *EMBO J.*, **4**, 979-984.
Fullilove,S.L. and Jacobson,A.G. (1978) in Ashburner,M. and Wright,T.R.F. (eds.), *The Genetics and Biology of Drosophila*, Vol. **2c**, Academic Press, NY, pp. 106-227.
Fyrberg,E.A., Mahaffrey,J.W., Bond,B.J. and Davidson,N. (1983) *Cell*, **33**, 115-123.
Garber,R.L., Kuriowa,A. and Gehring,W.J. (1983) *EMBO J.*, **2**, 2027-2036.
Green,M.R., Maniatis,T. and Melton,D.A. (1983) *Cell*, **32**, 681-694.
Hafen,E., Levine,M., Garber,R.L. and Gehring,W.J. (1983) *EMBO J.*, **2**, 617-623.
Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984) *Cell*, **37**, 833-841.
Hartenstein,V., Technau,G.M. and Campos-Ortega,J.A. (1985) *Wilhelm Roux's Arch. Dev. Biol.*, **194**, 213-216.
Ingham,P.W. (1984) *Cell*, **37**, 815-823.
Ish-Horowicz,D. and Burke,J.F. (1981) *Nucleic Acids Res.*, **9**, 2989-2998.
Kaufman,T.C., Lewis,R. and Wakimoto,B. (1980) *Genetics*, **94**, 115-133.
Kornberg,T., Siden,I., O'Farell,P. and Simon,M. (1985) *Cell*, **40**, 45-53.
Kuroiwa,A., Hafen,E. and Gehring,W.J. (1984) *Cell*, **37**, 825-831.
Langer,P.R., Waldrop,A.A. and Ward,D.C. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6633-6637.
Langer-Safer,P.R., Levine,M. and Ward,D.C. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4381-4385.
Laughon,A. and Scott,M.P. (1984) *Nature*, **310**, 25-31.
Lawrence,P.A. and Morata,G. (1977) *Dev. Biol.*, **56**, 40-51.
Levine,M., Hafen,E., Garber,R.L. and Gehring,W.J. (1983) *EMBO J.*, **2**, 2037-2046.
Levine,M., Rubin,G.M. and Tjian,R. (1984) *Cell*, **38**, 667-673.
Lewis,E.B. (1978) *Nature*, **276**, 565-570.
Lohs-Schardin,M. (1982) *Wilhem Roux's Arch. Dev. Biol.*, **191**, 38-36.
Mahowald,A.P. and Kambysellis,M.P. (1980) in Ashburner,M. and Wright,T.R.F. (eds.), *The Genetics and Biology of Drosophila*, Vol. **2d**, Academic Press, NY, pp. 141.224.
Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Conell,C., Quon,D., Sim,D.K. and Efstratiadis,A. (1978) *Cell*, **15**, 687-701.
McGinnis,W., Shermoen,A.W. and Beckendorf,S.K. (1983) *Cell*, **34**, 75-84.
McGinnis,W., Levine,M., 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.
McGinnis,W., Hart,C.P., Gehring,W.J. and Ruddle,F.H. (1984c) *Cell*, **38**, 675-680.
Melton,D.A., Krieg,P.A., Rebagliatti,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035-7056.
Müller,M.M., Carrasco,A.E. and DeRobertis,E.M. (1984) *Cell*, **39**, 157-162.
Nüsslein-Volhard,C. (1979) in Konigsberg,I. and Subtelny,S. (eds.), *Determinants of Spatial Organisation*, Academic Press, NY, pp. 185-211.
Pabo,C.O. and Sauer,R.T. (1984) *Annu. Rev. Biochem.*, **53**, 293-321.
Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37-43.
Rozek,C.E. and Davidson,N. (1983) *Cell*, **32**, 23-34.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Scott,M.P. and Weiner,A.J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4115-4119.
Shepherd,J.C.W., McGinnis,W., Carrasco,A.E., DeRobertis,E.M. and Gehring,W.J. (1984) *Nature*, **310**, 70-71.
Sonnenblick,B.P. (1950) in Demerec,M. (ed.), *Biology of Drosophila*, Wiley, NY, pp. 62-167.
Southern,E. (1975) *J. Mol. Biol.*, **98**, 503-517.
Steward,R., McNally,F.J. and Schedl,P. (1984) *Nature*, **311**, 262-265.
Struhl,G. (1981) *Nature*, **293**, 36-41.
Struhl,G. (1983) *J. Embryol. Exp. Morphol.*, **76**, 297-331.
Wharton,K.A., Yedvobnick,B., Finnerty,V.G. and Artavanis-Tsakonas,S. (1985) *Cell*, **40**, 55-62.
Wieschaus,E. and Gehring,W.J. (1976) *Dev. Biol.*, **50**, 249-263.

Received on 3 July 1985; revised on 19 August 1985