The EMBO Journal vol.7 no.8 pp.2569–2578, 1988

Molecular structure and spatial expression of a homeobox gene from the *labial* region of the Antennapedia-complex

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

Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland
[1]Present address: Laboratory of Biotechnology, University of Bergen, P.Box 3152 Arstad, N-5001 Bergen, Norway

Communicated by W.J.Gehring

We report the molecular characterization of a homeobox-containing gene that maps at 84A in the proximal region of the Antennapedia-complex. The structure and complete sequence are presented. Deletion analysis indicates that the cloned gene, F24, most likely corresponds to the *labial* (*lab*) gene. Northern blot experiments show a single ~3-kb transcript that is expressed at all embryonic stages from cellular blastoderm onwards and during larval development. The homeobox is split by an intron in the region which encodes the putative DNA-binding helix, a splicing position for homeobox-containing genes which is unique so far. The 5′ part of the gene contains four M-repeat sequences (CAA/G repeats) in the protein-coding region. *In situ* hybridization to the transcripts during embryogenesis reveals two domains of expression. The anterior one is located in parts of the developing head, mainly in the hypopharyngeal organ and in anterior parts of the mandibular lobe, and is restricted to the ectoderm. The posterior domain is part of the posterior midgut primordium (endoderm), that invaginates and later contacts the endoderm cells from the anterior midgut invagination.

Key words: Antennapedia-complex/homeobox/*in situ* hybridization/*Drosophila*/*labial*

Introduction

The spatial and temporal development of the *Drosophila* embryo is under the control of several sets of genes. Firstly, the maternal genes define the axes and the polarity of the egg (antero-posterior and dorso-ventral) before fertilization (for reviews, see Nüsslein-Volhard *et al.*, 1987; Anderson, 1987). After fertilization the embryo becomes first subdivided into metameric units by the zygotic segmentation genes (Nüsslein-Volhard and Wieschaus, 1980; for review, see Akam, 1987), and slightly later the precise expression of the homeotic selector genes leads to the correct segment identity in each of these units in the embryo, larva and the adult fly (for review, see Gehring and Hiromi, 1986; Akam, 1987).

Most of the homeotic selector genes have been genetically characterized and are clustered in two complexes on the 3rd chromosome. The bithorax-complex (BX-C) contains genes that are responsible for the correct identity of posterior thoracic and abdominal segments and compartments (Lewis, 1978). The Antennapedia-complex (ANT-C) appears to contain genes that are required for normal head and thoracic development (Kaufman, 1983; Kaufman and Abbott, 1984). More recently other genes with similar function(s) have been found outside these complexes, e.g. *caudal* (*cad*) (Mlodzik *et al.*, 1985; Macdonald and Struhl, 1986) or *spalt* (*sal*) (Jürgens, 1988; Frei *et al.*, 1988). The molecular cloning and analysis of regions of the BX-C (Bender *et al.*, 1983) and the ANT-C (Garber *et al.*, 1983; Scott *et al.*, 1983) led to the discovery of a conserved DNA segment in the coding region of some of these genes, which is called the homeobox (McGinnis *et al.*, 1984a,b; Scott and Weiner, 1984). Using this DNA segment as a probe, other related genes within and outside these complexes have been isolated and analyzed [e.g. *engrailed* (*en*), Fjose *et al.*, 1985; *cad*, Mlodzik *et al.*, 1985; *evenskipped* (*eve*), Harding *et al.*, 1986; Macdonald *et al.*, 1986; and *zerknüllt* (*zen*), Doyle *et al.*, 1986].

In the ANT-C several genes have been found to contain a homeobox: *Antennapedia* (*Antp*) and *fushi tarazu* (*ftz*) (McGinnis *et al.*, 1984a; Scott and Weiner, 1984), *Sex combs reduced* (*Scr*) (Kuroiwa *et al.*, 1985), *Deformed* (*Dfd*) (McGinnis *et al.*, 1984a; Regulski *et al.*, 1987), *zen* (Doyle *et al.*, 1986) and *bicoid* (*bcd*) (Frigerio *et al.*, 1986). All these genes show specific patterns of expression along the antero-posterior axis (e.g. Levine *et al.*, 1983; Hafen *et al.*, 1984; Kuroiwa *et al.*, 1985; Chadwick and McGinnis, 1987; Frigerio *et al.*, 1986; for review, see Akam, 1987) or are restricted to dorsal tissue (*zen*, Doyle *et al.*, 1986). It is thought that they all select and control developmental pathways by modulating the activity of so-called realizator genes (Garcia-Bellido, 1975), whose differential expression in each metamere determines segmental identity. Therefore, it appears reasonable to assume that at least some of the homeotic genes specifically activate or inhibit gene expression by binding to DNA. Indeed, the homeobox shares homology to DNA-binding proteins from yeast and procaryotes (Shepherd *et al.*, 1984; Laughon and Scott, 1984), and all the homeobox proteins analyzed so far are localized to the nucleus (White and Wilcox, 1984; Beachy *et al.*, 1985; DiNardo *et al.*, 1985; Carroll and Scott, 1985; Carroll *et al.*, 1986; Wirz *et al.*, 1986; Macdonald and Struhl, 1986; Mlodzik and Gehring, 1987; Frasch *et al.*, 1987; Rushlow *et al.*, 1987; Riley *et al.*, 1987; Mahaffey and Kaufman, 1987). Furthermore, the predicted control function(s) for the homeotic selector genes are supported by DNA binding studies with some of these proteins (U.Weber and W.J.Gehring, unpublished results).

Here we describe the isolation of a homeobox-containing gene from the proximal region of the ANT-C. Its molecular structure and sequence have been determined. Preliminary sequence data of the homeobox and transcript localization have been published previously (F90-2 in Hoey *et al.*, 1986; F121 in Gehring, 1987). By analysis of several deletions, we conclude that this transcription unit most likely

©IRL Press Limited, Oxford, England

corresponds to the *labial* gene (*lab*, Kaufman, 1983; Kaufman and Abbott, 1984). *In situ* localization of its transcripts reveals two regions of expression. The first domain is anterior to the stripe of *Dfd* expression (Chadwick and McGinnis, 1987) in parts of the developing head, and the second domain is in the very posterior somatic blastoderm tissue, which later gives rise to regions of the middle midgut.

Results

Isolation and cytological mapping of genomic clones

We screened the *D.melanogaster* genomic library of Maniatis *et al.* (1978) under reduced stringency conditions (McGinnis *et al.*, 1984a) with a homeobox-containing fragment from the *Ultrabithorax* (*Ubx*) gene as a probe. Beside clones originating from the *en* region (Fjose *et al.*, 1985) and *cad* (Mlodzik *et al.*, 1985), several other homeobox-containing clones were isolated. One of the, λF24, was found to hybridize to chromosomal band 84A within the ANT-C (Figure 1A). None of the previously well characterized homeobox genes of the ANT-C shares homology with this clone outside the homeobox or the M-repeats (see below) or has a similar restriction enzyme map. Adjacent genomic clones were isolated by the technique of genomic walking, and the distal portion of the walk is shown in Figure 2. To better define the position of our cloned DNA within the ANT-C, we performed *in situ* hybridization analysis to balanced deficiency (Df) chromosomes of the ANT-C. Df(3R)*Scr* (Wakimoto *et al.*, 1984) was found to uncover the entire cloned region. In contrast, the cloned segment at 84A was found to lie proximally outside Df(3R)Scx^{W+RX2}. The *labial* gene behaves identically concerning these deficiencies (Kaufman, 1983; Kaufman and Abbott, 1984; see Discussion and Figure 9). These results, together with the facts that all other homeotic genes in the ANT-C contain a homeobox and the spatial control of F24 transcription during embryogenesis is reminiscent of other expression patterns of homeotic genes (see below), strongly suggest that the cloned F24 region corresponds to the *labial* gene. However, mutant rescue experiments with the P-element-mediated germ-line transformation remain to be done. The proximal breakpoint of Df(3R)9A99 which deletes the entire ANT-C, except *Antp* (Jürgens *et al.*, 1984) was found in the proximal part of the F24 clone (see Figure 1B,C and Figure 2). The distal breakpoint has been mapped previously by Kuroiwa *et al.* (1985, for summary see Figure 9).

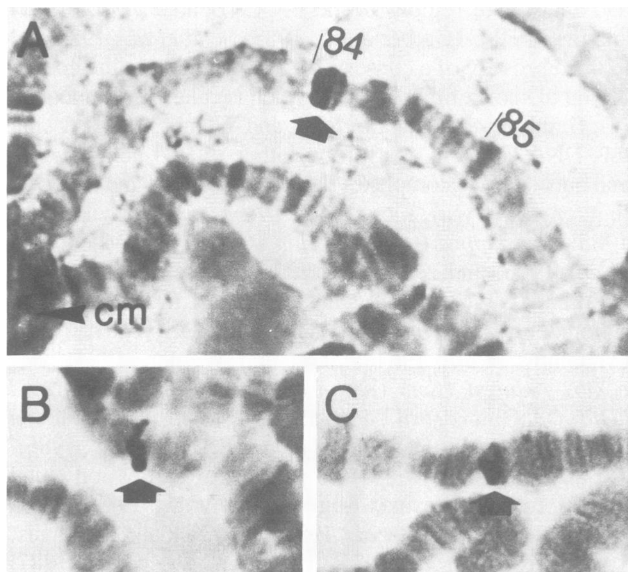

Fig. 1. *In situ* hybridization of biotinylated clones from the F24 region to salivary gland polytene chromosomes. (**A**) λF24 hybridized to chromosomes of Oregon R flies. The signal is at 84A. (**B**) and (**C**) Mapping of proximal breakpoint of Df(3R)9A99. λF24 (**B**) and λF24.2 (**C**), see Figure 2, were hybridized to chromosomes of balanced stocks Df(3R)9A99/TM6. Note that the breakpoint is located within clone F24 (for summary see Figure 2). The signals are indicated by arrows. cm: centromere.

Transcriptional organization at the F24 locus

The homeobox transcript of F24 is transcribed in the same orientation relative to the centromere as *Antp* and *Scr*, namely from distal to proximal. It spans ~18 kb and contains three exons, the 5′ exon being separated by ~14 kb from the two 3′ exons. The homeobox is located in both 3′ exons which are separated by a ~260-bp intron (see Figures 2 and 6). A transcript containing this homeobox can be detected on Northern blots and is ~3 kb long [including

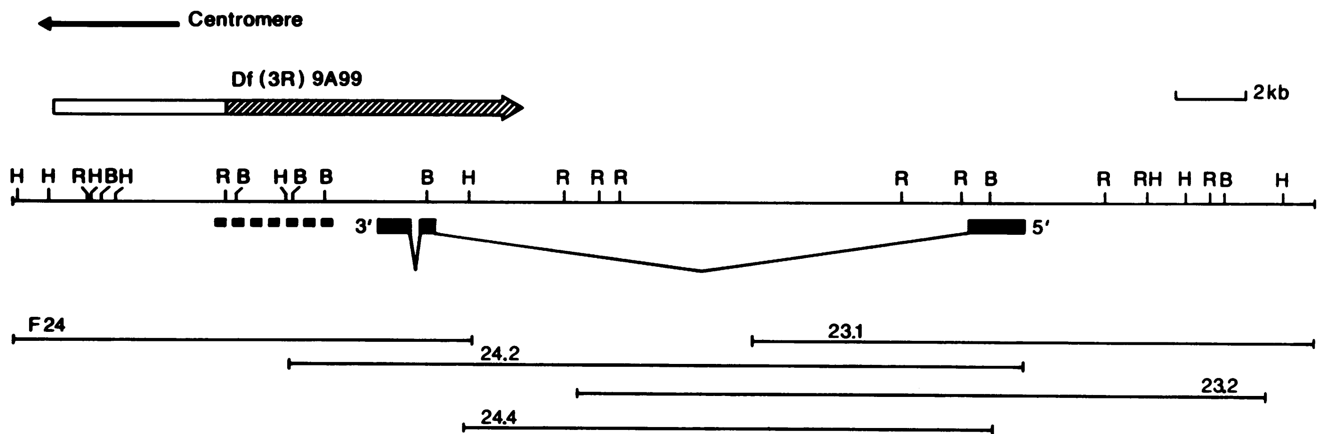

Fig. 2. Genomic organization of cloned DNA at 84A. Black boxes present exons of the F24 homeobox transcript. Broken line indicates second transcriptional unit; the exons were not mapped accurately for this transcript. Breakpoint of Df(3R)9A99 is indicated; open box represents region of uncertainty. The extent of the analyzed clones is shown below the restriction map. Centromere (see arrow) is to the left. Restriction enzymes: B = *Bam*HI, R = *Eco*RI, H = *Hin*dIII.

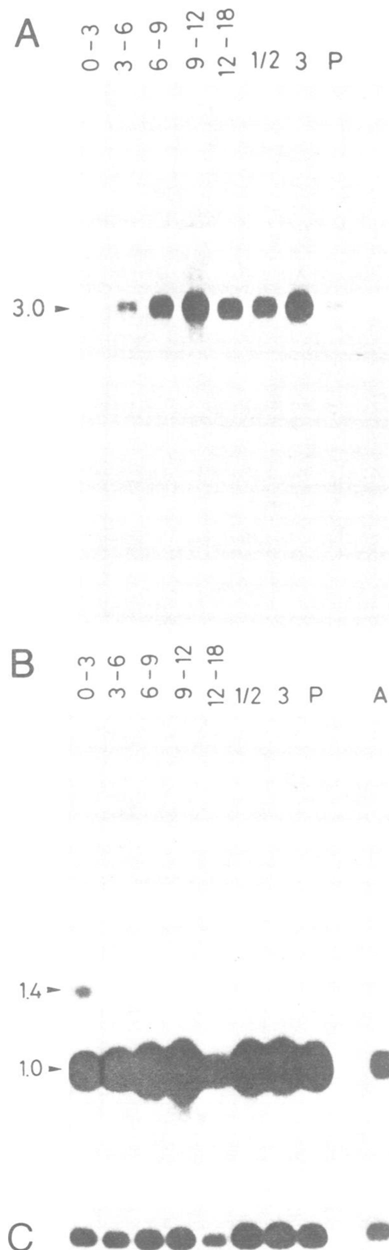

Fig. 3. Transcription from cloned 84A region during *Drosophila* development. Poly(A)$^+$ RNA (2 μg/lane) was subjected to Northern blot analysis and hybridized with specific probes at high stringency. (**A**) Blot probed with the F24 homeobox. A single mRNA species of ~3.0 kb [including poly(A) tail] is detected. The same band is detected by hybridization with the 5′ exon (see Figure 2) or the cDNA clone c241 (see Figure 6) (data not shown). Probes from the large intron (see Figure 2) did not detect any transcript. (**B**) The same blot rehybridized with a probe from adjacent proximal transcription unit. A major 1.0-kb and a minor 1.4-kb mRNA (0−3 h) are detected. (**C**) Control hybridization with *Drosophila* elongation factor 2. Embryonic stages (first 5 lanes) are indicated in hours. 1/2: 1st and 2nd instar larvae, 3: 3rd instar larvae, P: pupal, A: adults (males and females).

poly(A) tail, Figure 3A]. It starts to be expressed ~3 h after fertilization at the cellular blastoderm stage and accumulates throughout embryogenesis. A second peak of expression is observed in the 3rd instar larva. Weak expression can be detected in pupal stages and no signal is found in adult flies. However, we cannot exclude traces of expression at this stage which are below the limits of detection.

Within the clone F24 another transcriptional unit can be detected. It is located downstream of the homeobox transcript (referred to as F24 below) and it appears to be transcribed in the opposite direction to F24. Northern blot analysis reveals two mRNA species of ~1.4 and 1.0 kb [including poly(A) tail, see Figure 3B]. The 1.0-kb transcript is present at high levels throughout the *Drosophila* life cycle. However, the 1.4-kb RNA appears only in 0−3 h old embryos, and, when the blot is overexposed the transcript can be detected in adults as well (data not shown). This suggests that the minor RNA species is maternally transcribed and still present in young embryos. *In situ* hybridization to embryonic sections has revealed that these transcripts, in contrast to F24 (see below), do not show any specific localization, but are distributed uniformly throughout the developing embryo at all embryonic stages (data not shown).

Structure and sequence of the F24 gene

Using probes containing the F24 homeobox and flanking sequences we have isolated cDNA clones from a 3−12 h embryonic library (kindly provided by L.Kauvar). Several clones are from the homeobox region and of minor size, but one contains the complete 3′ end [poly(A)-tailed] and a large portion of the 5′ exon. This cDNA, c241, and the corresponding genomic clones were sequenced on both strands (see Materials and methods). The determined nucleotide sequence of the transcript including upstream sequences up to −390, and the deduced amino acid sequence, are shown in Figure 4. Terminal sequences of the ~14-kb intron and the complete second intron splitting the homeobox have been determined as well (data not shown, Figure 5C). The sequence of the homeobox and the small intron have been published previously (F90-2 by Hoey *et al.*, 1986; F121 by Gehring, 1987). To determine the 5′ end of the transcript we performed S1 protection and primer extension experiments. For S1 protection analysis the complementary strand was generated in an M13-vector using the *Pst*I fragment extending from −64 to +392 in Figure 4. Including M13 primer and polylinker sequences, it had a size of 536 nucleotides and the protected fragment was 393 nucleotides long (see Figure 5A). This result was confirmed by primer extension analysis using a synthetic oligonucleotide primer (21-mer) extending from +40 to +61 (inclusive), which was extended for 39 nucleotides with total and poly(A)$^+$ RNA from 12−18 h old embryos (data not shown). Both experiments place the initiation site at position +1 (±2) in Figure 4. The corresponding sequence ATCAGTC (underlined in Figure 4) is a perfect match to the proposed cap-site consensus for insect genes as proposed by Hultmark *et al.* (1986). However, we cannot exclude minor transcription start sites, which might not be detected at these levels of sensitivity. No TATA box homology is found at the appropriate distance from the proposed initiation site (see Figure 4). The untranslated leader of the transcript is 239 nucleotides long, and the open reading frame (ORF) starts at the first ATG at position 240 and is followed by 1887 translated nucleotides which code for a 629 (or 635, see below) amino acid long peptide (Figure 4 and 6). The trailer sequence is 673 bases long and contains two potential polyadenylation signals close to its 3′ end (underlined in Figure 4).

Sequence analysis of the intron−exon junction at the 5′ end of the second exon reveals two 16 nucleotide long direct repeats spaced by two additional nucleotides both sharing

```
-390
AGATCTGATCATCCCCAGCCAGTTCCACACCAGCTCATTGCATAATTTCAAGTTCCTTCATAGAACTCCCATCTTGAGTAGCCGTCAATCCGCTCCTGCCATGGCGACTAGCAAAAAAAA

-270
AAACGAGCAAACAGGGAACGAGAAAATCTAAATGGAAACGGCAACTGCCAAAAACCGGTAGTTGAAATACGGGTATTTGTATATACGTATGACTAACCCCCGAACCACTGATAAGGGCAG

-150
TCATCCGAAGAATCCCAGGGAGCGAGCGAGACGGCAGTCGCAACACGTATGTAGCACACACGAGAGGAAGCGGGACCACAGGCAGTCTGCAGAGGGGCGTGGCCAAGACCAGCGGTTGTG

-30                             1►
CGGTCTGAAAGAAACCGGGTTCGGGCCAGTAATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAA

90
CCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGTGATACACCGGTTATATCGGAGT

210                                                                                                                      30
                              MetMetAspValSerSerMetTyrGlyAsnHisProHisHisHisHisProHisAlaAsnAlaTyrAspGlyTyrSerThrThrThrAla
GGCGAGAAAGTGTGGTTCCGGCTGGACAATATGATGGACGTAAGCAGCATGTACGGCAACCACCCGCACCACCACCATCCACATGCCAATGCCTACGACGGCTACAGCACCACCACCGCG

330                                                                                                                      70
SerAlaAlaAsnAlaSerSerTyrPheAlaProGlnGlnHisGlnProHisLeuGlnLeuGlnGlnGlnGlnGlnHisGlnHisLeuGlnGlnProGlnGlnHisLeuThrTyrAsnGly
TCGGCGGCCAACGCCAGCAGCTACTTTGCTCCGCAGCAACACCAGCCCCACCTGCAGCTGCAGCAGCAGCAACAGCATCAGCATTTGCAGCAGCCCCAGCAGCACCTCACCTACAATGGC

450                                                                                                                     110
TyrGluSerSerSerProGlyAsnTyrTyrProGlnGlnGlnAlaGlnLeuThrProProProThrSerSerHisGlnValValGlnGlnHisGlnGlnGlnGlnGlnAlaGlnGlnGln
TACGAGAGCAGCTCCCCAGGCAACTACTATCCGCAGCAGCAGGCGCAGCTGACACCACCGCCCACCAGCAGCCACCAGGTGGTGCAGCAGCACCAGCAACAGCAGCAGGCGCAGCAGCAA

570                                                                                                                     150
GlnLeuTyrProHisSerHisLeuPheSerProSerAlaAlaGluTyrGlyIleThrThrSerThrThrThrGlyAsnProGlyLysProLeuHisProSerSerHisSerProAlaAsp
CAGCTCTATCCGCATTCGCATCTCTTTAGTCCCAGTGCGGCGGAGTATGGCATCACCACGAGTACGACCACGGGAAATCCGGGCAAGCCGCTCCATCCTAGTAGCCACTCTCCGGCGGAC
                                                                                      └c241

690                                                                                                                     190
SerTyrTyrGluSerAspSerValHisSerTyrTyrAlaThrAlaAlaValAlaThrValAlaProProSerAsnSerSerProIleThrAlaAlaAsnAlaSerAlaThrSerAsnThr
TCCTACTACGAAAGTGACTCGGTGCACTCGTACTACGCTACCGCCGCAGTGGCCACAGTTGCTCCACCCAGCAATAGCTCACCTATTACCGCCGCCAATGCAAGTGCCACCAGCAACACA

810                                                                                                                     230
GlnGlnGlnGlnGlnAlaAlaIleIleSerSerGluAsnGlyMetMetTyrThrAsnLeuAspCysMetTyrProThrAlaGlnAlaGlnAlaProValHisGlyTyrAlaGlyGln
CAACAACAGCAGCAACAGGCAGCAATCATCAGCTCCGAGAACGGGATGATGTACACCAACCTGGACTGCATGTATCCCACGGCCCAGGCTCAGGCTCCGGTTCACGGATATGCCGGCCAG

930                                                                                                                     270
IleGluGluLysTyrAlaAlaValLeuHisAlaSerTyrAlaProGlyMetValLeuGluAspGlnAspProMetMetGlnGlnAlaThrGlnSerGlnMetTrpHisHisGlnGlnHis
ATCGAGGAGAAATACGCCGCCGTCCTGCACGCCAGCTATGCACCCGGAATGGTGCTGGAGGATCAGGATCCAATGATGCAGCAGGCCACGCAGTCGCAGATGTGGCACCACCAGCAACAT

1050                                                                                                                    310
LeuAlaGlySerTyrAlaLeuAspAlaMetAspSerLeuGlyMetHisAlaHisMetHisHisGlyLeuProHisGlyHisLeuGlyAsnLeuAlaAsnAsnProHisGlnGlnGlnPro
CTGGCAGGCAGCTATGCCCTAGATGCCATGGACTCGCTAGGAATGCACGCCCACATGCATCACGGCCTGCCCCACGGACACCTGGGCAACTTGGCCAACAATCCGCATCAGCAACAGCCA

1170                                                                                                                    350
GlnValGlnGlnGlnGlnGlnProHisGlnGlnProGlnHisProGlnAsnGlnSerProAlaAlaHisGlnGlnHisHisGlnAsnSerValSerProAsnGlyGlyMetAsnArg
CAAGTCCAGCAGCAACAGCAGCAGCCGCACCAGCAACCGCAACATCCGCAGAACCAATCCCCAGCGGCACACCAGCAGCACCACCAGAACTCCGTGTCGCCCAACGGTGGAATGAATCGC

1290                                                                                                                    390
GlnGlnArgGlyGlyValIleSerProGlySerSerThrSerSerSerThrSerAlaSerAsnGlyAlaHisProAlaSerThrGlnSerLysSerProAsnHisSerSerSerIlePro
CAGCAGCGCGGGGGAGTGATCTCCCCGGGTAGCTCCACTTCCTCCTCCACCTCCGCCTCGAATGGAGCACATCCTGCCAGCACGCAATCAAAGTCGCCAAATCATTCCAGCAGCATCCCC

1410                                             ▼                                                                      430
ThrTyrLysTrpMetGlnLeuLysArgAsnValProLysProGlnAlaProLysLeuProAlaSerGlyIleAlaSerMetHisAspTyrGlnMetAsnGlyGlnLeuAspMetCysArg
ACCTACAAGTGGATGCAACTCAAGAGGAATGTTCCAAAGCCTCAAGCACCGAAACTACCTGCCAGTGGAATCGCCAGCATGCACGACTACCAGATGAATGGACAGCTGGACATGTGCCGC
                                         AlaProSerTyrPhePro
                                          CACCTTCTTACCTTCCAG

1530                                                                                                                    470
GlyGlyGlyGlyGlyGlySerAspValGlySerGlyProValGlyValGlyGlyAsnGlySerProGlyIleGlyGlyValLeuSerValGlnAsnSerLeuIleMetAlaAsnSerAla
GGTGGAGGAGGCGGGGGCAGCGACGTCGGGAGCGGTCCCGTGGGCGTTGGGGGCAACGGATCCCCGGGATCGGAGGCGTCCTTTCCGTGCAGAACTCCCTGATAATGGCGAACAGTGCG

1650                                                                                                                    510
AlaAlaAlaGlySerAlaHisProAsnGlyMetGlyValGlyLeuGlySerGlySerGlyLeuSerSerCysSerLeuSerSerAsnThrAsnAsnSerGlyArgThrAsnPheThrAsn
GCGGCGGCGGGAAGTGCCCATCCCAACGGCATGGGCGTGGGTCTGGGCAGCGGATCCGGACTGAGCAGCTGCAGCCTCTCCAGCAACACCAACAACTCCGGCCGGACGAACTTCACCAAC

1770                                                                                                          ▼         550
LysGlnLeuThrGluLeuGluLysGluPheHisPheAsnArgTyrLeuThrArgAlaArgArgIleGluIleAlaAsnThrLeuGlnLeuAsnGluThrGlnValLysIleTrpPheGln
AAGCAGCTGACCGAGCTGGAAAAAGAGTTCCACTTCAATCGCTACTTGACGCGGGCGCGCCGCATTGAAATCGCCAATACGTTGCAGCTTAATGAAACGCAGGTCAAAATCTGGTTCCAG

1890                                                                                                                    590
AsnArgArgMetLysGlnLysLysArgValLysGluGlyLeuIleProAlaAspIleLeuThrGlnHisSerThrSerValIleSerGluLysProProGlnGlnGlnGlnProGlnPro
AACCGCCGCATGAAGCAGAAGAAGCGCGTGAAGGAGGGCCTCATTCCGGCGGACATCCTGACGCAGCACTCCACGTCCGTGATCAGCGAGAAGCCGCCGCAGCAGCAGCAGCCACAGCCG

2010                                                                                                                    630
ProGluLeuGlnLeuLysSerGlnGlySerAspLeuGlyGlyAsnGluLeuAlaThrGlyAlaProSerThrProThrThrAlaMetThrLeuThrAlaProThrSerLysGlnSer***
CCGGAGCTGCAGTTGAAGTCGCAGGGCAGCGATCTGGGCGGGAATGAGCTGGCCACAGGAGCACCTTCGACACCCACGACAGCCATGACACTGACAGCACCCACGAGCAAGCAAAGTTGA

2130
AGGGCTTTCTTCGATGTTGGTTGTAATTGTTATGCTATTGTTGTTTCTGCCGCCAAAATATTGACGCATTTCGTTGTGATAAAATAGTAGAAAGTGCAGAGAGGATCAGAAGGAAATAAA

2250
CTAAGTAATTAAACAAATAATACATACAAATTATACACACACACATAGCAAATTGAAATCCGTAATCAAACGCTAGCAACTAGTGGAATCTAGGAGTCTAGGTGCCATATTCATTCACGT

2370
AGGGGCTGTAAATACTTACACTTACATTTCGCATTGCCGCAGCTGAAGTTTTTGGACTAGCTTAAACTGAACTCAGGGTTAAATTAAAATGTAACTTACTTAAACGCACGCAGAGCAATT

2490
GACAAAATACTGATGCAAACAGAATCAAAACAGAAATCGAAAATCTAAATCTTAACTATTTAAAACTTTCCAATAACTATAGCTAGTAGACCCCTAGTTATGTATATTCCACCCGCCACC

2610
ACAACTCGTCTGTAGTATTATTAGCTAAGTTCTAGAATGCCCAAGTTATTCGTTTTGTGCTTAAAAGTAATCGGTAAATCTTGATGTAAATCAACGTCTGTATATTTAATCTAGTCCTAA

2730
TTGTAATATAATTATAACTATACACCACTTGTATGAATAAACAATAAACGAAGCCCGATTGCAAATTGTT
```

Fig. 4. DNA and deduced amino acid sequence of the F24 homeobox gene. The sequences of the transcript (excluding the introns) and of flanking 5′ upstream genomic regions (to −390) are shown. The transcriptional start site is at position +1 (and horizontal arrowhead) as determined by S1 protection (Figure 5A) and primer extension (data not shown); the synthetic oligonucleotide (21-mer) used for primer extension extended from position +40 to +61, inclusive. The nucleotides are numbered at the left margin, and the deduced amino acids at the right margin. The positions of the introns are indicated by arrowheads. Note the use of two different acceptor sites at the 5′ end of the second intron, which give rise alternatively to a protein product with six additional amino acids (insert below major spliced form, see also Figure 5B,C and Figure 6). The extent of the longest cDNA isolated, c241, is indicated. Sequence upstream of the c241 clone has been determined on genomic DNA, the other parts have been sequenced on c241 and genomic clones. The cap site consensus for insect genes (Hultmark *et al.*, 1986; ATCAG/TTC/T) and the polyadenylation signals are underlined. The homeo domain is boxed. The sequence downstream of the *Xba*I site at position 2645 (see also Figure 6A) was only determined on the cDNA clone.

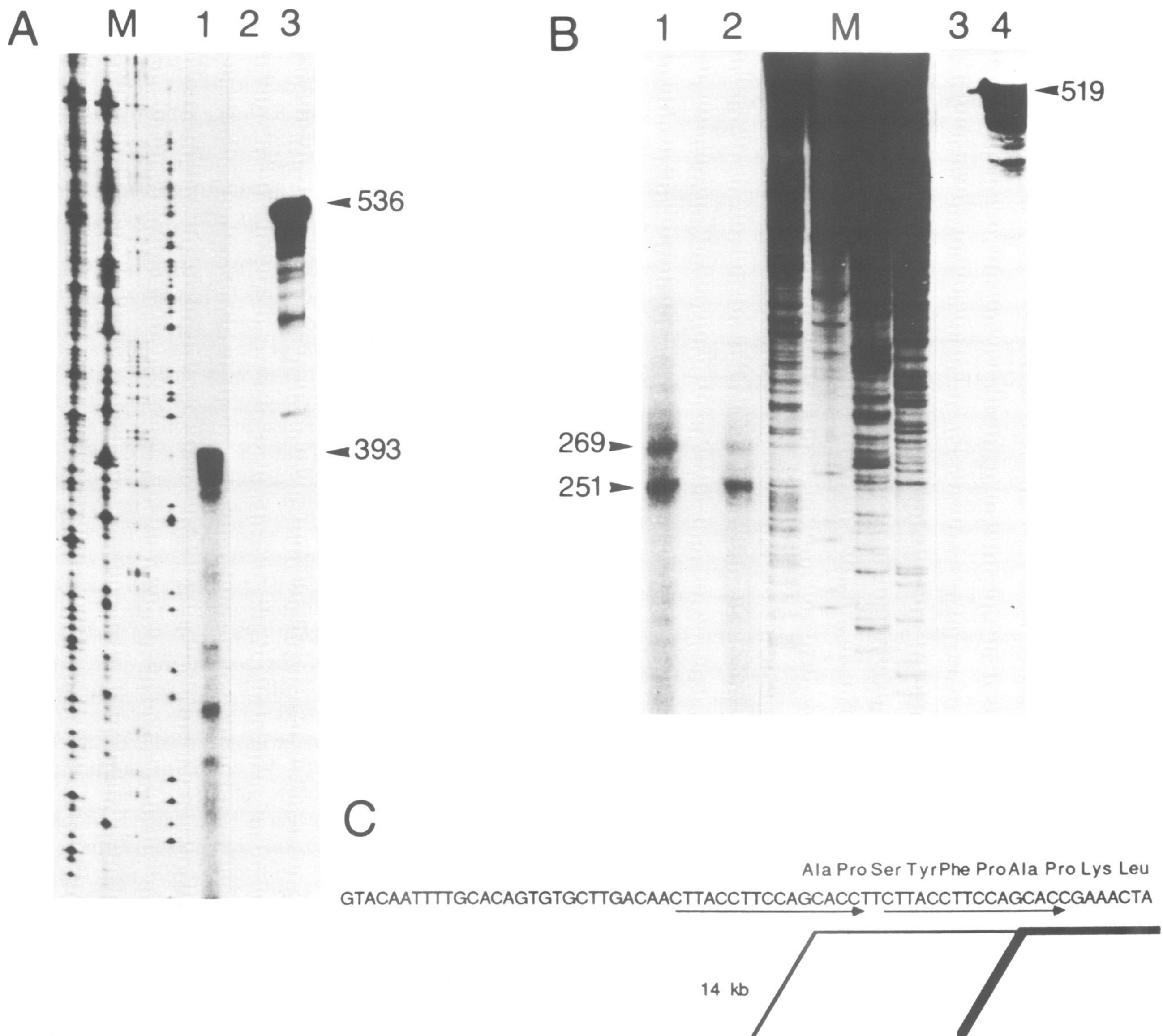

Fig. 5. Structural analysis of the F24 gene. (**A**) Mapping of the transcription initiation site by S1 protection analysis. The *Pst*I fragment containing the 5′ end (see Figure 6) of the F24 transcribed sequence was subcloned in an M13 vector and the complementary non-coding strand was generated (**lane 3**, 536 nucleotides including polylinker and primer from M13 vector). **Lane 1** shows the protected product (393 nucleotides) with 200 μg of total RNA from 12–18 h old embryos. **Lane 2**: control protection to S1 nuclease with 200 μg yeast RNA. No specific protection product is observed in the control experiment. The determined transcription initiation site was confirmed by primer extension analysis using a synthetic oligonucleotide (data not shown, see Results). (**B**) Mapping the two splice acceptor sites after the first intron by S1 protection experiments. The *Bam*HI/*Bgl*II fragment containing the appropriate genomic region (see Figure 6) was subcloned in an M13 vector; the generated single-stranded fragment was 519 nucleotides long (including polylinker and primer, see **lane 14**). **Lanes 1** and **2** show protected fragments of 251 and 269 nucleotides after hybridization to 10 μg poly(A)$^+$ or 200 μg total RNA respectively. Both RNA samples were isolated from 12–18 h old embryos. Note that the fragment of 251 nucleotides is ~5- to 10-fold higher in intensity than the 269 fragment. **Lane 3:** control experiment with 200 μg yeast RNA (no specific protection). (**C**) Sequence and summary of the two differential splice acceptor sites. The two direct repeats of 16 nucleotides are indicated by arrows. The proposed different splice acceptors are shown by lines, the thickness of which indicates the relative abundance (see panel **B**). The deduced amino acid sequence is given above the nucleotide sequence.

the same splice acceptor site (see Figure 5C). However, all of the isolated cDNA clones are spliced at the second acceptor site. To investigate the possibility of differential splicing, we performed S1 protection analysis with this genomic region. The opposite strand of the *Bgl*II–*Bam*HI fragment encompassing the acceptor sites (see Figure 6A) was generated in M13 (final size was 519 nucleotides including M13 primer and polylinker sequences). S1 analysis with this DNA piece revealed two protected fragments of 251 and 269 nucleotides (see Figure 5B). This result shows that both potential splice acceptor sites are being used. The additional six amino acids are shown in Figure 4 (below intron sign) and in Figure 5C. Nevertheless, we cannot exclude the possibility of a polymorphism of the *Oregon R*-strain used in this experiment, since it is not isogenic.

Features of the proposed protein sequence

The deduced amino acid sequence shown in Figures 4 and 6B summarizes some features of the F24 protein. The F24 homeo domain is split by an intron after position 44, in

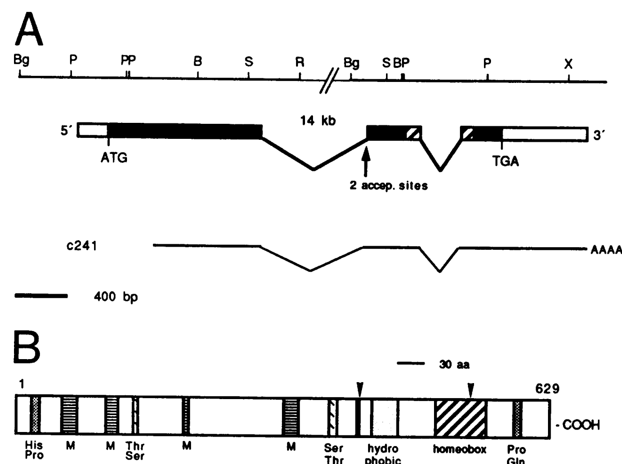

Fig. 6. Summary of structural features of the F24 transcript and the deduced protein. (**A**) The restriction map of the sequenced 5′ and 3′ regions is shown. The exons are indicated by open (untranslated) or black (coding) boxes. The homeobox is hatched. Note that there are two splice acceptor sites after the first intron. The extent of cDNA c241 is indicated. Restriction enzymes: B = *Bam*HI, Bg = *Bgl*II, R = *Eco*RI, P = *Pst*I, S = *Sma*I, X= *Xba*I. (**B**) Summary of the primary structure of the putative F24 protein. The homeo domain is hatched. Note that the 5′ exon contains four M-repeat regions (CAA/G-Gln stretches, indicated by M). Several stretches of two amino acids, e.g. His/Pro or Ser/Thr, and a hydrophobic region are indicated. Black bar at the first exon junction indicates the alternatively spliced six additional amino acids. The positions of the introns are shown by arrowheads.

contrast to *engrailed* and *invected* where the intron is located at position 17 (Poole *et al.*, 1985). Its homology to the *Antp* homeo domain is 67% overall (DNA level: 64%). Interestingly, the region with the highest conservation compared to *Antp* is not the putative recognition helix, proposed for DNA binding activity (position 42−50; Laughon and Scott, 1984; Shepherd *et al.*, 1984), but between positions 15 and 35 (95%, 20 out of 21). The F24 protein also contains four M-repeats (McGinnis *et al.*, 1984a) or *opa* sequences (Wharton *et al.*, 1985), all located in the 5′ exon (see Figure 6B). In addition, several stretches rich in His/Pro, Ser/Thr or Pro/Gln are found in the amino acid sequence as indicated in Figure 6B. A hydrophobic stretch of ~30 amino acids is located in the second exon.

During the course of our experiments we found that the fourth M-repeat sequence (see Figures 4 and 6B) is highly conserved in other genes of *Drosophila* and the honey bee. Low stringency Southern blots hybridized with the *Bam*HI−*Eco*RI fragment at the end of the 5′ exon revealed ~10−15 bands in the *Drosophila* genome and ~7 bands in the honey bee (M.Mlodzik, unpublished results), and not the typical smear usually observed with other M-repeats, e.g. that of the *Antp* gene (McGinnis *et al.*, 1984a). Using the same fragment we have screened a *Drosophila* genomic library and isolated several independent clones. One of these corresponds to the zygotic dorso-ventral polarity gene *twist* (*twi*), which was unambiguously identified by *in situ* hybridization to chromosomes and to tissue sections, and by comparison with restriction maps of previously published *twi* clones (Thisse *et al.*, 1987). The homologous DNA sequence lies within the fourth M-repeat (nucleotides 1176-1211 in Figure 4) and is strongly conserved in the same reading frame (33 out of 35 nucleotides, 11 out of 12 amino acids) in *twi* (M.Mlodzik, unpublished results).

Spatial expression during blastoderm and germ band extension

In the cellular blastoderm (~3 h after fertilization), transcripts from the F24 gene start to accumulate anteriorly in the cells between 74 and 83% egg length (EL, from the posterior pole), and also in the most posterior somatic cells (Figure 7A, the probes used in these experiments are discussed in Materials and methods). No transcripts are detectable in the syncytial blastoderm or at earlier stages. The anterior domain of expression appears to be located just anterior to the *Deformed* (*Dfd*) stripe of expression (Chadwick and McGinnis, 1987; Martinez-Arias *et al.*, 1987). This is also apparent at early gastrulation when *Dfd* expression lies in the region of the cephalic furrow, whereas F24 is expressed just anterior to the cephalic furrow (Figure 7B). However, cells expressing the F24 gene do not form a circumferential stripe around the blastoderm embryo, as is the case for *Dfd* (Chadwick and McGinnis, 1987; Martinez-Arias *et al.*, 1987) and other homeotic genes. The F24 expression is restricted to ventro-lateral regions on both sides of the blastoderm embryo. Expression is not found in the mesoderm, the anterior endoderm anlagen (ventral) or the dorsal tissue (Figure 8A,B). Comparing the domains of F24 expression to the blastoderm fate map of the head region, (Jürgens *et al.*, 1986), F24 is detected in the anlagen for the anterior part of the mandibular lobe and the hypopharyngeal organ (also called 'intercalary element'), and probably also in lateral parts of the procephalic lobe.

After gastrulation, when the mesodermal and endodermal primordia have invaginated the ectoderm and underlying mesoderm start the process of germ band extension (Sonnenblick, 1950). The germ band extends around the posterior of the embryo on the dorsal surface. This movement is completed at ~6 h after fertilization (Campos-Ortega and Hartenstein, 1985). During this process F24 transcripts are found in the cells just anterior to the cephalic furrow (but not in the mesoderm) and in parts of the invaginating posterior midgut rudiment. When the germ band is fully extended (~6 h after fertilization) the first signs of segmentation appear, and grooves have become obvious along the germ band of the embryo. In the head region the gnathal lobes start to form posterior to the stomodeum (Campos-Ortega and Hartenstein, 1985). At this stage, ectoderm cells posterior to the stomodeum but anterior to the cephalic furrow show F24 expression (Figure 7C,D). They belong most likely to the hypopharyngeal organ and anterior parts of the mandibular lobe. Parts of the posterior midgut invagination also exhibit constant F24 expression throughout these embryonic stages (Figure 7C).

F24 expression during later embryogenesis

At 8 h post-fertilization, the germ band begins to retract around the posterior pole to position all segmental primordia along the ventral surface. The segmental grooves and the anterior gnathal buds now become obvious landmarks in the developing embryo (Campos-Ortega and Hartenstein, 1985). Head involution starts during germ band retraction, and the visible gnathal lobes migrate anteriorly to border and join the larval mouth opening (Campos-Ortega and Hartenstein, 1985; Jürgens *et al.*, 1986). At this stage it is difficult to locate landmarks in the developing head, since the extensive morphogenetic movements largely reorganize the embryonic cephalon. The hypopharyngeal organ and parts of the

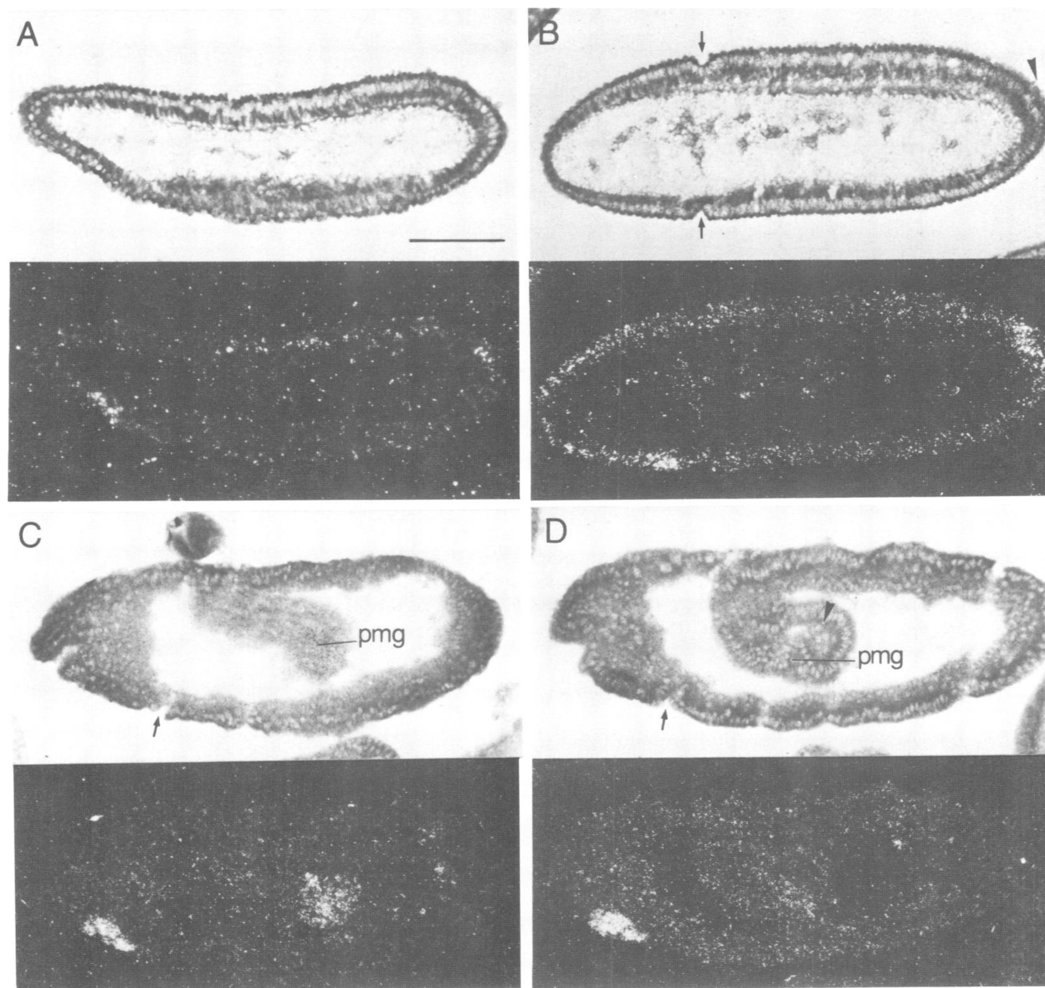

Fig. 7. Distribution of F24 transcripts during the cellular blastoderm through germ band extension stage. The anterior end of the embryos points always to the left, and dorsal, where visible, faces upwards. The sections were hybridized to tritiated c241 (see Figure 6) probe. (**A**) Brightfield and darkfield images of a completed cellular blastoderm embryo at the onset of gastrulation. Note the hybridization signal at ~75–83% EL (from the posterior pole) and in some of the very posterior somatic cells. (**B**) Brightfield and darkfield images of an embryo after beginning of gastrulation. Note hybridization signals anterior to the cephalic furrow (75–83% EL) and in somatic cells at the posterior pole under the pole cells. The cephalic furrow is indicated with an arrow, the pole cells by an arrowhead. Note that the anterior region of expression is not a circumferential stripe around the entire embryo. Expression is restricted to ventrolateral domains on both sides of the embryo, and is not found in the mesoderm primordia (see also panels **C** and **D**). (**C**) and (**D**) Brightfield and darkfield images of sagittal section through extended germ band embryos. Hybridization is detected anterior to the cephalic furrow (arrows), but posterior to the stomodeal lip. The expressing cells most likely belong to the anterior region of the mandibular lobe and the hypopharyngeal organ. The posterior domain of F24 expression can be detected in parts of the posterior midgut rudiment (pmg). The section in (**D**) does not pass through the region of expression in the pmg. Note that the anterior expression in parts of the head is restricted to the ectoderm; no F24 mRNA has been detected in mesodermal or anterior endodermal tissue. Pole cells are indicated by an arrowhead. The horizontal bar in (**A**) represents 0.1 mm.

mandibular and maxillary lobes move into the oral cavity to form parts of the pharynx. Most of the cells that express F24 take part in these movements and migrate into the stomodeal opening. However, some cells that express F24 transcripts stay at the surface (arrow in Figure 8A). These cells might correspond to an anterior sensory organ. In addition, parts of the condensing central nervous system, a portion of the supraoesphageal ganglion, show an accumulation of F24 transcripts (data not shown; also described by Hoey *et al.*, 1986).

The expression in regions of the developing midgut persists throughout the remainder of embryogenesis (arrowheads in Figure 8A,B). However, during and after germ band retraction (see above) the most posterior cells of the posterior midgut primordium (pmg) become localized in the middle region of the midgut. Therefore, F24-expressing midgut cells are localized in middle regions of the embryo (Figure 8A,B), anterior to the posterior midgut cells that express *caudal* (Mlodzik *et al.*, 1985, 1987; Macdonald and Struhl, 1986; Mlodzik and Gehring, 1987). It appears as if pmg cells expressing F24 are involved in meeting the cells of the anterior midgut and in mediating the fusion of the two endodermal primordia (Figure 8A). Later in development, the midgut completely surrounds the yolk sac, and during stages 15 and 16 three constrictions appear consecutively in the tube-like wall of the midgut (Campos-Ortega and Hartenstein, 1985). F24 is expressed in cells in the middle of the tube-like midgut and later, during and after the constrictions, in the middle region of the developing gut system (see Figure 8A,B).

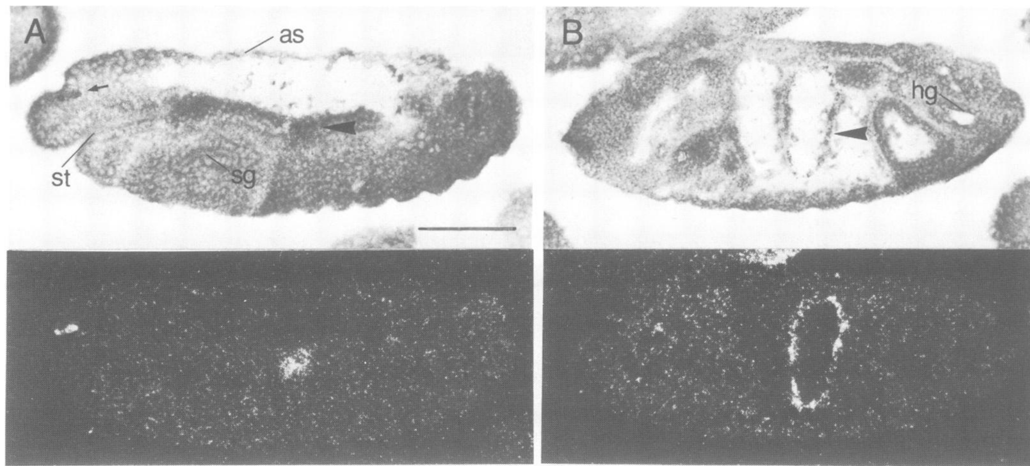

Fig. 8. Expression of the F24 gene during later embryogenesis. The anterior end of the embryos is left, in (**A**) dorsal is up. (**A**) Bright- and darkfield images of a sagittal section through a stage 13 embryo (after Campos-Ortega and Hartenstein, 1985). Note the hybridization in the middle midgut region (arrowhead) and in a few cells dorsal to the stomodeal invagination (small arrow). These cells might originate from parts of the procephalic lobe, in the anlage of which F24 is expressed during the blastoderm and the gastrula (stages 5 and 6). as, aminoserosa; sg, salivary gland; st, stomodeum. (**B**) Bright- and darkfield images of a slightly oblique horizontal section of a stage 16 embryo. Hybridization signal in the midgut region is indicated by an arrowhead. hg, hindgut. The horizontal bar in (**A**) represents 0.1 mm.

Discussion

To find a potential mutant candidate for the F24 gene within the ANT-C, we have analyzed our cloned segment at 84A with respect to several deletions that define all or parts of the ANT-C (Kaufman, 1983; Kaufman and Abbott, 1984). This analysis (see Results) demonstrates that the F24 gene most likely corresponds to *labial* (*lab*), a gene in the most proximal region of the ANT-C that is involved in head development (Kaufman, 1983). Figure 9 summarizes a genetic deletion analysis of the ANT-C (Kaufman, 1983) and indicates how the cloned gene behaves with respect to these deletions (arrowhead in Figure 9). We can exclude that F24 corresponds to another candidate, the homeotic head gene *proboscipedia* (*pb*, Kaufman, 1983). However, a detailed phenotypic description for *lab* is not available and therefore it remains to be seen how the spatial expression patterns correlate with the *lab* phenotype(s). The F24 (*lab*) gene appears to be identical to F90-2 (Hoey *et al.*, 1986). The sequences of the homeoboxes and the restriction maps of the F90-2 and F24 clone are identical.

Concerning its chromosomal location and structure, the F24 gene behaves similarly to most other homeotic genes. It is located within one of the two homeotic gene clusters, the ANT-C, and it contains a homeobox, which is located in the 3′ part of the gene, as in most other homeobox genes of the ANT-C and the BX-C. Its location in the proximal part of the ANT-C fits to the general observation that the position of genes within the complex is related to the order of their expression along the antero-posterior axis; that is *lab* (F24) is positioned proximal to and expressed anterior to *Dfd; Dfd* is proximal to and expressed anteriorly to *Scr* (Martinez-Arias *et al.*, 1987) and so on. This seems to be the general rule, except that within the ANT-C some genes containing a homeobox, but having rather different functions, like *bcd* (Frohnhöfer and Nüsslein-Volhard, 1986; Frigerio *et al.*, 1986), *zen* (Wakimoto *et al.*, 1984; Doyle *et al.*, 1986) or *ftz* (Wakimoto *et al.*, 1984; Hafen *et al.*, 1984) are interspersed between the homeotic genes. However, *lab*

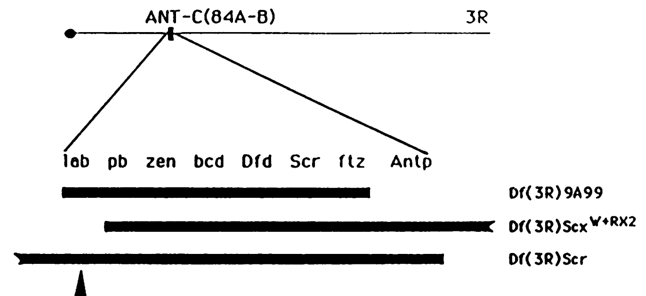

Fig. 9. Summary of the deletion analysis. The right arm of the third chromosome (3R) is schematically presented. The ANT-C region is enlarged and genes identified by mutagenesis are indicated. The extent of the deficiencies is shown. The cloned F24 segment at 84A behaves as shown by the arrowhead, which is identical to the *labial* gene (*lab*). Other abbreviations: pb, *proboscipedia*; zen, *zerknüllt* (the *zen* region contains two homeobox transcriptional units); bcd, *bicoid*; Dfd, *Deformed*; Scr, *Sex combs reduced*; ftz, *fushi tarazu*; Antp, *Antennapedia*.

(F24) is expressed also at the very posterior pole. The *lab* (F24) gene itself is similarly organized to the other analyzed homeotic genes of the two clusters, except that it contains an intron within the proposed DNA-contacting helix of the homeo domain. This has not previously been observed in any other homeobox gene. The 5′ part of the protein shares four regions of M-repeat homology (McGinnis *et al.*. 1984a). One of these M-repeats (the fourth) appears to be conserved at other genomic loci of developmental interest (e.g. *twist*, see Results). The presence of the two protein species due to the two splice acceptor sites opens up the possibility that the two variants have slightly different function(s), although polymorphism of the analyzed wild-type strain has not been excluded. A similar situation has been found in *Antp* (Schneuwly *et al.*, 1986; Stroher *et al.*, 1986; Laughon *et al.*, 1986).

Spatial expression

There are obvious differences in the expression patterns of

F24 (*lab*) compared to other homeotic genes of the ANT-C or the BX-C. Notably, F24 (*lab*) is not expressed in a circumferential stripe around the blastoderm embryo, unlike for all other analyzed homeotic genes (Levine *et al.*, 1983; Akam and Martinez-Arias, 1985; Chadwick and McGinnis, 1987; Martinez-Arias *et al.*, 1987). F24 is expressed in lateral parts of the blastoderm, where the ectodermal and neuro-ectodermal anlagen are localized, and not in all the germ layers, as is the case for other homeotic genes. In addition, F24 (*lab*) has two domains of expression, one in the anterior ectodermal anlagen and the second in the posterior endoderm. Only one other homeobox gene, *caudal* (*cad*), is also expressed in endodermal tissue, in that case in the posterior midgut (Mlodzik *et al.*, 1985, 1987; Macdonald and Struhl, 1986). However, the *lab* expression in endodermal primordia starts as early as at the cellular blastoderm stage (Figure 7A,B) and not during germ band extension as is the case for *cad*. The abundance of F24 (*lab*) transcripts appears to be low as compared to other transcripts such as *Dfd, cad* or *Antp*, since exposure times of hybridized tissue sections were generally substantially longer with probes of similar specific activity.

The anterior domain of F24 (*lab*) is located anterior to *Dfd* (Chadwick and McGinnis, 1987; Martinez-Arias *et al.*, 1987) during the cellular blastoderm, the gastrula stages and in later embryogenesis. Both anterior and posterior regions of expression are either anterior or posterior respectively to the corresponding domains of *spalt* (*sal*) expression (Frei *et al.*, 1988). *Sal* is an unusual homeotic gene as it affects posterior and anterior structures, and is mainly expressed outside the anlagen defined by the trunk namely outside the thoracic and abdominal segments (Jürgens, 1988; Frei *et al.*, 1988). Interestingly, during stages 9 and 10 (after Campos-Ortega and Hartenstein, 1985) the ectodermal domain of *Dfd* expression is transiently extended towards the hypopharyngeal organ and the anterior mandibular lobe and therefore overlaps with the expression of *lab* (F24) during these stages. It remains to be seen how the two proteins might interact in establishing the correct identity of these developing head parts.

The region of *lab* (F24) expression during early embryogenesis (stage 8) is most likely bordered by parasegmental bounderies in the parasegmental frame (Martinez-Arias and Lawrence, 1985). It is always localized just anterior to the area containing *Dfd*, transcripts which seem to be expressed at these stages in parasegment 1 (Martinez-Arias *et al.*, 1987). Since the width of both regions along the antero-posterior axis is about the same (5–6 cells at the cellular blastoderm), *lab* (F24) is most likely expressed in parasegment O. However, in later embryogenesis when the gnathal buds become obvious, genes like *Dfd* and *Scr* are expressed in the epidermis in a segmental and not a parasegmental frame (Chadwick and McGinnis, 1987; Martinez-Arias *et al.*, 1987). It may be that *lab* is expressed in the same register at this stage as well. Clarification of the overlapping region of expression between *lab* and *Dfd* (see above), and the definitive elucidation of the (parasegmental versus segmental) frame of expression will require double labelling techniques for the independent detection of *lab* and *Dfd* products. It also remains to be determined whether a phenotype or function for the midgut expression can be found.

Materials and methods

DNA and RNA methods

DNA from λ phages was isolated as described by Garber *et al.* (1983). Plasmid DNA was prepared using the alkaline extraction procedure as described by Frei *et al.* (1985). The original homeobox-containing phage λF24 was isolated from the genomic library of Maniatis *et al.* (1978). All other genomic phages were isolated from a library prepared in the λEMBL4 vector from an Oregon-R strain (kindly provided by U.Walldorf). cDNA clones were isolated from a 3–12 h embryonic library (kindly provided by L.Kauvar; Poole *et al.*, 1985). Genomic or cDNA fragments isolated from phage DNA were usually subcloned into pGem-vectors (Promega-Biotech) or Bluescript-vectors (Stratagene). Embryonic and larval RNA was isolated following the guanidium isothiocyanate method as described by Chirgwin *et al.* (1979).

Southern and Northern blot analysis

Southern blots were prepared as described by Southern (1975). The blots were hybridized and washed at the desired stringency as described by McGinnis *et al.* (1984a). Poly(A)$^+$ RNAs, 2 μg for all stages investigated, were fractionated by electrophoresis on a 1% agarose–formaldehyde gel. The RNA was transferred to nitrocellulose filters by the Southern method (1975). Probes were labelled with standard nick-translation techniques using ^{32}P-labelled nucleotides. Filters were hybridized at 42°C and washed at 60–70°C in solutions described by McGinnis *et al.* (1984a). Similar loadings of poly(A)$^+$ RNA per lane were checked by rehybridization of the blots with a *Drosophila* elongation factor probe (gift of Uwe Walldorf).

DNA sequencing

The DNAs were sequenced by the dideoxy nucleotide method of Sanger *et al.* (1977). All DNA sequences were analyzed at least in two independent clones on each strand. The cDNA clone c241 was subcloned in both orientations into Bluescript + with the KS-polylinker (Stratagene). Overlapping deletions were generated using the exonuclease III–mung bean nuclease method as described by the supplier (Stratagene). Genomic fragments were subcloned into appropriately cut M13 mp18 and mp19 vectors.

S1 mapping and primer extension experiments

For protected fragments of 400 nucleotides or more, S1 mapping was carried out as described by Kuroiwa *et al.* (1984). For precision S1 mapping on sequencing gels, the non-coding single-stranded DNA fragment was generated in M13 vectors according to the procedure of Burke (1984). Hybridization was performed in 1 M NaCl, at 60°C for 1–3 h. S1 digestions were done as described by the supplier (Boehringer Mannheim), and protected fragments were separated on 6% polyacrylamide sequencing gels. Primer extension experiments were performed with a ^{32}P-end-labelled synthetic oligonucleotide (21-mer) as described previously (Mlodzik and Gehring, 1987).

In situ hybridization techniques

In situ hybridization to polytene chromosomes was carried out with biotinylated probes. Materials used for this procedure were bio-dUTP (11 carbon linker arm) mainly synthesized as described by Langer *et al.* (1981). After hybridization the labelled DNA was detected, according to the supplier's instructions, with Detek-I-hrp (Enzo-Biochemicals), a peroxidase-conjugated avidin complex.

In situ hybridizations to embryonic tissue sections were performed essentially as described by Hafen *et al.* (1983). The specific activity of nick-translated fragments was $\sim 2 \times 10^8$ d.p.m./μg. Mean single-stranded DNA fragment length was 40–120 nucleotides. Young embryos, 0–6 h old, were pre-fixed, demembranated and post-fixed as described by Akam and Martinez-Arias (1985). The autoradiographic exposure was 28–35 days. Embryos are staged after Campos-Ortega and Hartenstein (1985).

The probes used for these experiments were either the cDNA clone c241, or a genomic homeobox probe. Both probes detect the same patterns throughout embryogenesis. Therefore, detection of cross-hybridization to the highly conserved M-repeat region in other genes, e.g. *twi* (see Results), can be excluded. These probes also detected a single transcript on Northern blots (see Results), so that both regions of expression must be from the same gene.

Acknowledgements

We are grateful to Michael Levine for sharing unpublished results, Larry Kauvar for providing cDNA library aliquots and Bill McGinnis for initial

discussions. We thank Paul Baumgartner and Roger Aeschbacher for help with establishing the genomic homeobox sequence and Greg Gibson and Clive Wilson for comments on the manuscript. Our work benefited greatly from discussions with and help of Uwe Walldorf and Ursula Weber. Finally, our appreciation goes to Daniel Wey for help with the photographs, Andreas Fritz for help with the arrangement for Figure 4 and Erika Wenger-Marquardt for efficient preparation of the manuscript. This work was supported by grants from the Swiss National Science Foundation and the Kantons Basel. A.F. was supported by a fellowship from the Norwegian Research Council NAVF.

References

Akam,M.E. (1987) *Development*, **101**, 1–22.

Akam,M.E. and Martinez-Arias,A. (1985) *EMBO J.*, **4**, 1689–1700.

Anderson,K. (1987) *Trends Genet.*, **3**, 91–97.

Beachy,P.A., Helfand,S. and Hogness,D.S. (1985) *Nature*, **313**, 545–551.

Bender,W., Akam,M.E., Karch,F., Beachy,P.A., Peifer,M., Spierer,P., Lewis,E.B. and Hogness,D.S. (1983) *Science*, **221**, 23–29.

Burke,J.F. (1984) *Gene*, **30**, 63–68.

Campos-Ortega,J.A. and Hartenstein,V. (1985) *The Embryonic Development of Drosophila melanogaster*. Springer Verlag, Berlin.

Carroll,S.B. and Scott,M.P. (1985) *Cell*, **43**, 47–57.

Carroll,S.B., Laymon,R.A., McCutcheon,M.A., Riley,P.D. and Scott,M.P. (1986) *Cell*, **47**, 113–122.

Chadwick,R. and McGinnis,W. (1987) *EMBO J.*, **6**, 779–789.

Chirgwin,J.M., Przybyla,A.E., McDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.

DiNardo,S., Kuner,J.M., Theis,T. and Levine,M. (1985) *Cell*, **43**, 59–69.

Doyle,H.J., Harding,K., Hoey,T. and Levine,M. (1986) *Nature*, **323**, 76–79.

Fjose,A., McGinnis,W.J. and Gehring,W.J. (1985) *Nature*, **313**, 284–289.

Frasch,M., Hoey,T., Rushlow,C., Doyle,H. and Levine,M. (1987) *EMBO J.*, **6**, 749–759.

Frei,E., Baumgartner,S., Edström,J.-E. and Noll,M. (1985) *EMBO J.*, **4**, 979–987.

Frei,E., Schuh,R., Baumgartner,S., Burri,M., Noll,M., Jürgens,G., Seifert,E., Nanber,U. and Jäckle,H. (1988) *EMBO J.*, **7**, 197–204.

Frigerio,G., Burri,M., Bopp,D., Baumgartner,S. and Noll,M. (1986) *Cell*, **47**, 735–746.

Frohnhöfer,H.G. and Nüsslein-Volhard,C. (1986) *Nature*, **324**, 120–125.

Garber,R.L., Kuroiwa,A. and Gehring,W.J. (1983) *EMBO J.*, **2**, 2027–2036.

Garcia-Bellido,A. (1975) *Ciba Found. Symp.*, **29**, 161–182.

Gehring,W.J. (1987) *Molecular Approaches to Developmental Biology*. Alan R.Liss, NY, pp. 115–129.

Gehring,W.J. and Hiromi,Y. (1986) *Annu. Rev. Genet.*, **20**, 147–173.

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.

Harding,K., Rushlow,C., Doyle,H.J., Hoey,T. and Levine,M. (1986) *Science*, **233**, 953–959.

Hoey,T., Doyle,H.J., Harding,K., Wedeen,C. and Levine,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4809–4813.

Hultmark,D., Klemenz,R. and Gehring,W.J. (1986) *Cell*, **42**, 237–247.

Jürgens,G. (1988) *EMBO J.*, **7**, 189–196.

Jürgens,G., Wieschaus,E., Nüsslein-Volhard,C. and Kluding,H. (1984) *Wilhelm Roux's Arch. Dev. Biol.*, **193**, 283–295.

Jürgens,G., Lehmann,R., Schardin,M. and Nüsslein-Volhard,C. (1986) *Wilhelm Roux's Arch. Dev. Biol.*, **195**, 359–377.

Kaufman,T.C. (1983) In Raff,R.A. and Jefferys,W. (eds), *Time, Space and Pattern in Embryonic Development*. Alan Liss, NY, pp. 365–383.

Kaufman,T.C. and Abbott,M.K. (1984) In Malacinsky,G.M. and Klein,W.H. (eds), *Molecular Aspects of Development*. Plenum Press, NY, pp. 182–218.

Kuroiwa,A., Kloter,U., Baumgartner,P. and Gehring,W.J. (1985) *EMBO J.*, **4**, 3757–3764.

Langer,P.R., Waldrop,A.A. and Ward,D.C. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6633–6637.

Laughon,A. and Scott,M.P. (1984) *Nature*, **308**, 25–31.

Laughon,A., Boulet,A.M., Bermingham,J.R., Laymon,R.A. and Scott,M.P. (1986) *Mol. Cell. Biol.*, **6**, 4676–4689.

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.

Macdonald,P.M. and Struhl,G. (1986) *Nature*, **324**, 537–545.

Macdonald,P.M., Ingham,P. and Struhl,G. (1986) *Cell*, **47**, 721–734.

Mahaffey,J.W. and Kaufman,T.C. (1987) *Genetics*, **117**, 51–60.

Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,D.K. and Efstratiadis,A. (1978) *Cell*, **15**, 687–701.

Martinez-Arias,A. and Lawrence,P.A. (1985) *Nature*, **313**, 639–642.

Martinez-Arias,A., Ingham,P.W., Scott,M.P. and Akam,M.E. (1987) *Development*, **100**, 673–683.

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.

Mlodzik,M. and Gehring,W.J. (1987) *Cell*, **48**, 465–478.

Mlodzik,M., Fjose,A. and Gehring,W.J. (1985) *EMBO J.*, **4**, 2961–2969.

Mlodzik,M., De Montrion,C.M., Hiromi,Y., Krause,H.M. and Gehring,W.J. (1987) *Gene Dev.*, **1**, 603–614.

Nüsslein-Volhard,C. and Wieschaus,E. (1980) *Nature*, **287**, 795–801.

Nüsslein-Volhard,C., Frohnhöfer,H.G. and Lehmann,R. (1987) *Science*, **238**, 1675–1681.

Poole,S., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.

Regulski,M., McGinnis,N., Chadwick,R. and McGinnis,W. (1987) *EMBO J.*, **6**, 767–777.

Riley,P.D., Carroll,S.B. and Scott,M.P. (1987) *Genes Dev.*, **1**, 716–730.

Rushlow,C., Frasch,M., Doyle,H. and Levine,M. (1987) *Nature*, **330**, 583–586.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Schneuwly,S., Kuroiwa,A., Baumgartner,P. and Gehring,W.J. (1985) *EMBO J.*, **5**, 733–739.

Scott,M.P., Weiner,A.J., Hazelrigg,T.J., Polisky,B.A., Pirrotta,V., Scalenghe,F. and Kaufman,T.C. (1983) *Cell*, **35**, 763–776.

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.

Stroher,V.L., Jorgensen,E.M. and Garber,R.L. (1986) *Mol. Cell. Biol.*, **6**, 4667–4675.

Thisse,B., El Massal,M. and Perrin-Schmitt,F. (1987) *Nucleic Acids Res.*, **15**, 3439–3453.

Wakimoto,B.T., Turner,F.R. and Kaufman,T.C. (1984) *Dev. Biol.*, **102**, 147–172.

Wharton,K.A., Yedvobnick,B., Finnerty,V.G. and Artavanis-Tsakonas,S. (1985) *Cell*, **40**, 55–62.

White,R.A.M. and Wilcox,M. (1984) *Cell*, **39**, 163–171.

Wirz,J., Fessler,L.I. and Gehring,W.J. (1986) *EMBO J.*, **5**, 3327–3334.

Received on April 18, 1988

Note added in proof

There is 85% sequence identity between the homeodomains encoded by *labial* and the mouse *Hox-1.6* gene (Baron,A., Featherstone,M.S., Hill,R.E., Hall,A., Galliot,B. and Duboule,D. (1987) *EMBO J.*, **6**, 2977–2986) indicating that these two domains are closely related in evolution, which has also been noticed by others (D.Duboule, in press). The sequence homology also extends a short distance 3′ to the homeobox. In addition, 10 out of 18 amino acids around the conserved pentapeptides of *labial* (amino acids 386 to 403) and *Hox-1.6* are identical. However, the N-terminal sequences of *Hox-1.6* are not yet known.