The EMBO Journal vol.14 no.21 pp.5306–5317, 1995

The *Drosophila fork head* domain protein crocodile is required for the establishment of head structures

Udo Häcker[1], Eckhard Kaufmann[2], Christine Hartmann[1], Gerd Jürgens[3], Walter Knöchel[2] and Herbert Jäckle[1,4]

[1]Max-Planck-Institut für Biophysikalische Chemie, Abteilung Molekulare Entwicklungsbiologie, Postfach 2841, D-37018 Göttingen, [2]Abteilung Biochemie, Universität Ulm, Albert-Einstein-Allee 11, D-89081 Ulm and [3]Lehrstuhl für Entwicklungsgenetik, Universität Tübingen, Spemannstrasse 37-39, D-72076 Tübingen, Germany

[4]Corresponding author

The *fork head* (*fkh*) domain defines the DNA-binding region of a family of transcription factors which has been implicated in regulating cell fate decisions across species lines. We have cloned and molecularly characterized the *crocodile* (*croc*) gene which encodes a new family member from *Drosophila*. *croc* is expressed in the head anlagen of the blastoderm embryo under the control of the anterior, the dorsoventral and the terminal maternal organizer systems. The *croc* mutant phenotype indicates that the *croc* wild-type gene is required to function as an early patterning gene in the anterior-most blastoderm head segment anlage and for the establishment of a specific head skeletal structure that derives from the non-adjacent intercalary segment at a later stage of embryogenesis. As an early patterning gene, *croc* exerts unusual properties which do not allow it to be grouped among the established segmentation genes. A single-site mutation within the *croc fkh* domain, which causes a replacement of the first out of four conserved amino acid residues thought to be involved in the coordinate binding of Mg^{2+}, abolishes the DNA binding of the protein *in vitro*. In view of the resulting lack-of-function mutant phenotype, it appears likely that metal binding by the affected region of the *fkh* domain is crucial for proper folding of the DNA-binding structure.

Keywords: *crocodile* gene/DNA binding/*Drosophila* fork head/head structures/protein folding

Introduction

The successful execution of the genetic program largely depends on the coordinate regulation of gene expression by mechanisms that control transcription precisely in time, space and level. In eukaryotes, this regulation operates through distinct interactions between transcriptional factors bound to *cis*-acting enhancers and the basal components of the transcription machinery (Lewin, 1990; Roeder, 1991; Gill and Tjian, 1992; Tjian and Maniatis, 1994). Genetics combined with the molecular analysis of early pattern formation during *Drosophila* embryogenesis revealed a rich body of functionally defined transcription factors that coordinate the genetic program underlying the biological pattern-forming processes. Most of the identified factors represent integral components of the segmentation gene cascade in the preblastoderm embryo, and they also appear to be required for tissue organization and organogenesis at later stages of embryogenesis (Akam, 1987; Ingham, 1988; St Johnston and Nüsslein-Volhard, 1992; Hoch and Jäckle, 1993; Pankratz and Jäckle, 1993). The molecular analysis of the *Drosophila* segmentation process has also led to the discovery of the homeodomain (Gehring, 1987; Scott *et al.*, 1989).

The discovery of the homeodomain has opened the search for genes that share common structural DNA-binding domains in other organisms. The basic theme of this approach was continued with a limited number of distinct DNA-binding motifs such as the homeobox, the zinc finger, the helix–loop–helix motif and the paired domain (for a review see Johnson and McKnight, 1989), leading to what now appears to be a large and growing number of transcription factors that act in determinative events or cell-fate decisions across species lines. The most recently found evolutionarily conserved DNA-binding domain of some 110 amino acids, termed the HNF-3/*fork head* (*fkh*) DNA recognition motif or *fkh* domain (Weigel and Jäckle, 1990), emerged from a sequence comparison of the hepatocyte-specific transcription factor HNF-3α of rat (Lai *et al.*, 1990) and the region-specific homeotic gene *fkh* of *Drosophila* (Weigel *et al.*, 1989). X-ray crystallography of the DNA-bound *fkh* domain revealed that the DNA-binding protein fold resembles the structure of histone H5 (Clark *et al.*, 1993).

Members of the family of *fkh* domain proteins, such as HNF-3α,-β and -γ of mouse (Ruiz i Altaba *et al.*, 1993; Ang and Rossant, 1994; Weinstein *et al.*, 1994), XFKH1/XFD-1/*pintallavis* of *Xenopus* (Dirksen and Jamrich, 1992; Knöchel *et al.*, 1992; Ruiz i Altaba and Jessell, 1992), *whn* of rodents (Nehls *et al.*, 1994) and *axial* of zebrafish (Strähle *et al.*, 1993), have been implicated in development and oncogenesis (Li and Vogt, 1993). In *Drosophila*, only two out of the eight known members of the *fkh* domain transcription factors have been identified functionally on the basis of a mutant phenotype, the prototype gene *fkh* (Weigel *et al.*, 1989) and the *fkh* domain gene pair *sloppy paired* (*slp*) (Grossniklaus *et al.*, 1992, 1994). While *fkh* activity is essential for ectodermal structures of the fore- and hindgut (Weigel *et al.*, 1989), the *slp* genes *slp1* and *slp2* appear to combine the functions of the gap, pair rule and segment polarity classes of segmentation genes. *slp* functions as a gap-like gene in the prospective head region of the embryo and exerts segment polarity function in both the head and trunk anlage (Grossniklaus *et al.*, 1994).

Here we present a molecular analysis of another member of the *Drosophila fkh* domain gene family, previously termed FD1 (Häcker *et al.*, 1992), which is identical to the *crocodile* (*croc*) gene identified by mutant alleles. We

show that *croc* expression at the blastoderm stage is controlled by the maternal coordinate genes. *croc* activity is required for the formation of the anterior-most head segment primordium and for the establishment of a distinct part of the head skeleton that derives from a different segment primordium of the head.

Results

Structure and sequence of the FD1 gene

Previous studies have identified a family of *fkh* domain-encoding transcripts in *Drosophila*, among which the FD1 sequence was localized to position 78F on the left arm of the third chromosome (Häcker *et al.*, 1992). We isolated the corresponding genomic DNA region from λ-phage libraries and two full-size cDNA clones from different cDNA libraries prepared from poly(A)$^+$ RNA of 0–4 and 4–8 h old embryos, respectively (Figure 1a). The FD1 transcription unit encodes a single transcript, as revealed by Northern blot analysis (Figure 1b). Comparison of cDNA with genomic sequences (Figure 2a) and primer extension analysis (results not shown) revealed that the primary FD1 transcript contains a single exon which initiates at the first nucleotide present in the cDNA sequence (Figures 1a and 2a).

Genomic DNA fragments covering 9 kb upstream and 7 kb downstream of the coding sequence, respectively, were fused upstream of a reporter gene construct containing the *Drosophila* hsp43 core promoter in front of the bacterial *lacZ* gene (Thummel and Pirrotta, 1992). Embryos containing the respective FD1 DNA fragments acting as enhancer elements in front of the reporter gene construct show all spatial aspects of the FD1 gene expression patterns during early and late stages of embryogenesis (see below). The different patterns of expression conducted by the 9 kb upstream and 7 kb downstream DNA fragments (Figure 1c) indicate that enhancer elements necessary for FD1 expression are separated by the transcribed region. The structure of the FD1 gene is summarized in Figure 1.

The FD1 transcript contains an open reading frame which codes for a putative 54 kDa protein (Figure 2a). Sequence comparison with known proteins revealed a single diagnostic protein motif, the *fkh* domain, which is located in the N-terminal region of the protein. The *fkh* domain of FD1 shows a higher degree of sequence identity to the *fkh* domains of FKH1 and MFH-1 of mouse (Kaestner *et al.*, 1993; Miura *et al.*, 1993) and XFD-4 of *Xenopus* (Knöchel *et al.*, 1992; Scheucher *et al.*, 1995) compared with the *fkh* domains of *fkh* and HNF-3α (Figure 2b).

FD1 transcript and protein expression patterns

FD1 expresses spatially and temporally restricted patterns during embryogenesis, as revealed by both *in situ* hybridization and staining with antibodies generated against the bacterially produced FD1-encoded protein (see Materials and methods). Figure 3 shows the *in situ* hybridization patterns of FD1 transcripts and anti-FD1 antibody staining during embryogenesis. The patterns of expression were virtually identical, suggesting that FD1 expression is mainly, if not exclusively, controlled at the level of transcription. Furthermore, the nuclear anti-FD1 antibody staining is consistent with the predicted function, supporting the idea that FD1 encodes a DNA-binding transcription factor.

FD1 is expressed initially in both the anterior and posterior regions of the embryo (Figure 3a and g). In the anterior region, FD1 transcripts appear in a ventrally shifted 'anterior cap', while the posterior expression domain consists of a transient ventrally located 'posterior spot'. During cellular blastoderm, the anterior cap retreats from the pole region and forms a tilted stripe which covers the anlagen of the stomodeal invagination up to the position of the clypeolabral anlage on the blastoderm fate map (Figure 3b, c, h and i). During gastrulation, the anterior cap retreats from the clypeolabral region. FD1 transcripts accumulate in cells associated with the developing foregut as well as a region corresponding to the intercalary segment anlage (Figure 3d and j). In the posterior region, FD1 expression is reinitiated in an area corresponding to the developing mesoderm adjacent to the hindgut (Figure 3c and i). During the extended germband stage, the anterior FD1-expressing cells form a cluster of cells in association with the developing foregut to eventually line the posterior pharynx wall. In addition, a number of FD1-expressing cells can be found in a metameric pattern within the developing mesoderm. A description of the late expression patterns outside the head region will be reported elsewhere.

Spatial regulation by synergistic interactions between maternal coordinate gene activities

The activation of the zygotic segmentation gene cascade requires the preceding activity of the maternal coordinate systems, including the anterior organizer gene *bicoid* (*bcd*), the posterior gene *nanos*, the terminal signal transduction pathway mediated by the receptor tyrosine kinase encoded by *torso* (*tor*) and the dorsoventral morphogen encoded by *dorsal* (*dl*; reviewed in St Johnston and Nüsslein-Volhard, 1992). As shown in Figure 4, the anterior, the dorsoventral and the terminal systems are functionally required for the activation of FD1 expression and for the spatial control of the anterior cap domain, while the posterior system is not required for the regulation of the FD1 expression pattern (data not shown).

In the absence of *bcd* activity, FD1 failed to be expressed in the anterior cap domain (Figure 4a). Conversely, an increase in *bcd* activity in the embryos led to an expansion of FD1 expression towards the posterior (Figure 4b). However, this expansion was one-sided with respect to the dorsoventral axis of the embryo. Thus, although the *bcd* product acts in a concentration-dependent manner, FD1 expression can only be expanded ventrally in the presence of *dl* activity. In fact, the lack of *dl* activity causes a strong reduction of the FD1 expression domain to a single spot, corresponding in position to the peak of *bcd* activity at the anterior pole (Figure 4c). Conversely, dorsal activity along the entire dorsoventral axis, as in embryos laid by *Toll* mutant females, causes an expansion of the FD1 expression domain towards the dorsal-most position (Figure 4d). In embryos lacking *tor* activity, FD1 expression was abolished in the dorsal region (Figure 4e). However, if *tor* is activated ectopically due to the dominant tor^{4021} mutation, the FD1 expression domains are expanded significantly on the ventral side (Figure 4f).

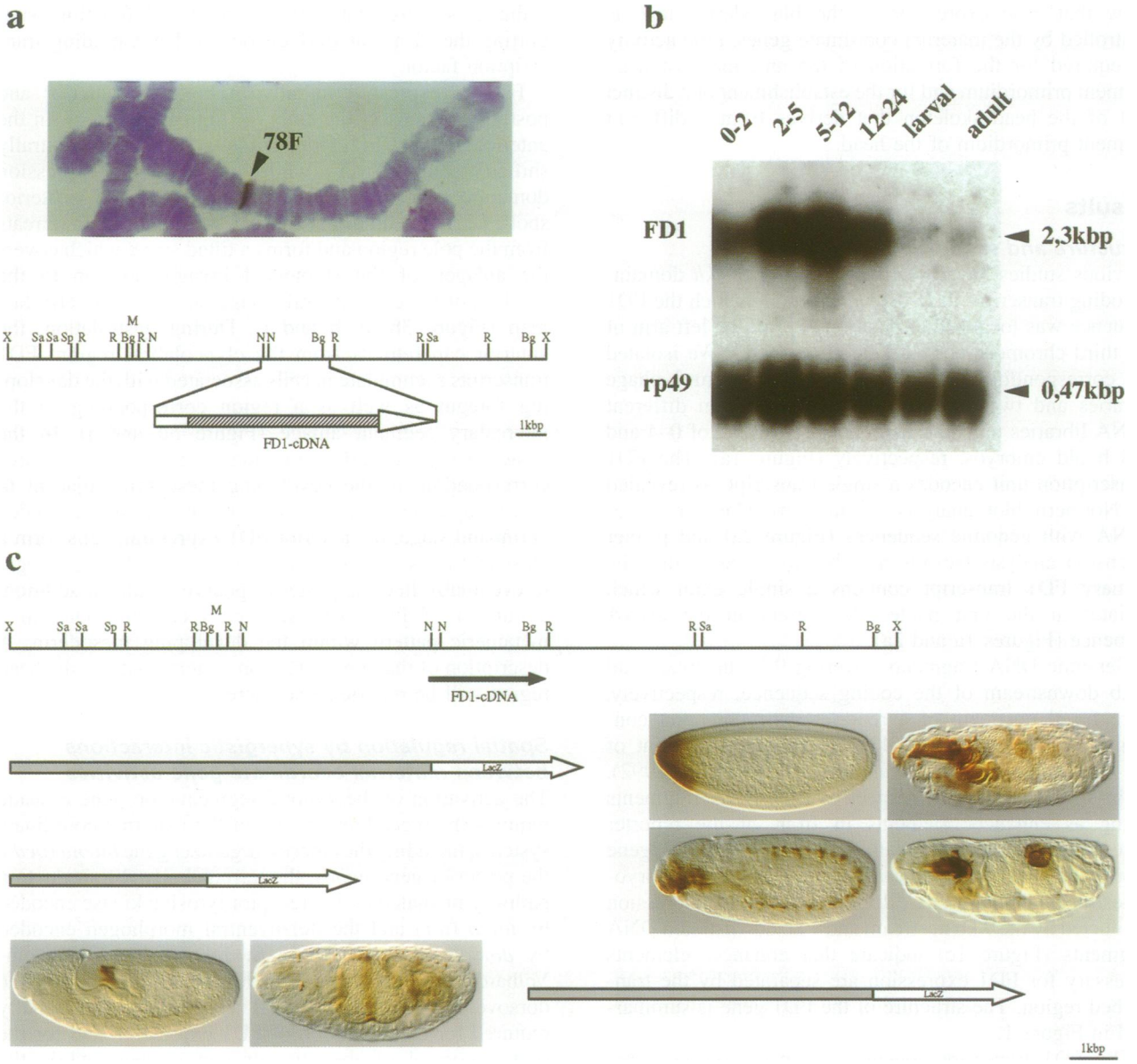

Fig. 1. Molecular characterization of the FD1 locus. (**a**) *In situ* hybridization of the FD1 cDNA to polytene chromosomes (top) determines the cytological location of the transcription unit to be 78F on the left arm of the third chromosome. The restriction map (bottom) shows the location of the cDNA within 18.3 kb of genomic DNA. The FD1 transcription unit is encoded by a single exon of 2.4 kb, shown enlarged below the map. The hatched box indicates the translated region. Bg, *Bgl*II; M, *Mlu*I; N, *Nsi*I; R, *Eco*RI; Sa, *Sac*I; X, *Xba*I. (**b**) Developmental Northern blot analysis of FD1 expression showing a single transcript of 2.3 kb throughout all stages of embryogenesis and its reduced amounts during larval and adult stages. rp49 DNA (O'Connell and Rosbash, 1984) was used to assess the RNA content of various lanes. Numbers above the blot refer to the time after egg laying at which RNA was prepared. (**c**) *Cis*-regulatory elements of the FD1 gene: three enhancer–*lacZ* fusion constructs encompassing 9, 4 and 7 kb (top to bottom) of genomic DNA flanking the transcript are shown below the restriction map. *lacZ* reporter gene expression driven by each construct is depicted beside the construct. The 9 kb *Xba*I–*Nsi*I fragment is sufficient to drive expression during early and late stages in the head region, as well as in several mesoderm-related expression domains in the trunk, while the 4 kb *Xba*I–*Bgl*II subfragment conducts the initial expression after gastrulation. The 7 kb *Bgl*II fragment (3′ to the transcription unit) drives expression in the posterior tip of the mesoderm and in the midgut constrictions. These *cis*-acting regions (shaded boxes) contain the necessary elements for the spatial aspects of FD1 expression, but neither the 9 kb DNA fragment 5′ to the transcript nor the 7 kb 3′ sequences (fused in front of the transcription unit) is sufficient to conduct transgene rescue of *croc* mutant embryos (for the identity of the FD1 and the *croc* gene see the text).

We also examined whether the FD1 anterior cap domain was affected in embryos which carry homozygous mutations for known *bcd* target genes such as *orthodenticles*, for *tor* target genes such as the terminal gap genes *tailless* (*tll*) and *huckebein* (*hkb*) and for the *dl* target genes *twist*, *snail* and *zerknüllt* (reviewed by St Johnston and Nüsslein-Volhard, 1992). The absence of each one of these gene activities did not affect FD1 expression.

FD1 mutations cause the croc mutant phenotype

By formal criteria, FD1 takes a position within the genetic hierarchy at the level of the known *Drosophila* gap genes. Gap genes are the first and small-sized zygotic segmentation genes which are locally expressed during the short nuclear division cycles prior to the cellular blastoderm (Rothe *et al.*, 1992 and references therein). In this view, the short size of the FD1 primary transcript,

a

```
-360  CTGGTGCGGGCGACGCTCTGCCGACGCAAACGCTTCTGCTCGATGCGAGCTGTGTGTTATTTCAGTTACACAGATTCAGTATCAGACGTG
-270  TTACTATACCGAACGGTTCTGCGCTCGCTCTGCGCTCCCATTTCCCCTGTACTACAAATAGATGCATATATATATATATATATAGATT
-180  TATATATCTGCACTTATGTGAGCGACTTGTTTAAGCAAATATTATCGCTAGAAATAATAAGAAGCAGCCAGTGGTGAAGTGATTAAGTTC
 -90  GAGAAAGAAAGCAAGACAAAAGCAAGAGAGTACGAAAATACCAGAAGGCGGCAGTGGAAAATCGATCAGAAATCCGCCGGAATCGCCGCC
   1  M  H  T  L  F  S  D  Q  N  S  F  T  R  H  Y  A  Q  T  A  A  G  Y  G  S  A  S  A  V  A  A
   1  ATGCATACGTTATTCAGCGACCAGAACTCCTTTACGCGCCACTATGCCCAGACCGCAGCCGGTTACGGATCCGCATCCGCGGTGGCGGCG
  31  A  S  S  A  S  A  A  A  A  A  H  Y  A  Y  D  Q  Y  S  R  Y  P  Y  S  A  S  A  Y  G  L  G
  91  GCCAGCAGCGCCTCCGCTGCAGCAGCGGCACATTACGCATACGACCAGTACTCCCGCTATCCGTACTCGGCCAGCGCCTACGGCCTGGGT
  61  A  P  H  Q  N  K  E  I  V  K  P  P  Y  S  Y  I  A  L  I  A  M  A  I  Q  N  A  A  D  K  K
 181  GCCCCGCACCAGAACAAGGAGATTGTAAAGCCACCTTACTCGTACATCGCCCTGATAGCCATGGCCATCCAGAATGCGGCGGACAAGAAG
  91  V  T  L  N  G  I  Y  Q  Y  I  M  E  R  F  P  Y  Y  R  D  N  K  Q  G  W  Q  N  S  I  R  H
 271  GTGACCCTGAACGGGATCTATCAGTACATCATGGAGCGATTTCCCTACTATCGCGACAACAAGCAGGGCTGGCAGAACTCCATTCGGCAC
 121  N  L  S  L  N  E  C  F  V  K  V  A  R  D  D  K  K  P  G  K  G  S  Y  W  T  L  D  P  D  S
 361  AATCTTAGCCTCAACGAGTGCTTCGTGAAGGTGGCCCGCGATGACAAGAAGCCCGGCAAGGGCTCCTACTGGACTCTGGATCCGGACTCC
 151  Y  N  M  F  D  N  G  S  F  L  R  R  R  R  R  F  K  K  K  D  V  M  R  E  K  E  E  A  I  K
 451  TACAACATGTTCGACAACGGCTCCTTCCTTCGCCGGCGTCGCCGCTTCAAGAAGAAGGATGTGATGCGGGAGAAGGAGGAGGCCATCAAG
 181  R  Q  A  M  M  N  E  K  L  A  E  M  K  P  L  K  L  M  T  N  G  I  L  E  A  K  H  M  A  A
 541  CGGCAGGCCATGATGAACGAGAAGCTGGCCGAGATGAAGCCCCTCAAGCTGATGACCAACGGCATACTGGAGGCCAAGCACATGGCCGCC
 211  H  A  A  H  F  K  K  E  P  L  M  D  L  G  C  L  S  G  K  E  V  S  H  A  A  M  L  N  S  C
 631  CATGCTGCCCACTTCAAAAAGGAGCCCCTAATGGACTTGGGTTGTCTAAGTGGCAAGGAGGTCTCCCATGCCGCAATGCTCAACTCCTGC
 241  H  D  S  L  A  Q  M  N  H  L  A  G  G  G  V  E  H  P  G  F  T  V  D  S  L  M  N  V  Y  N
 721  CACGACAGCTTGGCCCAAATGAACCACTTGGCTGGCGGCGGAGTGGAGCATCCCGGCTTCACCGTGGACTCACTCATGAATGTGTACAAC
 271  P  R  I  H  H  S  A  Y  P  Y  H  L  N  E  D  N  L  A  T  V  A  S  S  Q  M  H  H  V  H  H
 811  CCGCGGATCCATCACAGCGCCTATCCCTACCACTTGAACGAGGACAACTTGGCCACGGTGGCCAGCAGTCAGATGCATCATGTCCACCAT
 301  A  A  A  A  H  H  A  Q  Q  L  Q  R  H  V  A  H  V  A  H  P  L  T  P  G  G  Q  G  A  G  G
 901  GCTGCGGCGGCCCATCATGCCCAGCAACTGCAGCGTCATGTGGCGCACGTGGCCCATCCTCTGACCCCCGGCGGCCAAGGAGCGGGAGGC
 331  Q  S  S  G  H  S  P  T  T  I  S  T  P  H  G  P  A  H  G  G  W  Y  T  P  E  T  P  P  S  E
 991  CAATCGTCCGGACACTCGCCCACCACGATCTCAACTCCGCATGGTCCCGCGCACGGGGGTTGGTATACCCCGGAAACGCCACCATCGGAA
 361  P  V  P  H  N  G  Q  Q  G  T  P  T  H  P  G  H  N  N  N  N  S  S  S  V  L  N  H  N  G  V
1081  CCCGTTCCTCACAATGGGCAGCAGGGCACACCCACCCACCCGGGACATAATAACAACAACAGCAGCAGCGTACTCAACCACAATGGAGTG
 391  G  N  G  G  G  G  G  G  G  G  G  G  G  S  S  S  V  L  T  S  S  P  T  S  A  L  G  F  R  D
1171  GGCAACGGTGGCGGAGGCGGCGGCGGTGGCGGTGGAGGATCCTCCAGCGTGCTGACCAGCAGTCCCACATCGGCACTGGGTTTCCGCGAC
 421  M  I  F  E  Q  N  Q  S  C  Q  L  D  T  G  S  P  T  G  S  L  Q  S  A  S  P  P  A  S  A  S
1261  ATGATCTTCGAGCAGAACCAATCCTGCCAACTGGACACCGGTAGTCCCACCGGCAGCCTACAATCCGCCAGTCCACCTGCCTCCGCATCC
 451  V  A  A  A  S  A  A  A  A  A  A  A  V  I  S  S  H  H  H  H  H  H  H  H  A  A  L  S  G  N  L
1351  GTGGCAGCGGCATCAGCGGCAGCAGCGGCTGCGGTGATCAGTAGCCACCACCACCACCATCATCACCATGCGGCGCTTAGCGGGAACCTG
 481  G  Q  L  G  Q  L  S  N  L  S  H  Y  R  P  H  V  G  H  Y  Q  E  Y  G  I  K  Y  G  V  *
1441  GGCCAGCTGGGTCAGCTGAGCAATCTGAGCCACTACCGACCGCATGTGGGCCACTACCAGGAGTACGGCATCAAGTACGGGGGTGTAAGCC
1531  TGTGGGCTGGATTCTCGGCGACTGGGCTGGGCTCTCTTGTATATTCCTGTAAATACATACACCATATAATATACCCAAGAACGAGTCTAC
1621  ACAGCCTGGGGGTGGTGATCCCTCGGGGTCCAGAGTCTGCTATCTCAGATCTTTTTGTTTCCAACTTGTTTTCGATTTATCCGTATACGA
1711  AATAATGGGGGACATAGCTGTTAAGCAACCGCATTAACGTATTGTATTGTATTAACTGTATTTGTAATTGTAGCTCGTAATTGTTATCGA
1891  ATTTAGGCATAGTCCCAATTGTCTGCTACGCAATTTGTGATGGGGTTTTGCTACACAGAATATAAACTTAAATTAGAGCATTTTGTCAAT
1891  AACTAAAACTACCGCATCGGATCGATCTATCATCGCCTAGTACACTAGCCTAGCCCACGTAAGCAATAATAAATTATACAACTAAATCGT
1981  AGCTATAACATAACTATCAAAACTAGCGTACCACCAATATGAGCCGAACAAAAAGTGTATTGAAAATATGTTAAAATATCTCAAATCGTG
2071  TGTTTTATTTGGCCCCTTCATATTCGGAGCGACTTTCGCTATTCCATCTGGCCATTGAGCATTTGGCTGGCAAAACAGACTTTTTACCCC
```

b

```
          --1-------10--------20--------30--------40--------50--------60--------70--------80--------90--------

FD1       IVKPPYSYIALIAMAIQNAADKKVTLNGIYQYIMERFPYYRDNKQGWQNSIRHNLSLNECFVKVARDDKKPGKGSYWTLDPDSYNMFDNGSFLRRRRRFK
FKH1      M----------T------P---I-------F--D---F------------------------P----------------------E-----------  92%
MFH-1     L----------T------PE--I-------F--D---F--E----------------------P----------------------E-----------  90%
XFD-4     L--------G--T------P---I-------F--D---F--E----------------------P----------S---------E-----------  89%
fkh       HA-------S--T-----NPTRML--SE---F--DL--F--Q-Q-R-------S--F-D----IP-TPD------F---H---G---E--CY---QK---  66%
HNF-3α    HA-------S--T----Q-PS-ML--SE---W--DL-----Q-Q-R-------S--F-A-------SPD----------H---G---E--CY---QK---  71%
```

Fig. 2. (**a**) Genomic nucleotide and deduced amino acid sequence of FD1. The transcribed sequence (2436 nucleotides), as revealed by a full size cDNA, is shown; the numbering of the nucleotides and amino acids is indicated on the left-hand side by standard and bold numbers, respectively. The starts and ends of cDNAs are shown by small boxes; the *fkh* domain (see also b) is underlined. (**b**) Amino acid sequence comparison of the *fkh* domain of FD1 and related genes. The FD1 *fkh* domain is most similar to the *fkh* domains of MFH-1, FKH1 from mouse and XFD-4 from *Xenopus*. *fkh* domains of the prototype genes *fkh* and HNF-3α are also shown. Identical amino acids are depicted as '–'. '%' refers to sequence identity relative to the FD1 sequence; numbers indicate amino acid positions within the *fkh* domain.

the spatial aspects of the expression domain as well as the genetic requirement necessary for anterior cap expression suggest that FD1 might have an early function in the establishment of the larval head.

In the search for a corresponding gene function, we examined the mutation *croc*. The three *croc* mutant alleles $croc^{5f59}$, $croc^{75\text{-}3}$ and $croc^{UH}$ cause embryonic lethality associated with severe defects in the pregnathal head region of the homozygous embryo (see below). Complementation analysis, including the deficiency PC^{CP1} and the transposition *Tp(3,Y) J151* (Lindsley and Zimm, 1992), were used to map the *croc* gene to the chromosome interval 78E5–78F (data not shown) covering the cytogenetic location of the FD1 transcription unit. To link the FD1 transcript to *croc* function, we examined whether the FD1 transcription unit was altered in the *croc* mutant alleles.

The $croc^{5f59}$ mutation resulted in a transversion of T to A, causing a stop codon after the 13th amino acid of the putative FD1 protein. Anti-FD1 antibody staining of $croc^{5f59}$ mutant embryos showed that although the transcript is expressed normally, no FD1 protein could be detected (data not shown). These results are consistent with the argument that the $croc^{5f59}$ mutation leads to a premature truncation of the FD1 protein.

$croc^{75\text{-}3}$ DNA contains two point mutations. They cause the replacement of an alanine residue by valine at position 453 and the replacement of a leucine residue by phenylalanine at position 122 of the putative FD1 protein. While the first replacement lies within a functionally unidentified region of the FD1 protein, the second involves an amino acid residue within the *fkh* domain which is conserved among all known family members. In homozygous $croc^{75\text{-}3}$

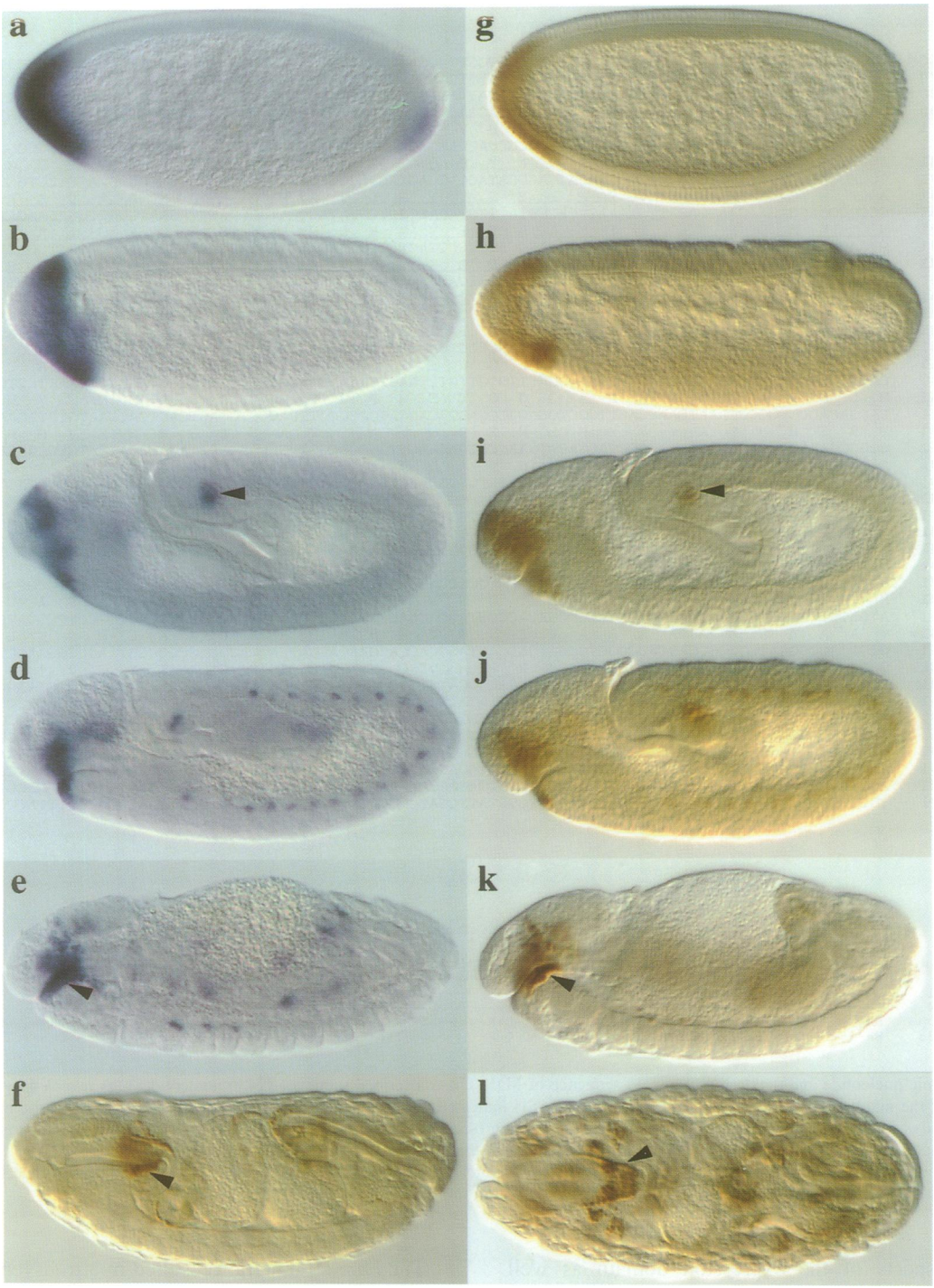

Fig. 3. Expression of FD1 during embryogenesis, as revealed by *in situ* hybridization (**a–e**) or anti-FD1 antibody staining (**f–l**) to whole-mount embryos. (a and g) Embryos at the syncytial blastoderm stage (stage 5). (b and h) Embryos at the onset of gastrulation (stage 6). Note the retraction of the anterior expression domain from the pole. The posterior domain has disappeared. (c and i) Embryos at the early elongated germband stage (stage 9). Note the additional expression domain at the posterior end of the germband adjacent to the hindgut (arrowhead). (d and j) Embryos at the late elongated germband stage (stage 11). Note the additional expression in a segmentally repeated pattern in the mesodermal part of the germband. In the head region, transcripts and protein are concentrated in the ventral epithelium of the stomodeum. (e and k) Embryos at the retracted germband stage (stage 13). Note the persistent expression in the ventral epithelium of the invaginating foregut (arrowhead). (f and l) Terminally developed embryos at the end of embryogenesis (stage 17). Note the expression of FD1 protein in the ventral posterior epithelium of the pharynx (arrowhead). All embryos are oriented with their anterior pole to the left, dorsal side up, except in (l), which shows a dorsal view, with anterior to the left. Staging is according to Campos-Ortega and Hartenstein (1985).

mutant embryos, both transcript and protein expression appear to be normal, as assayed by *in situ* hybridization and anti-FD1 antibody detection, respectively. However, the *fkh* domain of the $croc^{75\text{-}3}$ mutant protein is unable to bind DNA (see below).

The FD1 coding sequence of $croc^{UH}$ mutant DNA was identical to the wild-type sequence. However, neither FD1 transcripts nor FD1 protein could be detected in homozygous $croc^{UH}$ embryos (data not shown). This suggests that the corresponding mutation affects the control

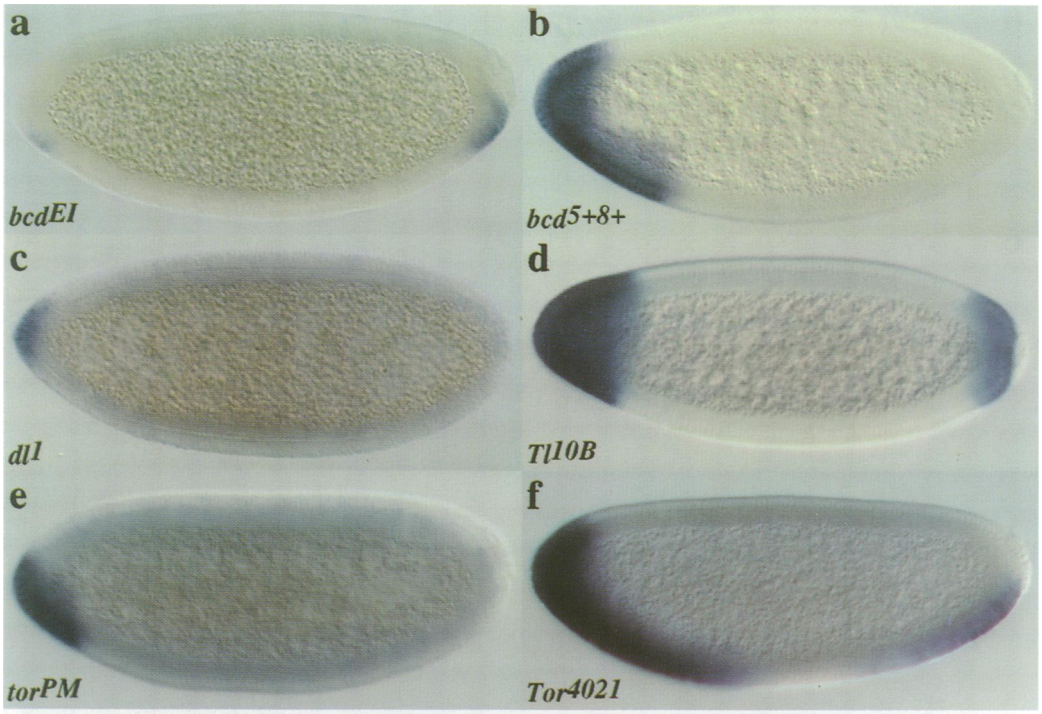

Fig. 4. Regulation of FD1 by key genes of the maternal organizer systems shown by *in situ* hybridization to whole-mount embryos. (**a**) Embryo derived from a homozygous mutant *bcd* mother. The posterior expression domain of FD1 is duplicated at the anterior pole. (**b**) Embryo derived from female containing multiple copies of *bcd*. The anterior expression domain of FD1 is expanded towards the posterior on the ventral side. The posterior expression is absent. (**c**) Embryo derived from a homozygous *dl* mutant mother. The anterior expression domain of FD1 is strongly reduced. The posterior expression is absent. (**d**) Embryo derived from a mother carrying a dominant allele of the *Toll* gene. Because of the uniformly distributed activity of dl protein in this embryo, FD1 is expressed symmetrically around the circumference of the embryo at both poles. (**e**) Embryo derived from a homozygous *torso* mutant mother. The anterior expression of FD1 is reduced, and the posterior expression is absent. (**f**) Embryo derived from a mother carrying a dominant *torso* allele. FD1 expression is expanded along the ventral side of the embryo at both poles. Embryos are at the syncytial blastoderm stage and are oriented with the anterior pole to the left and the dorsal side upwards.

region of the FD1 gene. Thus, each one of the *croc* mutant alleles shows that the corresponding mutation severely interferes with the functional expression of the FD1 gene product in a manner likely to cause the lack of FD1 function. The latter assignment is consistent with the findings that the mutant phenotypes of homozygous $croc^{5f59}$, $croc^{75\text{-}3}$ and $croc^{UH}$ mutant embryos are almost identical, as judged by morphological criteria, and that the same phenotype was observed when each of the mutant alleles was examined *in trans* to one of the other alleles or to the deficiency PC^{CP1} (data not shown). Taken together, these results establish that FD1 encodes *croc* function. Therefore we refer to the FD1 transcription unit as the *croc* gene.

croc requirement for segment patterning and head skeletal structure formation

To establish the biological function of *croc* gene activity, we asked whether regions of the embryo corresponding to the anterior cap domain of *croc* expression were affected in *croc* mutant embryos. Initially, this domain covers the anlagen of the clypeolabrum, the anterior midgut including the esophagus and part of the intercalary segment on the blastoderm fate map (Hartenstein *et al.*, 1985). During gastrulation, it retracts from the clypeolabrum and the gut anlagen but remains within the intercalary segment anlage (Figure 3b, c, h and i). Using the expression pattern of the segment polarity genes as a molecular marker for the establishment of head segment equivalents (Schmidt-Ott and Technau, 1992), we found that the expression patterns of *engrailed* (*en*) and *wingless* (*wg*) were altered in a single position in *croc* mutant embryos (Figure 5). Within the clypeolabral segment anlage, *en* expression is absent and the adjacent *wg* expression domain is significantly expanded and covers the area normally expressing *en* in addition to its normal expression domain (Figure 5c and d). Wild-type *croc* activity is therefore required for the control of segment polarity gene expression within a single segment equivalent, i.e. the anterior-most segment equivalent of the head anlage of the embryo.

As judged by the limited morphological markers which can be used to distinguish head segments (Cohen and Jürgens, 1991), the clypeolabrum of *croc* mutant embryos is at least partially differentiated because the characteristic labral sensory organs and the labrum itself can be observed in *croc* mutant larva. We also examined the internal clypeolabral structures using ectodermal and mesodermal cell markers. As a marker for ectodermal derivatives, we used the enhancer trap line P1618 (Karpen and Spradling, 1992). This enhancer trap line expresses β-galactosidase exclusively in the muscle attachment sites, termed apodemes, which are located on the inner side of the epidermis (Volk and VijayRaghavan, 1994). Figure 6a and b shows that the two dorsal rows of apodemes are present in *croc* mutant embryos, while the single ventral row of apodemes is missing. In addition, the palisade-like structure of dorsopharyngeal muscles never forms, as revealed by anti-myosin heavy chain antibody staining which labels muscle

Fig. 5. Segment polarity gene expression in *croc* mutant embryos. (**a** and **b**) Anti-*en* antibody staining, indicating that the *en* expression domain in the clypeolabrum anlage of wild-type embryos (a; arrowhead) is absent in $croc^{5f59}$ mutant embryos (b; arrowhead) at stage 13. (**c** and **d**) Double *in situ* hybridization *en* (brown) and *wg* (blue) patterns of wild-type (c) and *croc* mutant embryos (d) at stage 11 of embryogenesis, indicating that the clypeolabral expression domain of *wg* in *croc* mutants expands and thereby covers the domain corresponding to the *en* expression of wild-type (see a and c; arrowheads).

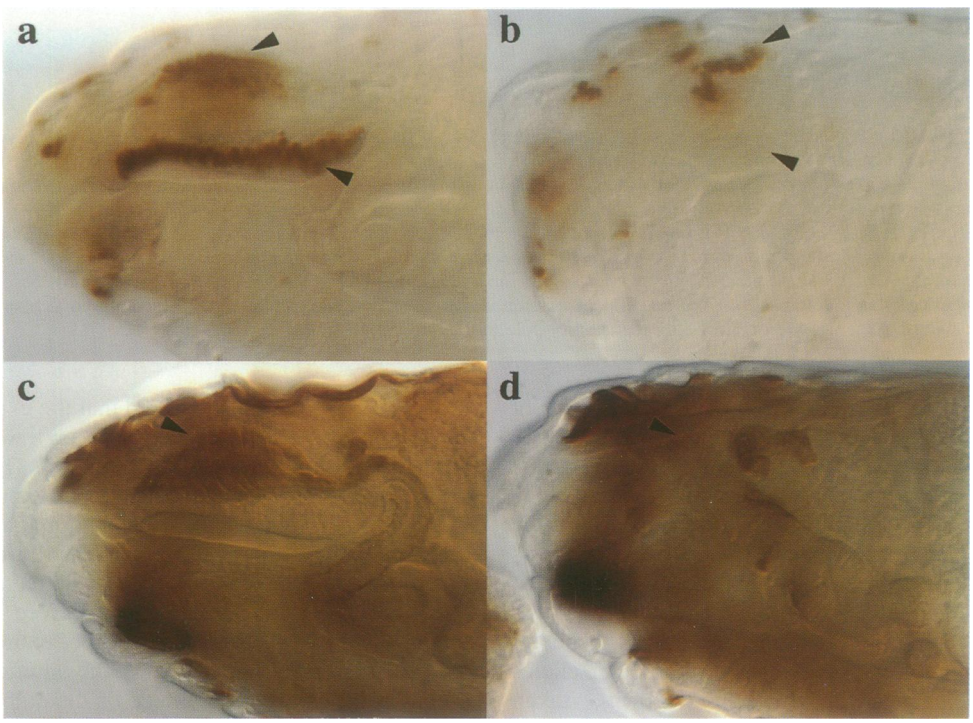

Fig. 6. Defects in the clypeolabral derivatives of *croc* mutant embryos. (**a** and **b**) Apodeme-specific β-galactosidase expression pattern of the P element insertion line P1618 (Karpen and Spradling, 1992). Arrowheads indicate the dorsal (top) and ventral (below) apodemes (epidermal muscle attachment sites) of the dorsal pharyngeal muscles of wild-type embryos (a) and their corresponding position in a $croc^{5f59}$ mutant embryo (b) at stage 16. Note that the ventral apodemes are absent in the $croc^{5f59}$ mutant embryo. The dorsal apodemes are composed of two rows of apodemes; one of them is out of focus. (**c** and **d**) Expression of myosin heavy chain, as revealed by anti-myosin heavy chain antibody staining at stage 17. Arrowheads indicate the dorsal pharyngeal muscles in wild-type (c) and their absence in the $croc^{5f59}$ mutant embryos. Orientation of the enlarged head regions: anterior pole left, dorsal side up; stages according to Campos-Ortega and Hartenstein (1985).

cells specifically (Figure 6c and d). Taken together, these results suggest that *croc* expression in the clypeolabral anlage is functionally required for the normal establishment of ectodermal and mesodermal derivatives in the corresponding segment.

croc expression in the clypeolabrum anlage is temporally restricted to blastoderm and early gastrulation (Figure 3). To link this early and transient phase of *croc* expression to defective structures observed in late *croc* mutant embryos, we used the FD1–*lacZ* transgene causing β-galactosidase reporter gene expression in the *croc*-expressing cells at the blastoderm stage. Due to the stability of β-galactosidase, the *lacZ* reporter gene-expressing cells can be traced throughout development, i.e. initial *croc* expression in the clypeolabral anlagen eventually ends up within the dorsopharyngeal muscles (Figure 1c). This finding and the lack of the corresponding body structure in *croc* mutant embryos indicate that the transient *croc* expression in the clypeolabrum anlage is required for the establishment of the normal dorsopharyngeal muscle pattern.

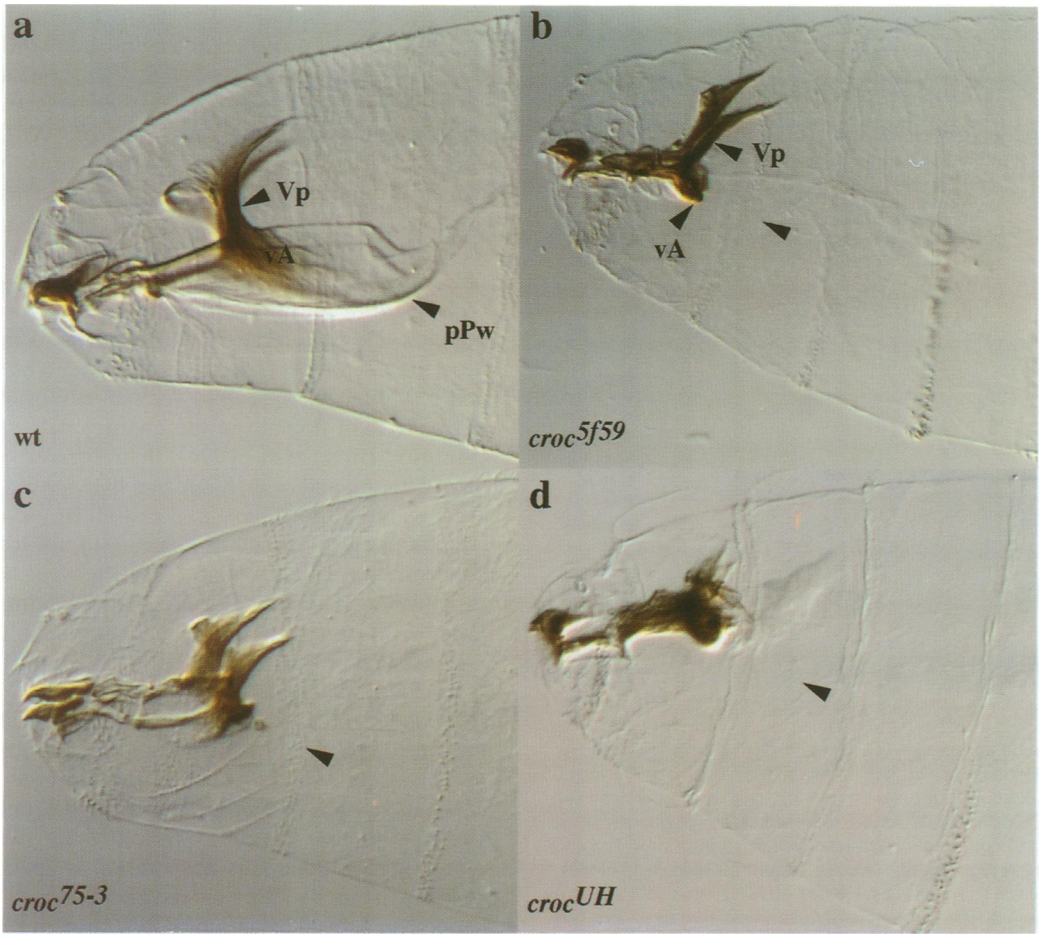

Fig. 7. Cuticle preparations showing first instar larval heads of wild-type and *croc* mutants. (**a**) Cephalopharyngeal skeleton of a wild-type larva (wt). The following skeletal elements are indicated: pPw, posterior pharynx wall; vA, ventral arm; Vp, vertical plate. (**b–d**) Different *croc* mutant larvae; *croc* alleles are indicated in the left-hand bottom corner. Note that the posterior wall of the pharynx is missing from the cephalopharyngeal skeleton (arrowhead) and that the ventral arm (arrowhead in b) is strongly reduced. Note that the ventral arm and the Lateralgräte are also reduced in the *croc* mutant larvae. These structures derive from the mandibular segment (ventral arms, Lateralgräte) and from the acron (dorsal arm), respectively, where *croc* is not expressed. We assume that the reduction of structures derived from the mandibular segment and the acron are secondary defects caused by the improper juxtaposition of the headlobes.

In contrast to its early and transient expression in the clypeolabral anlage, *croc* expression is maintained in intercalary derivatives, as eventually seen in a row of cells lining the posterior ventral part of the pharynx (Figure 3e, f, k and l) where the posterior wall of the pharynx normally forms. Larval cuticle preparations show that the posterior wall of the pharynx is absent in *croc* mutants and the ventral arm of the cephalopharyngeal skeleton is strongly reduced (Figure 7). Since the *wg* and *en* expression patterns were normal in the intercalary segment anlage of *croc* mutant embryos, it appears that *croc* plays a dual role at different levels during embryonic head development. Its early activity is transiently required for determinative events in establishing the posterior portion of the clypeolabrum, as reflected in the altered segment polarity gene expression patterns of *en* and *wg*, while its activity in the developing intercalary segment is not required for the establishment of the segment anlage *per se* but rather for the differentiation of specific structural elements.

DNA-binding properties of croc wild-type and mutant protein

To investigate the effect of the single residue (L122 → F122) exchange in the *fkh* domain of the mutant protein on its DNA-binding properties, we first determined the target sequence for the *fkh* domain of the wild-type protein and then performed a binding analysis of the corresponding mutant. Proteins encompassing the *fkh* domains were obtained by expression in *Escherichia coli* and subsequent affinity purification. Target analysis of the wild-type domain was achieved by a PCR/band shift supported selection procedure (Blackwell and Weintraub, 1990) from a mixture of deoxyoligonucleotides being degenerated at 16 consecutive positions. After eight cycles of binding site selection and enrichment, the targets were cloned and sequenced. Approximately 90% of >60 sequenced targets share a 7 bp consensus motif, the flanking regions of which do not show obvious similarity to each other. An arbitrary selection of sequences is shown in Figure 8a.

Two of the randomly selected binding sites were subjected to a gel retardation assay to study their ability to bind to the wild-type and mutant *croc fkh* domains. Figure 8b shows that these targets are efficiently shifted by the wild-type domain but that the single-point mutation in the mutant protein completely abolishes its DNA binding property. DNase footprinting verified this finding by showing that the *croc* wild-type domain protects the selected target site from hydrolysis (Figure 8c).

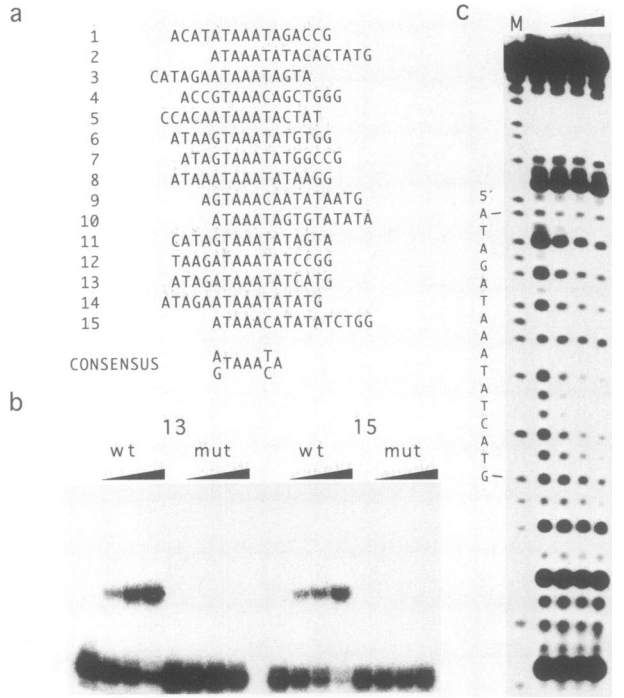

Fig. 8. DNA binding properties of *croc* wild-type and mutant *fkh* domains. (**a**) Alignment of *fkh* domain recognition sites. *In vitro* target sites have been obtained after eight cycles of selection and enrichment by means of a PCR-supported selection procedure employing a mixture of deoxyoligonucleotides degenerated at 16 positions. The 15 templates shown were arbitrarily selected from >60 sequenced clones sharing the indicated 7 bp consensus motif. (**b**) Gel retardation analysis of *croc* wild-type and mutant *fkh* domains. Two templates (13, 15) have been randomly selected for binding analysis. (Left) Increasing amounts (0, 10, 20 and 50 ng; see triangle) of wild-type (wt) and mutant (mut) protein have been applied to labeled template 13. (Right) Same as left but probing sequence 15. (**c**) DNA footprint analysis. Template 13 has been 3′ end-labeled, incubated with increasing amounts (0, 10, 30 and 90 ng; see triangle) of the wild-type *croc fkh* domains and subsequently digested with 100 ng DNase. M, Maxam–Gilbert reaction specific for dG and dA residues; sequence is shown to the left.

Discussion

We provide evidence that the *croc* gene of *Drosophila* encodes a *fkh* domain protein which is functionally required for the normal establishment of two non-adjacent larval head segments. *croc* acts as an early patterning gene in the anterior-most head segment anlage and plays a critical role in the differentiation of intercalary segment structures at a later stage of embryogenesis. Based on its control function within a single head segment anlage, *croc* cannot be grouped among the known segmentation genes. It may represent a new type of head patterning gene or an adapted gene that emerged from an ancestor with a more general function in head development of more primitive insects (see below).

Control and function of croc

croc fails to be expressed in the absence of *bcd* activity. In the absence of either *tor* or *dl* activity, *croc* expression becomes restricted to the anterior-most position of the embryo where *bcd* activity is highest. When *bcd* activity is increased because of extra *bcd* copies in the females, *croc* expression expands posteriorly, as expected if *bcd* acts in a concentration-dependent manner. Because this expansion is restricted to the ventral side (where *dl* activity is functional; see review by St Johnston and Nüsslein-Volhard, 1992), it appears that *croc* expression is dependent on at least two factors, *bcd* and *dl*. When *dl* is also ectopically active in the dorsal region of embryos derived from *Toll* females, the *croc* expression domain covers the anterior pole position including the dorsal side. Because *dl* activity is not sufficient for *croc* activation and does not depend on *bcd* activity, these results suggest that *bcd* requires *dl* to set the spatial limit of the *croc* anterior cap domain. This observation is analogous to the previously described synergistic interaction between *hunchback* (*hb*) and *bcd*, which is required for the setting of the zygotic *hb* expression domain in the anterior half of the embryo (Simpson-Brose *et al.*, 1994). In both cases, i.e. activation of zygotic *hb* and *croc*, it appears that *bcd* is the activating component of target gene expression. However, it requires a helper molecule, such as *hb* or *dl*, to lower the threshold concentration above which *bcd*-dependent activation can be achieved. The notion of *bcd* and *dl* binding sites within the *croc* upstream region (U.Häcker, unpublished results) is consistent with this proposal.

The terminal maternal system is initially required for *bcd*-dependent *croc* activation in the dorsal position of the anterior cap and may function in a manner analogous to *dl* on the ventral side. After blastoderm, the *croc* expression domain withdraws from the anterior cap in response to *tor* activity, which then functions in a repressing fashion. This observation is again reminiscent of what is seen with zygotic *hb* expression. Thus, *tor*-dependent phosphorylation of *bcd* protein (Ronchi *et al.*, 1993) may cause a functional switch of the activator into a repressor, or the *bcd* protein may thereby just lose its function as an activator (Sprenger and Nüsslein-Volhard, 1992). These phenomena, which are not yet understood in molecular terms, and the observation that zygotic target genes of the maternal organizers had no affect on the pattern of initial *croc* expression, suggest that the initial phase of *croc* expression is controlled exclusively by maternal factors.

Based on the maternal requirement for the spatial regulation of gene expression, *croc* can formally be placed at the level of the zygotic gap genes. However, while the absence of gap gene activities causes the lack of adjacent segment equivalents, *croc* affects only a portion of a single segment (defined by the absence of a single *en* expression domain in the head anlage and the corresponding expansion of the adjacent *wg* domain, which might be a result of the 'self-organizing' properties of the class of segment polarity genes once they are activated; Martinez-Arias, 1993). In the trunk, segment polarity gene activation and its spatial control are initiated by the preceding gap and pair-rule gene activities. In the case of the clypeolabrum anlage, however, *en* expression appears to be controlled either directly or indirectly by *croc* activity. The *croc* mutant phenotype is consistent with a specific requirement of the wild-type gene function in setting up the developmental fate of cells giving rise to only part of the clypeolabrum, because labral sensory organs and the labrum itself are formed, while specific muscles as well as their epidermal attachment sites are missing. These findings and the preceding lack of *en* expression in the clypeolabral anlage

suggest that the precursor cells giving rise to the posterior part of the clypeolabrum are not determined properly. Therefore *croc* may act to regulate the position-dependent posterior cell fate within a single head segment, as reflected in the altered *en* and *wg* expression patterns in the clypeolabrum anlage.

In the clypeolabral anlage, *croc* is expressed only transiently and its activity vanishes once segment polarity gene expression domains are established. However, in the intercalary segment, *croc* is not required for the proper expression of segment polarity genes and its expression is maintained. The different modes of *croc* expression in the two head segment anlagen are consistent with the finding that within the intercalary segment, a distinct head structure, the posterior wall of the pharynx, is absent. This implies that *croc* functions at different levels of pattern organization in two non-adjacent head segments, i.e. it is required for early pattern-forming processes in the clypeolabrum but later it acts as a differentiation gene to establish a specific structural element. Thus, *croc* functions in pattern formation in a manner different from that of the prototype gene of the *fkh* domain gene family, *fkh*, which acts as a region-specific homeotic gene (Weigel *et al.*, 1989), and the recently identified gene pair *slp*, which combine gap and segment polarity gene functions in the *Drosophila* head region (Grossniklaus *et al.*, 1992, 1994). While the region-specific function of *fkh* is clearly different from *croc* function, a possible link between *croc* and *slp* function could be made if one assumes that an ancient gap gene function of *croc* is evolutionarily reduced and segment polarity function became restricted to the anterior-most head segment in the case of *croc*. In addition, there is also a possibility that *croc* function can be placed within the current scenario of known head genes of *Drosophila* on the basis of the initial expression pattern and its control during the preblastoderm stage.

croc is also expressed in the posterior region forming the posterior spot, and therefore vaguely reminds one of the expression patterns of the terminal gap genes such as *hkb* and *tll* (Pignoni *et al.*, 1990; Brönner *et al.*, 1994). Because we have not found a corresponding defect in *croc* mutant embryos, *croc* may carry a redundant or no function in the posterior expression domain. One can therefore imagine that *croc* or an ancestor gene might have served as a terminal gap gene which by now has adopted specific functions in the anterior pole region of *Drosophila*. In this context, it is worth mentioning that the clypeolabrum of *Drosophila* derives from two distinct units which are still separable in more primitive dipterans (Snodgrass, 1935). Alternatively, *croc* may represent a founding member of a group of head genes which has escaped discovery so far.

DNA binding impaired by a single-site mutation

The *croc fkh* domain was found by an *in vitro* target site selection to bind to a set of recognition sites encompassing 16 bp. All targets share a 7 bp consensus sequence. A highly similar core motif has been found for the DNA-binding regions from rodent, human and *Xenopus* winged helix proteins (Overdier *et al.*, 1994; Pierrou *et al.*, 1994; Kaufmann *et al.*, 1995). It is interesting to note that different *fkh* domains displaying a wide range of amino acid sequence diversity share a rather similar recognition core sequence. This may be explained by the fact that those residues of the HNF-3γ/*fkh* domain involved in direct base contacts (Clark *et al.*, 1993) are highly conserved. However, this 7 bp sequence on its own is necessary but not sufficient for DNA binding (Kaufmann *et al.*, 1995). It is likely that regions flanking the core motif are not only required for binding but also confer recognition specificity to particular proteins. The single amino acid replacement (Leu → Phe) at position 122 of *croc* protein prevents the binding of the corresponding *fkh* domain to DNA. The resulting lack-of-function mutant phenotype of this mutation, the nuclear localization and the lack of *en* expression in the clypeolabral segment anlage are therefore consistent with the argument that *croc* functions as a transcriptional regulator which has lost the ability to interact with its specific target genes in the mutant condition. Therefore, the amino acid replacement must have caused an alteration within an essential region of the *fkh* domain. Based on the structure of the *fkh* domain which was obtained with HNF-3γ, the replaced leucine is the first of four conserved amino acid residues which are involved in metal binding (Clark *et al.*, 1993). The divalent Mg^{2+} is coordinated by four carbonyl oxygen atoms from the polypeptide backbone (including those of the replaced Leu171 and of Ser172, Asn174 and Phe177) and two appropriately placed water molecules, which give a total of six occupied coordination sites (Clark *et al.*, 1993). While any one replacement could provide a carbonyl oxygen atom as a coordination site, the size difference between Leu and its replacement Phe may interfere with the stabilization of the compact structure of the *fkh* domain by preventing the coordinate binding of Mg^{2+} which normally tethers the third α-helix to the N-terminus of the β-strand S2 (for details on the *fkh* domain structure see Clark *et al.*, 1993, figure 4). The interpretation that the steric effects caused by the size difference between Phe and Leu might interfere with the stability of the DNA-binding structure is consistent with the previous finding that a replacement of Phe177 and the two subsequent amino acids Val and Lys (by Val, Ala and Met, respectively) cause the lack of DNA binding of HNF-3γ in a manner analogous to the *croc* mutant protein (Clevidence *et al.*, 1993). Whether this steric effect prevents the water molecules becoming included in the *fkh* domain structure or whether the replacement affects the β-strand/α-helix positioning more directly remains to be solved by structural analyses.

Materials and methods

Fly stocks, mutagenesis and cuticle preparations

The three *croc* alleles (*th ri* $croc^{75\text{-}3}$ p^{p}/*TM3 Sb*; *ru* $croc^{5f59}$ *cu ca*/*TM3 Sb*; *ru st* $croc^{UH}$ *e ca*/*TM3 Sb*) were generated by EMS-induced mutagenesis, as described by Jürgens *et al.* (1984). The stock *Df(3R)* Pc^{CP1} was kindly provided by M.Ashburner. Cuticles were prepared for microscopic inspection as described by Nüsslein-Volhard *et al.* (1984).

Molecular characterization of the croc locus

The λ-phage encompassing 18.3 kb of DNA containing the FD1 *fkh* domain was isolated by screening a genomic *Drosophila*-Canton S DNA library (kindly provided by D.Tautz) prepared in the λ-Fix vector (Stratagene) using as a probe a 6.3 kb *Eco*RI fragment cloned in the original isolation of FD1 (Häcker *et al.*, 1992). Using the same probe, two identical cDNA clones were isolated from a 0–4 and 4–8 h cDNA library. Both libraries were prepared in the vector pNB40 (Brown and Kafatos, 1988). All DNA fragments isolated from phages or cDNA

clones were subcloned into Bluescript vectors (Stratagene). DNA was sequenced by the dideoxynucleotide procedure of Sanger *et al.* (1977). Genomic or cDNA fragments were subcloned into M13 mp18 or mp19 vectors, and sequencing was carried out on both strands. All recombinant DNA techniques have been described by Sambrook *et al.* (1989).

Analysis of expression patterns

In situ hybridization of whole-mount embryos using the digoxigenin-labeled FD1 cDNA as a probe was performed according to Tautz and Pfeifle (1989), with modifications of the labeling reaction. A detailed protocol is available upon request. Antibody stainings were performed as described previously (Macdonald and Struhl, 1986) using the Vectastain ABC Elite horseradish peroxidase system and the modifications described in Weigel and Jäckle (1990). The polyclonal rabbit anti-β-galactosidase antibody was obtained from Cappel. Double-label *in situ* hybridizations to whole-mount embryos using digoxigenin-labeled *wg* cDNA and biotin-labeled *en* cDNA as probes were carried out as described recently (Hartmann and Jäckle, 1995).

Molecular characterization of croc mutant alleles

For each of the analyzed *croc* alleles, genomic DNA was prepared from ~10 individual embryos identified as homozygous mutant for the *croc* locus according to their cuticle phenotype. Using this DNA as a template, the coding sequence of the *croc* gene was amplified by PCR (Saiki *et al.*, 1988) using the deoxyoligonucleotides 5′-CGGAATCGCCGCCATGCATACGTTATCAGCG-3′ and 5′-CTGAGCAATCTGAGCCACTACCGACCGCA-3′ flanking the *croc* coding region as primers. PCR products were subcloned into M13 vectors and sequenced. To exclude mistakes resulting from the amplification reaction, all experiments were performed in triplicate.

Construction of croc–lacZ fusion genes

A *Xba*I–*Nsi*I fragment, including 75 bp of the *croc* leader and 9 kb of upstream sequence, was cloned into the vector pCaSpeR hs43 βGal (Thummel and Pirrotta, 1992). A 4 kb *Xba*I–*Bgl*II subfragment of the above 9 kb fragment was cloned into the same vector. From the region 3′-adjacent to the *croc* transcription unit, a 7 kb *Bgl*II fragment, including 396 bp of the *croc* 3′-untranslated region, was also cloned into pCaSpeR hs43 βGal. All three recombinant plasmids were introduced into the genome by germline transformation. Several independent transformant lines were established for each construct, and expression of the *lacZ* gene was analyzed by anti-β-galactosidase antibody staining.

Preparation of anti-croc antibodies

A 1450 bp genomic *Nsi*I–*Pvu*II fragment encoding amino acids 1–482 of the putative *croc* protein was cloned into the *E.coli* expression vector pRSET (Invitrogen), allowing the production of a histidine-tagged fusion protein. The *E.coli* strain BL21(DE3) carrying the recombinant plasmid expressed a novel protein of the predicted size of 56 kDa. This protein was purified on ProBond™ Resin (Invitrogen) according to the manufacturer's protocol and analyzed by SDS–PAGE. Polyacrylamide slices containing 200 mg of recombinant protein were cut from the gel, air dried for several days and sent to Eurogentec (Seraing, Belgium) for the immunization of rabbits. Polyclonal anti-*croc* serum was enriched from total serum of immunized rabbits by immunoaffinity chromatography on Affigel 10/Affigel 15 (3:1) matrix (Bio-Rad) according to the manufacturer's protocol.

Expression and purification of the croc fkh domain

A 1.6 kb DNA fragment encompassing the coding sequence for the wild-type or mutant protein was obtained by the PCR-supported amplification of various *croc* alleles and subsequently cloned into the vector pCRII. Again, a region corresponding to the *croc* wild-type or mutant *fkh* domains spanning residues A(67) to K(180) was amplified from the recombinant pCRII plasmids employing specific primers, permitting the integration of the product into the *E.coli* expression vector pRSET. After the careful assessment of the DNA sequences, the expression of proteins was induced by 1 mM IPTG and continued for 1.5 h at 37°C. Purification was performed by affinity chromatography on a nickel trinitrilo acetic acid matrix, as described previously (Hoffmann and Roeder, 1991).

In vitro target site selection for the fkh domain of croc

A deoxyoligonucleotide, encompassing 16 consecutive residues of random sequence flanked by two regions of 17 bases each (SK and reverse complement of KS primer: 5′-TCTAGAACTAGTGGATC-3′ and 5′-CGAATACCGTCGACCCCG-3′), was converted into the double-stranded form and subsequently labeled with ^{32}P (Blackwell and Weintraub, 1990). 400 ng of purified *croc fkh* domain were incubated with ~0.5 ng of the gel-purified probe. After electrophoresis, the fragments corresponding to the retarded protein–DNA complex were excised, and the eluted DNA was purified and amplified after prior optimization of Mg^{2+} concentration by PCR employing KS and SK primers. A PCR without template did not produce a signal. The amplified DNA was labeled and used again in binding and amplification. Subsequent rounds of enrichment for target sequences were performed in the same manner, but with diminished amounts of protein to select for high-affinity sites (Gogos *et al.*, 1992). Finally, the DNA was cloned into the Bluescript vector (Stratagene) and target sequences were determined on an ABI 373 sequencer.

All gel retardation experiments were performed at 4°C on 7% polyacrylamide gels in 0.5× TBE. The protein–DNA complex was allowed to form within a period of 30 min in a total volume of 30 ml 20 mM Tris (pH 7.5), 50 mM KCl, 1 mM β-mercaptoethanol, 1 mM $MgCl_2$ and 10% (v/v) glycerol. For DNase footprinting studies, the target site of interest was excised from the recombinant Bluescript vector and labeled on one strand by a fill-in reaction with [α-^{32}P]dCTP and Klenow DNA polymerase. Formation of the protein–DNA complex was achieved under the same conditions as described above. The concentration of Mg^{2+} was finally raised to 5 mM, and 50–100 ng DNase were added at room temperature for up to 90 s. Hydrolysis was stopped and the products were subjected to denaturing PAGE (Galas and Schmitz, 1978).

Acknowledgements

We would like to thank our colleagues in the two laboratories for their various contributions during all parts of the work. We thank especially S.Hille, G.Dowe, H.Taubert and B.Guilliard for skilful technical assistance. The work was supported by grants from the Deutsche Forschungsgemeinschaft (SFB 271 to H.J. and Kn 200/4-1 to W.K.).

References

Akam,M. (1987) The molecular basis for metameric pattern in the *Drosophila* embryo. *Development*, **101**, 1–22.

Ang,S.-L. and Rossant,J. (1994) HNF3β is essential for node and notochord formation in mouse development. *Cell*, **78**, 561–574.

Blackwell,T.K. and Weintraub,H. (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science*, **250**, 1104–1110.

Brönner,G., Chu-LaGraff,Q., Doe,C.Q., Cohen,B., Weigel,D., Taubert,H. and Jäckle,H. (1994) Sp1/egr-like zinc-finger protein required for endoderm specification and germ layer formation in *Drosophila*. *Nature*, **369**, 664–668.

Brown,N.H. and Kafatos,F.C. (1988) Functional cDNA-libraries from *Drosophila* embryos. *J. Mol. Biol.*, **203**, 425–437.

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

Clark,K.L., Halay,E.D., Lai,E. and Burley,S.K. (1993) Co-crystal structure of the HNF-3/*fork head* DNA-recognition motif resembles histone H5. *Nature*, **364**, 412–420.

Clevidence,D.E., Overdier,D.G., Tao,W., Qian,X., Pani,L., Lai,E. and Costa,R.H. (1993) Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/*forkhead* DNA-binding-domain family. *Proc. Natl Acad. Sci. USA*, **90**, 3948–3952.

Cohen,S.M. and Jürgens,G. (1991) *Drosophila* headlines. *Trends Genet.*, **7**, 267–272.

Dirksen,M.L. and Jamrich,M. (1992) A novel, activin-inducible, blastopore lip-specific gene of *Xenopus laevis* contains a *fork head* DNA-binding domain. *Genes Dev.*, **6**, 599–608.

Galas,D. and Schmitz,A. (1978) DNase footprinting: a simple method for the detection of protein–DNA binding specificity. *Nucleic Acids Res.*, **5**, 3157–3170.

Gehring,W.J. (1987) Homeo boxes in the study of development. *Science*, **236**, 1245–1252.

Gill,G. and Tjian,R. (1992) Eukaryotic activators associated with the TATA box binding protein. *Curr. Opin. Genet. Dev.*, **3**, 234–241.

Gogos,J.A., Hsu,T., Bolton,J. and Kafatos,F.C. (1992) Sequence discrimination by alternatively spliced isoforms of a DNA binding zinc finger domain. *Science*, **257**, 1951–1955.

Grossniklaus,U., Pearson,R.K. and Gehring,W.J. (1992) The *Drosophila sloppy paired* locus encodes two proteins involved in segmentation

that show homology to mammalian transcription factors. *Genes Dev.*, **6**, 1030–1051.

Grossniklaus,U., Cadigan,K.M. and Gehring,W.J. (1994) Three maternal systems cooperate in the patterning of the *Drosophila* head. *Development*, **120**, 3155–3171.

Häcker,U., Grossniklaus,U., Gehring,W.J. and Jäckle,H. (1992) Developmentally regulated *Drosophila* gene family encoding the *fork head* domain *Proc. Natl Acad. Sci. USA*, **89**, 8754–8758.

Hartenstein,V., Technau,G.M. and Campos-Ortega,J.A. (1985) Fate-mapping in wild-type *Drosophila melanogaster*. III. A fate map of the blastoderm. *Roux's Arch. Dev. Biol.*, **194**, 213–216.

Hartmann,C. and Jäckle,H. (1995) Spatiotemporal relationships between a novel *Drosophila* stripe expressing gene and known segmentation genes by simultaneous visualization of transcript patterns. *Chromosoma*, in press.

Hoch,M. and Jäckle,H. (1993) Transcriptional regulation and spatial patterning in *Drosophila*. *Curr. Opin. Genet. Dev.*, **3**, 566–573.

Hoffmann,A. and Roeder,R.G. (1991) Purification of His-tagged protein in non-denaturing conditions suggests a convenient method for protein interaction studies. *Nucleic Acids Res.*, **19**, 6337–6338.

Ingham,P.W. (1988) The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature*, **335**, 744.

Johnson,P. and McKnight,S.L. (1989) Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.*, **58**, 799–839.

Jürgens,G., Wieschaus,E., Nüsslein-Volhard,C. and Kluding,H. (1984) Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. II. Zygotic loci on the third chromosome. *Roux's Arch. Dev. Biol.*, **193**, 283–295.

Kaestner,K.H., Lee,K.H., Schlöndorff,J., Hiemisch,H., Monaghan,A.P. and Schütz,G. (1993) Six members of the mouse *forkhead* gene family are developmentally regulated. *Proc. Natl Acad. Sci. USA*, **90**, 7628–7631.

Karpen,G.H. and Spradling,A.C. (1992) Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics*, **132**, 737–53.

Kaufmann,E., Müller,D. and Knöchel,W. (1995) DNA recognition site analysis of *Xenopus* winged helix proteins. *J. Mol. Biol.*, **248**, 239–254.

Knöchel,S., Lef,J., Clement,J., Klocke,B., Hille,S., Köster,M. and Knöchel,W. (1992) Activin A induced expression of a *fork head* related gene in posterior chordamesoderm (notochord) of *Xenopus laevis* embryos. *Mech. Dev.*, **38**, 157–165.

Lai,E., Prezioso,V.R., Smith,E., Litvin,O., Costa,R.H. and Darnell,J.E. (1990) HNF-3α, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes Dev.*, **4**, 1427–1436.

Lewin,B. (1990) Commitment and activation at polII promoters: a tail of protein–protein interactions. *Cell*, **61**, 1161–1164.

Li,J. and Vogt,P.K. (1993) The retroviral oncogene *qin* belongs to the transcription factor family that includes the homeotic gene *fork head*. *Proc. Natl Acad. Sci. USA*, **90**, 4490–4494.

Lindsley,D.L. and Zimm,G.G. (1992) *The Genome of Drosophila melanogaster*. Academic Press, San Diego, CA.

Macdonald,P. and Struhl,G. (1986) A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. *Nature*, **324**, 537–545.

Martinez-Arias,A. (1993) Development and patterning of the larval epidermis of *Drosophila*. In Martinez-Arias,A. and Bate,M. (eds), *The Development of Drosophila*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 517–608.

Miura,N., Wanaka,A., Tohyama,M. and Tanaka,K. (1993) MFH-1, a new member of the *fork head* domain family, is expressed in developing mesenchyme. *FEBS Lett.*, **326**, 171–176.

Nehls,M., Pfeifer,D., Schorpp,M., Hedrich,H. and Boehm,T. (1994) New member of the winged-helix protein family disrupted in mouse and rat nude mutations. *Nature*, **372**, 103–107.

Nüsslein-Volhard,C., Wieschaus,E. and Kluding,H. (1984) Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. I. Zygotic loci on the second chromosome. *Roux's Arch. Dev. Biol.*, **193**, 267–282.

O'Connell,P. and Rosbash,M. (1984) Sequence, structure and codon preference of the *Drosophila* ribosomal protein 49 gene. *Nucleic Acids Res.*, **12**, 5495–5513.

Overdier,D.G., Porcella,A. and Costa,R.H. (1994) The DNA-binding specificity of the hepatocyte nuclear factor 3/*forkhead* domain is influenced by amino-acid residues adjacent to the recognition helix. *Mol. Cell. Biol.*, **14**, 2755–2766.

Pankratz,M.J. and Jäckle,H. (1993) Blastoderm segmentation. In Martinez-Arias,A. and Bate,M. (eds), *The Development of Drosophila*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 467–516.

Pierrou,S., Hellqvist,M., Samuelsson,L., Enerbäck,S. and Carlsson,P. (1994) Cloning and characterization of seven human *fork head* proteins: binding site specificity and DNA bending. *EMBO J.*, **13**, 5002–5012.

Pignoni,F., Baldarelli,R.M., Steingrimsson,E., Diaz,R.J., Patapoutian,A., Merriam,J.R. and Lengyel,J.A. (1990) The *Drosophila* gene *tailless* is expressed at the embryonic termini and is a member of the steroid receptor superfamily. *Cell*, **62**, 151–163.

Roeder,R.G. (1991) The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem.*, **16**, 402–407.

Ronchi,E., Treisman,J., Dostatni,N., Struhl,G. and Desplan,C. (1993) Down-regulation of the *Drosophila* morphogen *bicoid* by the *torso* receptor-mediated signal transduction cascade. *Cell*, **74**, 347–355.

Rothe,M., Pehl,M., Taubert,H. and Jäckle,H. (1992) Loss of gene function through rapid mitotic cycles in the *Drosophila* embryo. *Nature*, **359**, 156–159.

Ruiz i Altaba,A. and Jessell,T.M. (1992) *Pintallavis*, a gene expressed in the organizer and midline cells of frog embryos: involvement in the development of the neural axis. *Development*, **116**, 81–93.

Ruiz i Altaba,A., Prezioso,V.R., Darnell,J.E. and Jessell,T.M. (1993) Sequential expression of HNF-3β and HNF-3α by embryonic organizing centers: the dorsal lip/node, notochord and floor plate. *Mech. Dev.*, **44**, 91–108.

Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) DNA sequencing with the chain terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Scheucher,M., Dege,P., Lef,J., Hille,S. and Knöchel,W. (1995) Transcription of four different *fork head*/HNF-3 related genes (XFD-4, 6, 9 and 10) in *Xenopus laevis* embryos. *Roux's Arch. Dev. Biol.*, **204**, 203–211.

Schmidt-Ott,U. and Technau,G.M. (1992) Expression of *en* and *wg* in the embryonic head and brain of *Drosophila* indicates a refolded band of seven segment remnants. *Development*, **116**, 111–125.

Scott,M.P., Tamkun,J.W. and Hartzell,G.W.I. (1989) The structure and function of the homeodomain. *Biochim. Biophys. Acta*, **989**, 25–48.

Simpson-Brose,M., Treisman,J. and Desplan,C. (1994) Synergy between the *hunchback* and *bicoid* morphogens is required for anterior patterning in *Drosophila*. *Cell*, **78**, 855–865.

Snodgrass,R.E. (1935) *Principles of Insect Morphology*. Cornell University Press, London, UK.

Sprenger,F. and Nüsslein-Volhard,C. (1992) *Torso* receptor activity is regulated by a diffusible ligand produced at the extracellular terminal regions of the *Drosophila* egg. *Cell*, **71**, 987–1001.

St Johnston,R.D. and Nüsslein-Volhard,C. (1992) The origin of pattern and polarity in the *Drosophila* embryo. *Cell*, **68**, 201–219.

Strähle,U., Blader,P., Henrique,D. and Ingham,P.W. (1993) *Axial*, a zebrafish gene expressed along the developing body axis, shows altered expression in cyclops mutant embryos. *Genes Dev.*, **7**, 1436–1446.

Tautz,D. and Pfeifle,C. (1989) A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. *Chromosoma*, **98**, 81–85.

Thummel,C. and Pirrotta,V. (1992) New pCaSpeR P-element vectors. *Dros. Inf. Serv.*, **71**, 150.

Tjian,R. and Maniatis,T. (1994) Transcriptional activation: a complex puzzle with few easy pieces. *Cell*, **77**, 5–8.

Volk,T. and VijayRaghavan,K. (1994) A central role for epidermal segment border cells in the induction of muscle patterning in the *Drosophila* embryo. *Development*, **120**, 59–70.

Weigel,D. and Jäckle,H. (1990) The *fork head* domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell*, **63**, 455–456.

Weigel,D., Jürgens,G., Küttner,F., Seifert,E. and Jäckle,H. (1989) The homeotic gene *fork head* encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. *Cell*, **57**, 645–658.

Weinstein,D.C., Ruiz i Altaba,A., Chen,W.S., Hoodless,P., Prezioso,V.R., Jessell,T.M. and Darnell,J.E. (1994) The winged-helix transcription factor HNF3β is required for notochord development in the mouse embryo. *Cell*, **78**, 575–588.

Received on June 19, 1995; revised on July 21, 1995