collier, a novel regulator of *Drosophila* head development, is expressed in a single mitotic domain

Michèle Crozatier, Denise Valle*, Laurence Dubois, Saad Ibnsouda† and Alain Vincent‡

Background: Segmentation of the *Drosophila* embryo is based on a cascade of hierarchical gene interactions that is initiated by maternal morphogens; these interactions define spatially restricted domains of zygotic gene expression within the blastoderm. Although the hierarchy of the segmentation genes that subdivide the trunk is well established, the patterning of the head is less well understood. Seven head segments can be assigned on the basis of metameric patterns of segment-polarity gene expression and internal sensory organs. The domains of expression of head gap-like genes broadly overlap, with their posterior margins out of phase by one segment. Taken together with the lack of pair-rule gene expression in the head, these observations led to the suggestion that head gap genes act in a combinatorial manner, determining head segmental borders and segmental identity at the same time.

Results: We have identified a new *Drosophila* gene, *collier* (*col*), whose expression at the blastoderm stage is restricted to a single stripe of cells corresponding to part of the intercalary and mandibular segment primordia, possibly parasegment 0. Reduction of *col* activity in early gastrula embryos by antisense RNA expression results in a specific lack of head structures derived from these segments. The expression of *col* coincides with a mitotic domain, which supports the proposal that cells in this domain undergo a concerted mitotic and differentiation program that is orchestrated at the transcriptional level. Col is an ortholog of mammalian early B-cell factor/Olfactory-1. These proteins define a new family of transcription factors that contain a helix–loop–helix dimerization motif and a new type of DNA-binding domain that is highly conserved during evolution.

Conclusions: Here we describe Col, the first *Drosophila* member of a new family of transcription factors. Col may act as a 'second-level regulator' of head patterning. The structural conservation of Col during evolution raises the questions of its conservation of function in head specification and its interactions with other factors conserved between insects and vertebrates.

Address: Centre de Biologie du Developpement, UMR 9925 CNRS/UPS, 118, route de Narbonne, 31062 Toulouse Cedex, France.

Present addresses: *Fundacao Oswaldo Cruz, Depto. Bioquimicia E Biologica Molecular, Av. Brasil 4365 Manguinhos CX Postal 926 Rio de Janeiro R.J. Brazil CEP 21040. †Departement de Biologie, Universite Sidi Mohamed Ben Abdellah, Faculte des Sciences Saiss, Fes, Maroc.

Temporary address (until August 1996): ‡Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA.

Correspondence: Alain Vincent.
E-mail: vincent@cict.fr

Received: 25 March 1996
Revised: 29 April 1996
Accepted: 29 April 1996

Current Biology 1996, Vol 6 No 6:707–718

 ISSN 0960-9822

Background

The organization of a repetitive body pattern is a fundamental aspect of embryonic development in many animals. In *Drosophila*, segmentation is based on a cascade of hierarchical gene interactions that is initiated by maternal morphogens. These gene interactions define spatially restricted domains of segmentation gene expression at the blastoderm stage. The gap, pair-rule and segment-polarity genes serially subdivide the trunk into reiterated units, and each unit acquires its identity through the action of homeotic genes (reviewed in [1–3]). Although the segmental organization of the embryonic head is morphologically obscured, the expression patterns of the segment-polarity genes *engrailed* (*en*) and *wingless* (*wg*) and the identification of sensory neurons indicate the existence of seven head segments [4–6]. Only the formation of the two posterior-most segments, the labial and maxillary, depends on the segmentation and homeotic genes used in the trunk. A different mechanism has been proposed for the formation of the antennal, intercalary and mandibular segments [4] on the basis of mutant phenotypes and patterns of expression of the head gap-like genes *orthodenticle* (*otd*), *empty spiracles* (*ems*) and *buttonhead* (*btd*). These three genes encode transcription factors that are expressed in partly overlapping stripes that roughly correspond to the head-segment anlagen affected in loss-of-function mutants [7–10]. A combinatorial mode of gap-like gene function could, in principle, partition the head anlage into a fixed number of segmental units and activate a unique pattern of homeotic gene expression in each segment primordium with no need for second-level regulators, such as pair-rule genes, in the trunk [4,11,12].

An intricate pattern of mitotic domains has also been described in the embryonic head [13]. Mitotic domains correspond to defined groups of newly formed cells that

enter mitosis 14 synchronously. The highly complex pattern of the mitotic domains reflects the transcription of *string* (*stg*), the *Drosophila* homolog of *Schizosaccharomyces pombe cdc25*, which releases cells from G2 arrest [14]. Although the developmental role of such a precise patterning of mitoses remains elusive, mitotic domains are thought to be an early manifestation of the commitment of cells to a specific developmental fate [13,15]. However, this view has not yet been substantiated because no regulatory gene expressed specifically in a single mitotic domain has, so far, been identified.

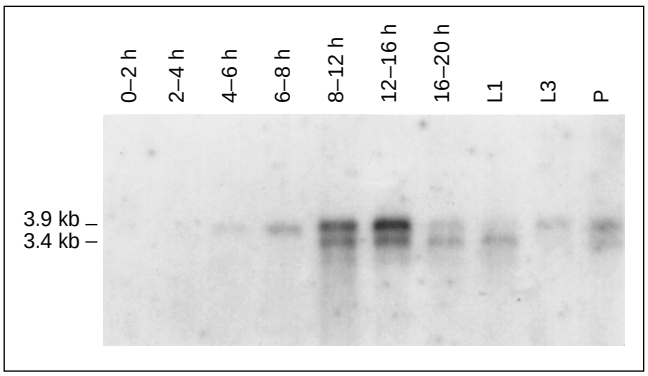
Here we report the isolation of a new *Drosophila* gene, *collier* (*col*), whose expression at the blastoderm stage in a single patch of cells coincides with a single mitotic domain (MD2) and overlaps part of the mandibular and intercalary segment primordia. This region of expression may correspond to parasegment 0, as drawn on the blastoderm fate map [16]. At the onset of gastrulation, *col* activity is required for the correct differentiation of head structures derived from these primordia. The predicted amino-acid sequence of the Col protein shows a striking homology to the mammalian transcription factor EBF (early B-cell factor; [17]), which was also independently isolated as Olf-1 (Olfactory-1, [18]). Col and EBF/Olf-1 define a new family of helix-loop-helix transcription factors. The high degree of structural conservation of Col during evolution suggests that this protein may be an ancestral head-patterning transcription factor which may interact with other evolutionarily conserved transcription factors that control head formation in both insects and vertebrates [19].

Results

collier, a *Drosophila* ortholog of the vertebrate transcription factor EBF/Olf-1

The *collier* gene was obtained from a molecular screen performed to isolate potential target genes of the Sry δ zinc finger transcription activator ([20]; see Materials and methods), and named from its early embryonic pattern of expression (see below). We obtained two classes of *col* cDNAs from 4–8 hour and 12–20 hour embryonic cDNA libraries. The cDNAs were 3.9 kb and 3.4 kb long, respectively, which correlated with the length of the two *col* transcripts detected on northern blots (Fig. 1 and data not shown). The 3.9 kb RNA was present from around 3 hours of embryogenesis, with a peak of accumulation between 8 and 16 hours post-fertilization; it persisted at very low level in first instar larvae and accumulated again in third instar larvae and pupae. The 3.4 kb transcript first accumulated after 8 hours of embryogenesis, peaked in first instar larvae and was present at low levels in third instar larvae and pupae (Fig. 1). Both *col* transcripts were detected at very low levels in male and female adults and were absent from dissected ovaries (data not shown). The 3.4 kb and 3.9 kb cDNAs differ from each other by 465 nucleotides (between positions 2098 and 2563), which are removed by a developmentally regulated alternative splicing event (Figs 1 and 2). The 3.9 kb cDNA contains an open reading frame of 1725 nucleotides, which predicts a protein (Col isoform 1) of 575 amino acids with a calculated molecular weight of 69 kDa (Fig. 2). The 3.4 kb cDNA contains an open reading frame of 1671 nucleotides, predictive of a second Col isoform of 557 amino acids with a calculated molecular weight of 60 kDa. The two Col isoforms have the same 528 amino-terminal amino acids but their sequences differ at the carboxy-terminal ends. In Col isoform 1, the carboxy terminal region of sequence divergence is 47 amino acids long, and in Col isoform 2 this region is 29 amino acids long (Fig. 2).

Figure 1



Developmental profile of *col* transcript accumulation. A developmental northern blot was probed with labeled 3.9 kb *col* cDNA. The developmental stage of embryonic RNA is denoted in hours after egg laying. L1 and L3 stand for first and third instar larvae, respectively; P stands for late pupae. Two *col* transcripts of 3.9 kb and 3.4 kb are detected, which vary in relative amounts throughout development. Each lane contains 2 μg of poly(A)$^+$ RNA. We used the rp49 probe [48] as an internal standard for quantitation of deposited RNA (data not shown).

Comparison of the predicted amino-acid sequence of Col with other protein sequences in the current databases revealed an extensive similarity with the vertebrate transcription factors EBF and Olf-1 (Fig. 3a). Although EBF was isolated as a mouse B-lymphocyte-specific transcription factor [17] and Olf-1 as a rat transcription factor regulating the expression of specific genes in olfactory neurons and their precursors [18], they probably represent products of homologous genes.

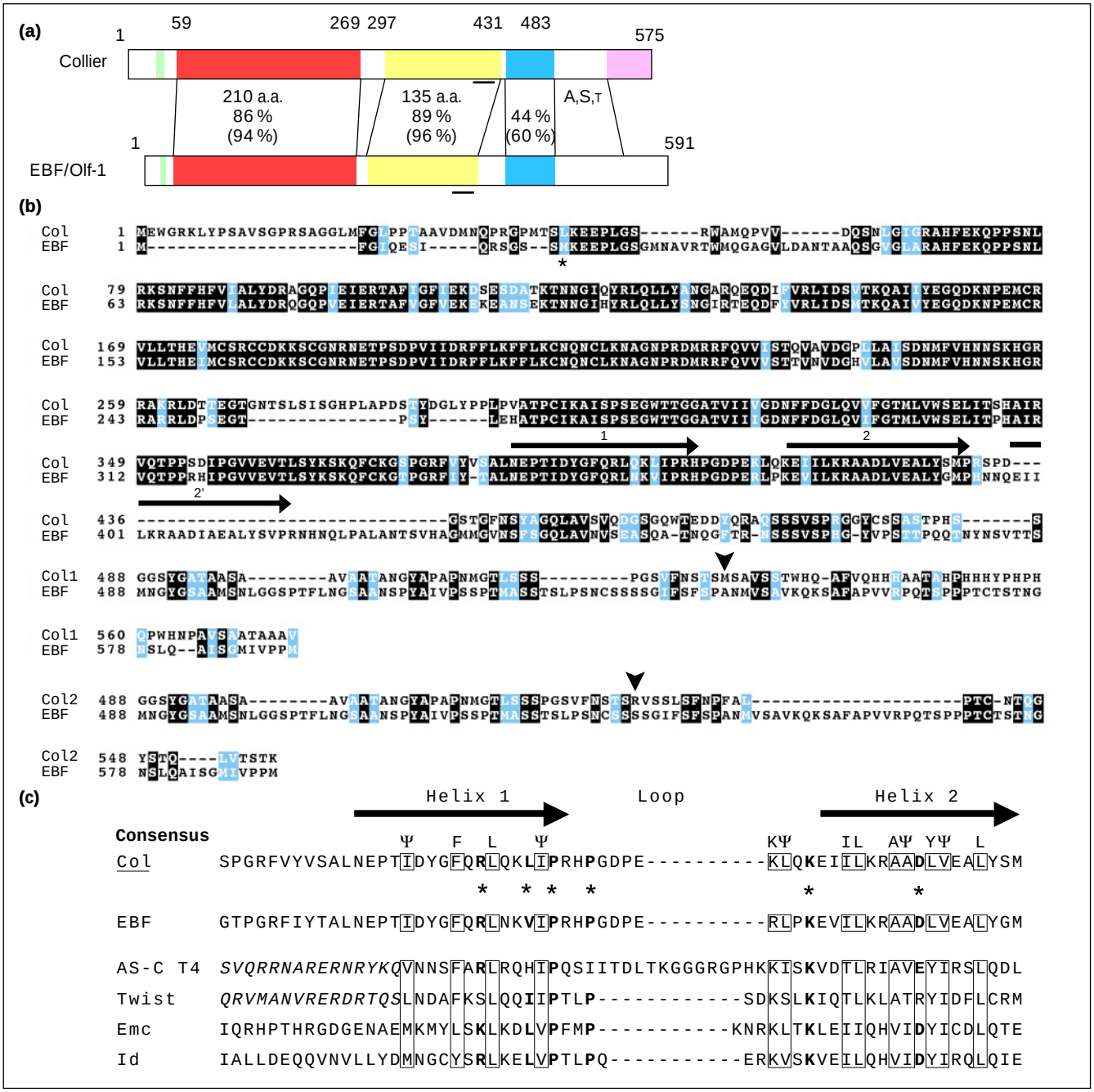
Col and EBF/Olf-1 define a new family of helix–loop–helix transcription factors

Two protein regions are particularly well conserved between Col and EBF/Olf-1 (Fig. 3a,b). The first one, which is 210 amino-acids long and lies between residues 59 and 269 of Col, shows 86 % identity (94 % similarity) and corresponds to the DNA-binding domain of EBF, which has been mapped biochemically [17,18]. The second

Figure 2

Sequence of *col*. The nucleotide sequence of the full-length 3.9 kb (3917 bp) *col* cDNA (EMBL accession number X97803) has an open reading frame which predicts a protein of 575 amino-acids (Col isoform 1). Amino acids are represented by the three-letter code. The sequence between nucleotides 2098 and 2563 (shown in italics) is absent from the 3.4 kb cDNA, this cDNA contains an open reading frame that predicts a protein of 557 amino acids (Col isoform 2). Col isoforms 1 and 2 are identical between amino-acid residues 1 and 528 and differ at their carboxy-terminal ends, which are underlined and labelled 1 and 2, respectively. Two potential polyadenylation signals (in bold letters) are located 30 and 14 nucleotides upstream of the polyadenylation site, respectively.

Figure 3



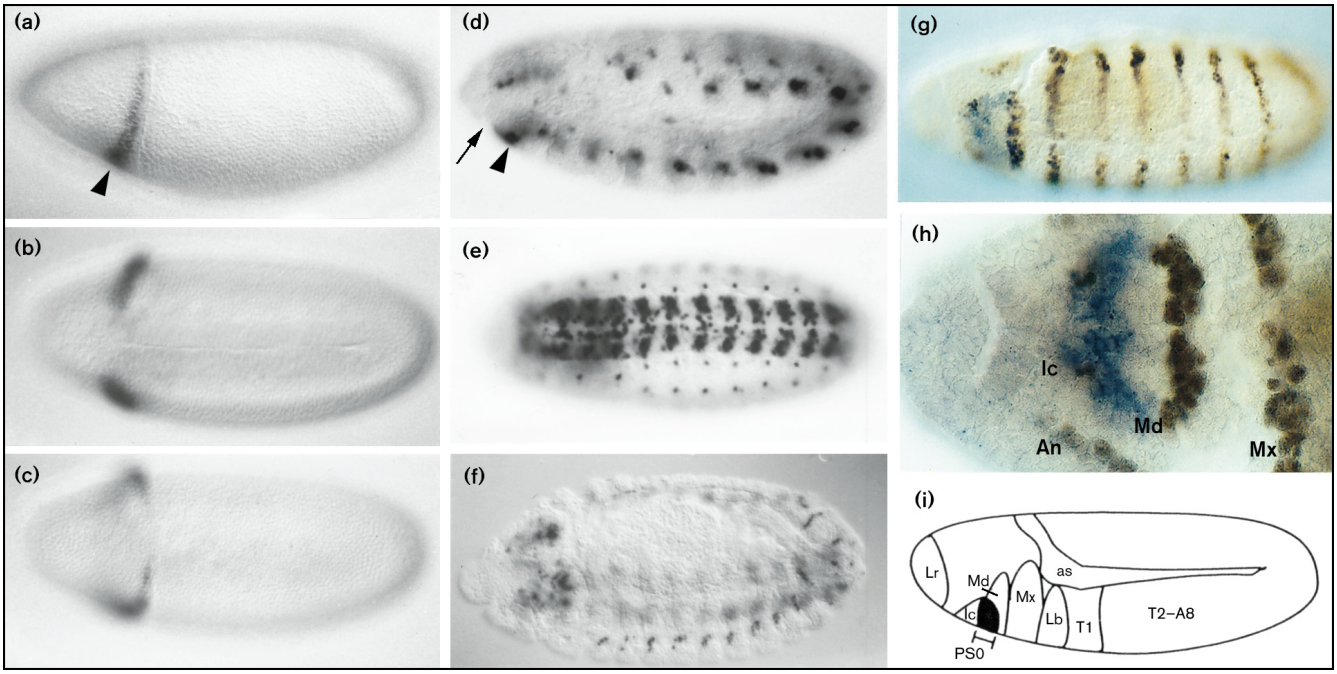
Col is a helix–loop–helix protein homologous to mammalian EBF/Olf-1 (a) Diagramatic alignment of the Col and EBF/Olf-1 proteins. The red box corresponds to the DNA-binding domain as defined in EBF [21] and the yellow box corresponds to a second highly conserved region which contains a helix–loop–helix motif indicated by the small black bar. The blue box represents a third region of significant homology. The percentages of identity and similarity (in brackets) are indicated. 'A,S,T' denotes the region rich in alanine, serine and threonine residues. The pink boxed area in the carboxy-terminal region of Col denotes the region where the two Col isoforms diverge. (b) Sequence alignment of Col and EBF/Olf-1. Identical and similar amino-acids are indicated by black and blue boxes, respectively. The DNA-binding domain of EBF (amino acids 50–251) includes a zinc-binding motif (between residues 151 and 177) [21]. The horizontal arrows indicate the positions of helix 1 and helix 2 of a predicted helix–loop–helix motif which is conserved between Col and EBF (see part (c)). The arrowhead indicates the position where the two Col isoforms diverge. The asterisk denotes the position of the methionine initiator in Olf-1. (c) Sequence alignment between Col, EBF and functional helix–loop–helix motifs in the following proteins: *Drosophila* AS-C T4 [49], Twist [50], Emc [51,52] and vertebrate Id [53]. The consensus motif is taken from [54]. The asterisk denotes residues (in bold) identical or similar in five out of the six helix–loop–helix proteins listed in addition to boxed consensus positions. ψ represents hydrophobic animo acids. The amino-acid sequence of the basic region of basic helix–loop–helix proteins is italicized.

region (Col residues 297 to 431) shows 89 % identity (96 % similarity) and partly overlaps a region of EBF sufficient for homodimerization *in vitro* [17]. We noticed a consensus helix-1–loop–helix-2 motif in this region, which was also conserved in the rodent proteins (Fig. 3c). Originally, no helix 1 but two potential helices 2 were reported in EBF and Olf-1 [17,18]. The second helix 2 is missing in Col. The helix–loop–helix dimerization motif is not preceeded by a basic region, consistent with the presence of an independent DNA-binding domain. The carboxy-terminal region is rich in alanine, serine and threonine residues and probably represents a transcription activation domain [21]. Except for the helix–loop–helix motif, no significant similarity was found between Col, EBF/Olf-1 and other described transcription factors. These two proteins therefore define a new family of helix–loop–helix transcription factors with a novel type of DNA-binding domain.

Spatial embryonic expression of col

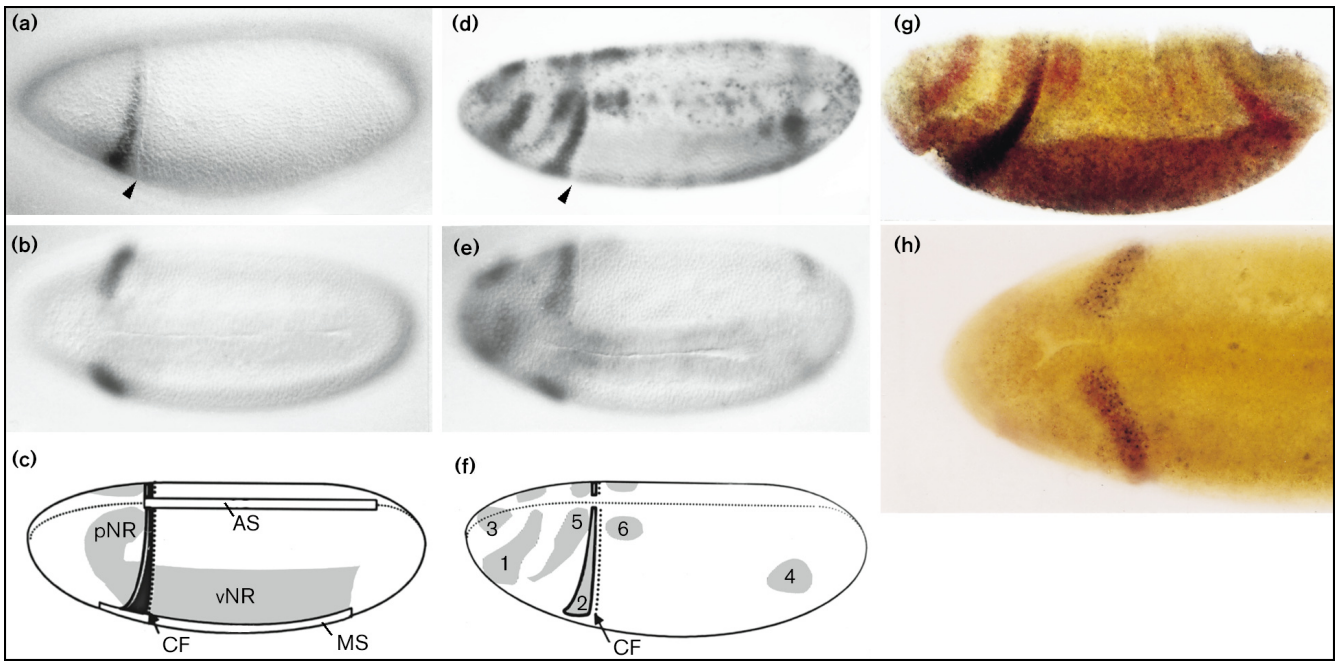
We first detected *col* expression at the beginning of mitotic cycle 14 in two laterally symmetrical stripes whose posterior limit was one or two rows of cells anterior to the position where the cephalic furrow forms, at about 70 % egg length (EL, measured from the posterior end of the embryo, [22]; data not shown). This restricted expression persisted during early gastrulation (Fig. 4a–c), when the cephalic furrow demarcates the head from the trunk and the posterior gnathal region (maxillary and labial primordia). At the same time, mesodermal precursor cells start to invaginate to form the ventral furrow, which runs longitudinally along the ventral side between 20 and 70 % EL. From dorsal to ventral, the stripe of *col* expression widened from one to four cells. However, neither the ventral-most mesodermal precursor cells (Fig. 4b) nor the dorsal-most amnioserosa precursor cells (Fig. 4c) expressed *col*. Comparison with the

Figure 4



col mRNA expression in wild-type embryos. **(a–f)** mRNAs were detected by whole-mount *in situ* hybridization. In all figures embryos are oriented with the anterior pole to the left. Stages are according to [22]. (a–c) Lateral, ventral and dorsal views of an early-gastrula stage embryo, respectively. *col* transcripts are detected as two laterally symmetrical stripes located anterior to the cephalic furrow (indicated by an arrowhead in (a). Neither the mesodermal precursor cells, a ventral band of cells approximately 18 cells wide, which undergo apical flattening [55] (visible in (b)), nor the amnioserosa precursor cells, which form a narrow stripe along the dorsal midline approximately five cells wide [22] (visible in (c)), express *col*. (d) Lateral view of a stage-11 embryo. Expression of *col* persists in the mandibular bud (arrowhead) which is now located posterior to the stomodeal invagination (arrow). Secondary expression sites include segmentally repeated groups of cells in the trunk and a few cells of the procephalon. (e) Ventral view of a stage-13 embryo after completion of germ-band shortening. *col* is now expressed in specific, segmentally reiterated lateral cells as well as in the ventral nerve chord. (f) Dorsal view of a stage-14 embryo shows *col* expression in a few cells in each brain hemisphere. **(g)** *In situ* hybridization to *col* mRNA (blue) and immunostaining with anti-En antibodies (brown) of an embryo at the germ-band extension stage (stage 10, ventral view). **(h)** Enlargement of the head region. En is expressed in the antennal (An), intercalary (Ic), mandibular (Md) and maxillary (Mx) segments; *col* is expressed in the anterior part of the mandibular segment and posterior part of the intercalary segment. **(i)** Schematic representation of an embryo at the germ-band extension stage, with the *col* expression domain indicated by the black area. Head-segments labral (Lr), intercalary (Ic), mandibulary (Md), maxillary (Mx), labial (Lb), trunk segments (T1, T2–A8), amnioserosa (as) and the position of putative parasegment 0 (PS0) are indicated.

Figure 5



Comparison of *stg* and *col* expression in early gastrula embryos. *In situ* hybridization with *col* (a,b) and *stg* (d,e) on an early-gastrula embryo. *stg* expression anticipates mitosis in mitotic domains (MDs) 1 to 6. MD2 is crescent-shaped and is anterior to the cephalic furrow (indicated by the arrowhead in (a,d)). (e) Ventral view showing the ventral limit of MD2. (c) Schematic fate map of the lateral view of a blastoderm-stage embryo, modified from [16]. MS; mesoderm; AS, amnioserosa; CF, cephalic furrow. The procephalic (pNR) and ventral (vNR) neurogenic regions are shaded grey. The *col* expression domain (thin black line) is shifted anteriorwards by one or two rows of cells compared to the mandibular segment anlage (shaded black) [16]. (f) Positions of MD1 to MD6 at the onset of gastrulation drawn from data shown in (a,d) and from [13]. MD2 (outlined in black) overlaps the *col* expression domain. (g,h) Double *in situ* hybridization with *stg* (red) and *col* (black) on early gastrula embryo. Lateral view (g); ventral view in detail (h). Black dots corresponding to *col* primary transcripts are detected in the cells that express *stg* in MD2. Not every nuclear dot signal is visible because of the focus.

blastoderm fate map [16,23] indicated that *col* expression extensively overlapped the mandibular segment anlage but was slightly displaced anteriorly (see Fig. 5c). By stage 11, the ventral invagination of the stomodeum has occurred and the germ band is almost fully extended, with the first signs of segmentation appearing in the epidermis. At this time, *col* transcripts were still present in the mandibular bud region but additional expression was now also seen in a few cells of the procephalon and in segmentally repeated groups of cells in the trunk (Fig. 4d). In stage 13 and 14 embryos, *col* was expressed in a segmentally reiterated pattern in the ventral nerve cord and in lateral and dorsal groups of cells, including cells of the peripheral nervous system (Fig. 4e), as well as in patches of cells in the brain (Fig. 4f).

In order to examine the relationship between expression of *col* and the segment anlagen in more detail, we compared the expression patterns of *col* and Engrailed (En) protein in germ-band extended embryos by a double *in situ*–immunostaining procedure. Each spot of En-expressing cells in the head is proposed to define the posterior limit of an individual segment [4,5]. Figure 4g–i shows that *col* mRNA was expressed in cells of the 'En intercalary spot' as well as in cells directly posterior to it. Based on this result and the results of *in situ* hybridizations on blastoderm stage embryos, *col* appears to be specifically expressed in posterior cells of the intercalary segment and in anterior cells of the mandibular segment, a region that possibly corresponds to parasegment 0 [23,24].

There is a striking similarity between the early stripe of *col* expression and the position of a specific mitotic domain at cycle 14, mitotic domain 2 (MD2). Mitotic domains are defined as groups of cells that enter mitosis 14 synchronously and out of synchrony with other groups of cells [13]. The pattern of *string* (*stg*) transcription anticipates the pattern of cycle 14 mitoses [14]. We therefore compared the expression of *stg* and *col* during early gastrulation, in either singly or doubly labelled embryos [25]. Figure 5 shows that, at the onset of gastrulation, *stg* and *col* were simultaneously expressed in a group of cells that correspond to MD2, suggesting that these cells not only share a mitotic fate, but also share a specific gene-expression program.

col is required for normal embryonic head morphogenesis

The *col* gene maps at 51B10-C1, between deficiencies Df(2R)Trix (51A1–51B6) and Df(2R)JP1 (51C3–52F5-9) (data not shown). There is currently no existing deficiency

Table 1

***col* phenocopies resulting from ectopic expression of *col* antisense RNA in wild-type embryos at the early gastrula stage.**

				Cuticles		
	Laid embryos	Hatched larvae	Undeveloped	Only head skeleton defects	Head skeleton defects and trunk defects	Only trunk defects
HscolA	67	1	19*	34	12	1
	196	2	34*	71	89	–
Total	263	3 (1 %)	53*	105 (40 %)	101 (38 %)	1
w^-	180	162	12	–	–	6
	240	219	18	–	–	3
Total	420	381 (91 %)	30	–	–	9

Embryos have either one copy of the transgene (HscolA) or no transgene (w^-). After 3 h of development at 25 °C, the embryos were heat-treated for 1 h at 37 °C. After heat treatment the embryos were left to develop at 25 °C and cuticles prepared. The asterisks indicate embryos which either did not develop a cuticle or were lost during cuticle preparation.

removing DNA in the 51B6–51C3 interval. Because we lacked *col* mutants, we attempted to eliminate, or at least reduce, *col* activity in early embryos by using an antisense RNA approach. To introduce antisense *col* RNA into the embryo, we made transgenic flies expressing this RNA under the control of the heat-shock-inducible hsp70 promoter (strain HscolA). A one-hour heat treatment at 37 °C applied between 4 and 6 hours of development had no observable effect on either control (w^-) or HscolA embryos, whereas a one-hour treatment between 3 and 4 hours resulted in specific defects in HscolA embryos. The control embryos developed normally, whereas almost all of the HscolA embryos failed to hatch. However, 80 % of the HscolA embryos developed to the point of making a cuticle, and in these embryos the only defects that we consistently observed were in the head skeleton (Table 1). A cuticle preparation of a wild-type first instar larva is shown in Figure 6a,c. In heat-treated HscolA embryos, the most prominent defect (seen in 80 % of the cuticles that were examined) was the absence or drastic reduction of the lateral gräten (Fig. 6b,d). All the other skeletal structures appeared normal. The lateral gräten are thought to originate from the mandibular segment [6,13]. To examine the consequence of HscolA expression earlier in development, we looked at the pattern of En expression in germ-band extended embryos. In heat-treated HscolA embryos, the En intercalary spot was either reduced (Fig. 6b) or was sometimes missing, whereas the mandibular En stripe appeared to be unaffected. Two HsColA transgenic lines were tested and gave similar results. Together with the expression data (Fig. 4), these results suggested that *col* function is specifically required for the formation of structures originating from part of the mandibular and intercalary segment primordia (the head region where *col* is expressed at the onset of gastrulation).

col expression is controlled by head gap genes

Development of head and thorax depends on the correct input of the localized maternal morphogen Bcd (review in [26]). The activity of *bcd* is required for zygotic activation of the head gap genes which subdivide the intermediate head region into partly overlapping domains along the antero-posterior axis [3,7,8,10,27]. The initial *col* expression domain was included in that of the head gap genes *btd* and *slp*, and was slightly posterior to that of *ems* (reviewed in [16]. *In situ* hybridization to *col* mRNA showed no detectable signal in embryos from *bcd* mothers or in *btd* homozygous mutant embryos. In *ems* mutant embryos, *col* expression was expanded ventrally to include mesodermal precursor cells, whereas normal expression was observed in embryos lacking both *slp* genes (*slp1* and *slp2*; data not shown). Taken together, these results indicate that *col* acts downstream of the head gap genes in the transcription regulatory cascade that patterns the anterior part of the *Drosophila* embryo head, with *btd* activity being absolutely required for *col* activation (Fig. 7). In *stg* mutant embryos, *col* expression during gastrulation was the same as that in wild-type embryos, indicating that the expression of *col* is independent of the mitotic program of cells in MD2.

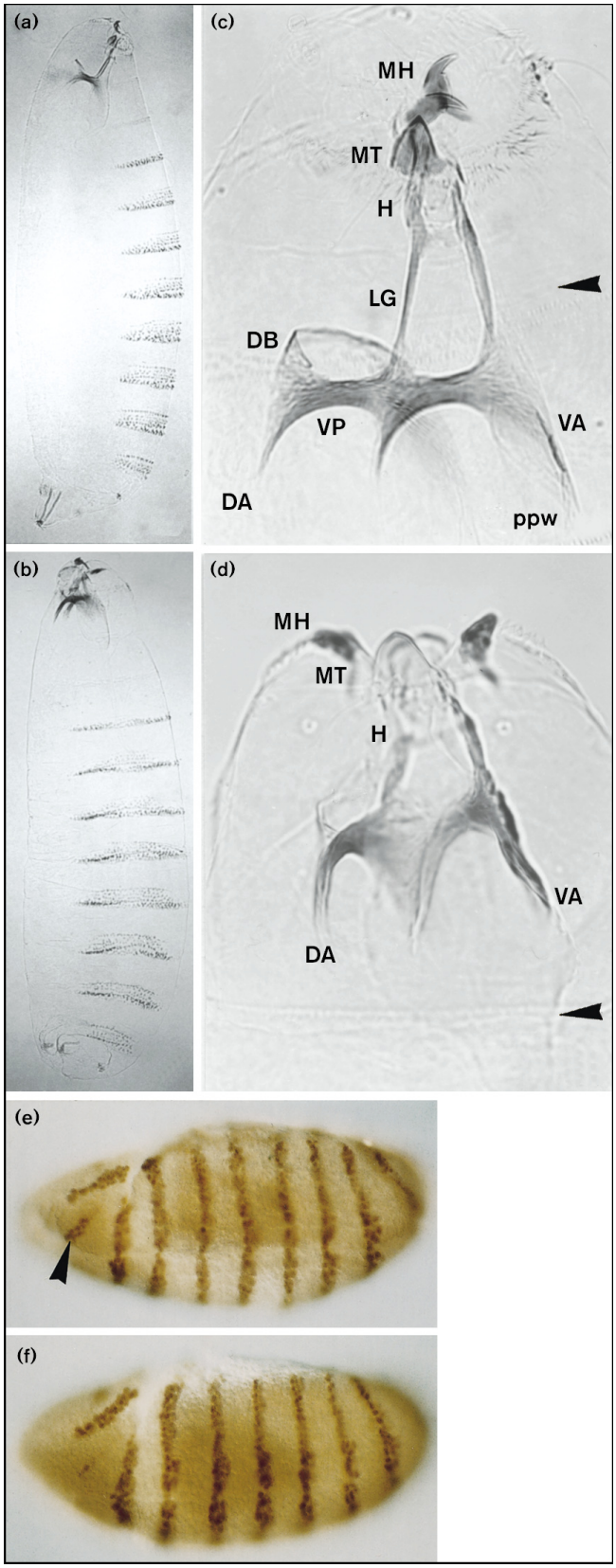
Discussion

The *Drosophila* embryonic head is composed of seven segments [6]. Only formation of the two posterior-most segments, the labial and maxillary, depends on the segmentation and homeotic genes used in the trunk. Formation of other segments has been proposed to depend on combinatorial inputs from head-specific gap genes expressed in partly overlapping domains at the blastoderm stage [4,11,12,16]. Here we report the identification and characterization of a new transcription-factor-encoding gene, *col*. We postulate that *col* represents a second-level regulator in the patterning of the embryonic head.

collier, a new head-patterning gene

Reducing *col* activity at the onset of gastrulation, through expression of antisense RNA, results in the specific absence of lateral gräten. The results from cell-ablation

experiments [23] suggested that this structure derives from the mandibular segment. The intercalary spot of



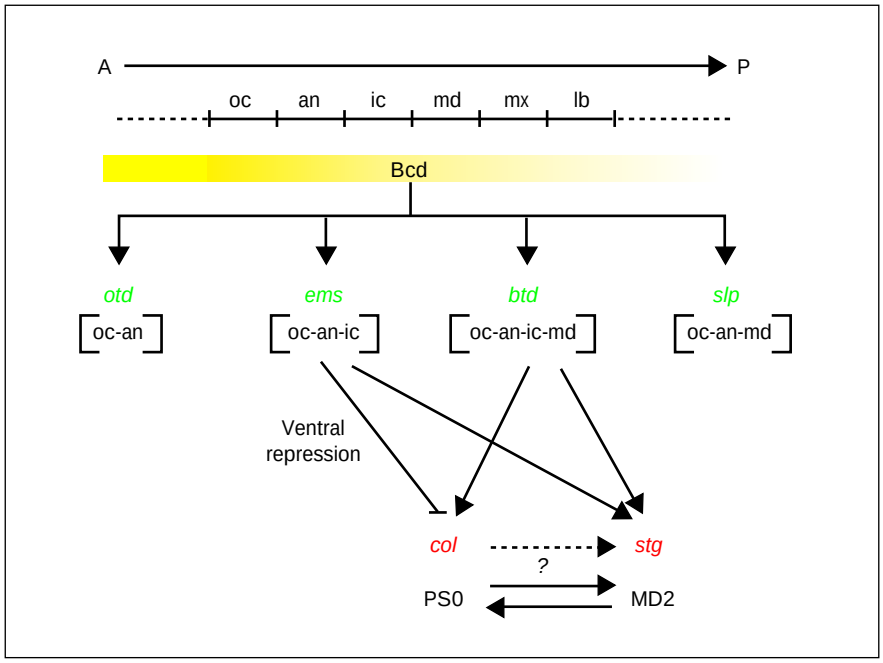
En-expressing cells is often partially missing in heat-treated HscolA embryos, whereas the mandibular spot is not affected. Both of these observations are consistent with the early pattern of *col* expression, which overlaps the anterior and posterior regions of the mandibular and intercalary segment primordia, respectively. The position, size and shape of the *col* stripe is sharply defined at early cycle 14, when *col* transcription is first observed; only neuroectodermal precursor cells express *col* — aminoserosa or mesodermal precursor cells do not.

The results described above indicate that information established along the anterior–posterior and dorsal–ventral axes prior to cycle 14 acts in concert to control *col* expression. This expression is dependent upon the activity of two head gap genes, *btd* and *ems*, and precedes expression of *en* and *wg* in the head. There is no change of *col* expression in embryos mutant for either the bZIP *cap'n'collar* (*cnc*) gene, which has been postulated to act as an homeotic selector in the mandibular parasegment [28], or the homeobox-containing genes of the *Hox* complex, *labial* and *Deformed*, which are expressed in the intercalary and mandibular plus maxillary segments, respectively [29,30] (data not shown). These results, taken together with the deletion of structures that is observed in HscolA embryos (which represent a subset of the defects seen in *btd* mutant embryos), lead us to suggest that *col* corresponds to a second-level regulator of embryonic head patterning that acts in parasegment 0. This study, together with the recent characterization of *crocodile* (a gene required for the formation of structures derived from the intercalary segment, the posterior wall of the pharynx and the ventral arm of the cephalopharyngeal skeleton [31]) indicates that a complex network of transcription factors acts downstream of head gap genes in controlling morphogenesis of the embryonic head. However, it seems that *crocodile (croc)* is not required for the establishment of the intercalary segment anlage *per se*, and the possible relationship between *col* and *croc* expression and function remains to be investigated.

Figure 6

col phenocopies produced by ectopic expression of *col* antisense RNA. **(a)** Wild-type cuticle from a w^- heat-treated embryo **(b)** Cuticle from a heat-treated HscolA embryo; note that head morphogenesis is abnormal, whereas the trunk is unaffected. **(c,d)** Enlargement of the head region of a wild-type (c) and a *col* phenocopy (d). The head skeleton structures, mouth hook (MH), median tooth (MT), H piece (H), lateral-gräten (LG), dorsal bridge (DB), vertical plate (VP), ventral arm (VA), dorsal arm (DA) and posterior pharynx wall (ppw) are indicated. In a *col* phenocopy the LG are missing. The arrowhead indicates the position of denticles corresponding to T2. **(e,f)** Expression of En in (e) heat-treated control (w^-) embryos and (f) HscolA embryos at the germband extension stage (stage 10). The intercalary spot of En-expressing cells (indicated by an arrowhead in (e)) is either reduced (f) or missing in heat-treated HscolA embryos.

Figure 7



Regulation of *col* expression by head gap genes. Data from [16,27] and experiments performed in our laboratory have been combined and are represented schematically. The relative positions of the head-segment anlagen along the anterio-posterior axis of the blastoderm stage embryo and the gradient of Bcd protein product are schematically represented at the top. The segments affected in lack-of-function mutants [4,27] are indicated under each head gap gene. *col* head-specific expression shows absolute requirement for *btd* and is modulated by *ems*. Expression of *stg* in MD2 also requires *btd* and *ems* ([28] and data not shown); potential regulation of *stg* by *col* has not been determined, as indicated by a broken arrow. oc, ocular; an, antennal; ic, intercalary; md, mandibular; mx, maxillary; lb, labial.

The difficulty in characterizing specific head defects might explain why no *col* mutant was isolated from the large-scale genetic screen performed on chromosome 2 by Nüsslein-Volhard *et al.* [32]. If the only ectodermal phenotype of *col* mutations is the head phenotype described here, *col* mutants would have been discarded, as they would have belonged to the small subclass (0.6 %) of embryonic lethal mutations that produced 'subtle' phenotypes, including defects in head skeleton differentiation.

Expression of *col* in a single mitotic domain: a transcriptional link between mitosis and morphogenesis?

In the *Drosophila* embryo, post-cellularization mitoses follow an invariant spatio-temporal pattern [13]. This pattern is governed by the expression of *stg* [14]. The transcription of *stg* itself is controlled by patterning genes through separate *cis*-regulatory regions that direct its expression in different mitotic domains [33]. Foe *et al.* [15] have proposed two models. The first proposes that *stg* regulatory sequences directly integrate pattern formation, and the second proposes that 'master regulatory genes' integrate this information and, in turn, control *stg* transcription. Because *col* encodes a transcription factor and is specifically expressed in a single mitotic domain, it could conceivably be a 'master gene' that controls the expression of *stg* in this domain. However, we favor another hypothesis, namely that *col* and *stg* respond to the same patterning information and act in parallel, with *col* assigning a specific gene-expression program to cells in MD2. In support of this, expression of both *stg* and *col* in MD2 is concomitant and specifically requires *btd* ([33] and data not shown). Conversely, expression of *col* in the head is unmodified in embryos mutant for *stg*, consistent with the observation that cell differentiation and morphogenetic events can occur in these embryos, even though they fail to produce a wild-type larval cuticle [13]. Furthermore, we detected *stg* expression in MD2 in heat-treated HscolA embryos (data not shown). The overlapping expression of *col* and *stg* in MD2 argues for a genetically orchestrated mitotic and differentiation schedule acting in concert for head morphogenesis.

A new family of transcription factors

Col shows extensive sequence similarity to the rodent transcription factor EBF/Olf-1. Unlike most cases of sequence conservation between *Drosophila* and vertebrate transcription factors, this similarity stretches over most of the Col and EBF/Olf-1 protein sequences. One highly conserved region corresponds to the unusually large DNA-binding domain of 210 amino acids defined in EBF; this domain also has homodimerization and transactivation potential and includes a novel zinc coordination motif essential for DNA recognition [21]. Thus far, the second highly conserved region, adjacent to the DNA-binding domain, has no precisely assigned function. It does, however, contain a sequence clearly related to the helix–loop–helix motif characteristic of the helix–loop–helix transcription factors, although this was not initially recognized in EBF/Olf-1 [17,18]. Helix–loop–helix motifs mediate homodimerization and/or heterodimerization [34].

Col and EBF/Olf-1 therefore define a new family of transcription factors that contain a novel type of DNA-binding domain associated with a helix–loop–helix motif; this motif is embedded in a larger region of 135 amino acids that is extremely well conserved between *Drosophila* and mouse. This structural conservation, in turn, strongly suggests that Col and EBF interact with other evolutionarily conserved factors and raises the question of the conservation of their functions from insects to vertebrates.

Are the functions of Col and EBF/Olf-1 conserved throughout evolution?

Recent analyses of basic developmental processes, such as neurogenesis and myogenesis, in different eukaryotes suggest that major developmental pathways and regulatory factors have been conserved throughout evolution [35]. The homeotic gene clusters are the most striking example of conservation between *Drosophila* and mammals of both genomic organization and related patterns of expression along the anterio-posterior axis [36]. The embryonic patterns of *Hox* gene expression in the central nervous system and branchial arches support the segmental structure of the hindbrain, as is morphologically apparent when the hindbrain is transiently subdivided into seven or eight rhombomeres. However, *Hox* genes are not expressed anterior to the midbrain–hindbrain boundary (reviewed in [37]), although there is circumstantial evidence that the forebrain might also be segmented [38]. This suggests that other classes of transcription factors are responsible for patterning of the more anterior head regions.

Although head specification was once thought to have arisen independently in the invertebrate and vertebrate lineages during evolution (discussed in [39]), the sequence conservation and related patterns of expression of the *otd* and *ems* (*otx* and *emx*) gene families in *Drosophila* and vertebrates suggests that patterning of the head was established earlier in a primitive ancestor [19]. Both mouse and *Xenopus col/EBF* orthologs show early localized expression in the neuroectoderm of the embryonic head (P. Charnay, personal communication; L.D., unpublished observations). This raises the possibility that the function of *col* in head patterning derives from an ancestral function. Other possible conserved functions for *col* might be related to its expression in specific sets of neurons in the central nervous system. Targeted disruption of EBF in mice leads to an early arrest of B-cell differentiation [40]. This tissue-specific phenotype occurs despite the fact that EBF is expressed in a variety of embryonic tissues ([18], P. Charnay, personal communication), suggesting that there is functional redundancy between EBF and closely related proteins in other tissues. In *Drosophila*, however, we did not find evidence for more than one member of the Col/EBF family (data not shown). Genetic analysis of *col* function and identification of interacting genes or genes that act upstream and downstream of *col* should help to elucidate a developmental role for this new family of transcription factors which are highly conserved between insects and vertebrates.

Conclusions

The hierarchy of segmentation genes which subdivides the trunk region of the early *Drosophila* embryo is well established. In contrast, patterning of the head region is less well understood. Here we report the identification of a new gene, *col*, whose expression at blastoderm is restricted to a narrow stripe of lateral cells overlapping part of the mandibular and intercalary segment primordia that possibly corresponds to parasegment 0. The activity of *col* is required at the onset of gastrulation for the correct formation of the mandibular segment. We propose that *col* may correspond to a second-level regulator of head morphogenesis. The expression of *col* matches a mitotic domain (MD2), suggesting the existence of a concerted mitotic and differentiation program in the cells of this domain at the transcriptional level. Col is the *Drosophila* ortholog of mammalian EBF/Olf-1, and together these proteins define a new family of transcription factors characterized by a novel type of DNA-binding domain associated with a helix–loop–helix motif. The extremely high degree of structural conservation of these proteins during evolution raises the issues of conservation of function and possible interactions with other transcription factors that control head formation in both insects and vertebrates.

Materials and methods

Fly strains

The btd^{XG81} and btd^{XA} mutant strains were obtained from S. Cohen, the *bcdE1*, ems^{9Q64}, stg^{4B51} and stg^{7B69} strains were from the Tübingen Stock Center and the $slp^{\Delta 34}$ deficiency strain was from U. Grossniklaus and W. Gehring. The *Df(2R)* Trix and *Df(2R)* JP1 deficiency strains were obtained from the Umea Stock Center.

Cloning and sequencing of col *cDNAs and northern blotting*

A small DNA fragment encoding an *in vitro* Sry δ binding site [20] was used to isolate 15 kb of surrounding genomic DNA. This DNA was used in turn to screen 4–8 h and 12–20 h embryonic cDNA libraries [41]. cDNA inserts were subcloned into Bluescript (Stratagene) and sequenced in both orientations using the exonuclease directional technique described in [42]. Isolation of poly(A^+) RNA from staged embryos and northern hybridization with *col* cDNA fragments were performed as described in [43].

Accession number

The EMBL accession number for the *col* sequence is X97803.

Heat-shock col *constructs and transgenic fly strains*

The 3.9 kb *col* cDNA was subcloned into a (w^+, pCas per) P-element vector [44] at the *Hpa*I site. DNA containing the *col* cDNA in an antisense orientation (HscolA construct) was injected into *white* (w^-) embryos and transformed fly strains were established according to standard methods [45]. Males homozygous for the HscolA transgene were crossed with w^- females, and w^- males were used as controls. The embryos resulting from these crosses were left to develop for 3 h at 25 °C before heat-shock treatment for 1 h at 37 °C. The embryos were then allowed to develop at 25 °C until they formed cuticles, which were prepared as described in [46].

In situ hybridization and antibody staining

Whole-mount *in situ* hybridization to embryos was performed with digoxigenin-labeled DNA prepared with the Genius kit from Boehringer Mannheim, using the procedure described in [47]. For double-labelling using *col* and *stg* probes simultaneously, we used the procedure described in [25]. Embryos were hybridized with a mixture of the fluorescein-labelled probe for *stg* and digoxigenin-labelled probe for *col*. *stg* hybridization was visualized by primary incubation with anti-fluorescein-alkaline phosphatase and staining with Vector Red (red). The *col* probe was revealed in a secondary incubation with anti-digoxigenin-alkaline phosphatase and staining with 4-nitro blue tetrazolium chloride/5-bromo-4-chloro-3 indolyl phosphate (black). To better distinguish between the two signals, we chose to use a nuclear *col* (intron) probe and a cytoplasmic *stg* (cDNA) probe (a gift from P. O'Farrell). For double immunostaining and *in situ* hybridization the embryos were first subjected to immunostaining as described in [30], using a polyclonal anti-Engrailed antibody prepared by F. Payre in our laboratory.

Acknowledgements

We are grateful to S. Cohen, U. Grossniklaus and W. Gehring, and the Tübingen and Umea *Drosophila* stock centers for mutant strains, and to R. Grosschedl and P. Charnay for communication of results prior to publication. We thank E. Wimmer for his help with re-examination of cuticle preparations, L. Bally-Cuif, C. Desplan and J. Smith for their comments on the manuscript and members of our laboratory for helpful suggestions and discussions throughout the course of this work. We also thank C. Ardourel for excellent technical assistance. This work was supported by CNRS, Association pour la Recherche sur le Cancer (ARC grant 6457) and a grant from Groupement de Recherches et Etudes du Génome (GREG). D. Valle was supported by fellowships from CNPq-Brasilia/Brasil and CNRS. S. Ibnsouda was supported by a fellowship from ARC and L. Dubois was supported by a fellowship from the Ministère de la Recherche et de l'Enseignement Supérieur (MRES).

References

1. Ingham PW: The molecular genetics of embryonic pattern formation in ***Drosophila***. *Nature* 1988, 335:25–34.
2. St Johnson D, Nüsslein-Volhard C: The origin of pattern and polarity in the ***Drosophila*** embryo. *Cell* 1992, 68:201–219.
3. Pankratz MJ, Jäckle H: Blastoderm segmentation. In *The development of* Drosophila melanogaster. Edited by Bate M, Martinez-Arias A. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1993:467–516.
4. Cohen S, Jürgens G: Mediation of ***Drosophila*** head development by gap-like segmentation genes. *Nature* 1990, 346:482–485.
5. Schmidt-Ott U, Technau GM: Expression of ***en*** and ***wg*** in the embryonic head and brain of ***Drosophila*** indicates a refolded band of seven segment remnants. *Development* 1992, 116:111–125.
6. Schmidt-Ott U, Gonzalez-Gaitan M, Jäckle H, Technau GM: Number, identity and sequence of the ***Drosophila*** head segments as revealed by neural elements and their deletion patterns in mutants. *Proc Natl Acad Sci USA* 1994, 91:8363–8367.
7. Finkelstein R, Perrimon N: The ***orthodenticle*** gene is regulated by ***bicoid*** and ***torso*** and specifies ***Drosophila*** head development. *Nature* 1990, 346:485–488.
8. Dalton D, Chadwick R, McGinnis W: Expression and embryonic function of ***empty spiracles***: a ***Drosophila*** homeobox gene with two patterning functions on the anterior–posterior axis of the embryo. *Genes Dev* 1989, 3:1940–1956.
9. Walldorf U, Gehring WJ: ***Empty spiracles***, a gap gene containing a homeobox involved in ***Drosophila*** head development. *EMBO J* 1992, 11:2247–2259.
10. Wimmer EA, Jäckle H, Pfeifle C, Cohen SM: A ***Drosophila*** homolog of human Sp1 is a head-specific segmentation gene. *Nature* 1993, 366:690–694.
11. Cohen S, Jürgens G: ***Drosophila*** headlines. *Trends Genet* 1991, 7:267–271.
12. Finkelstein R, Perrimon N: The molecular genetics of head development in ***Drosophila melanogaster.*** *Development* 1991, 112:899–912.
13. Foe VE: Mitotic domains reveal early commitment of cells in ***Drosophila*** embryos. *Development* 1989, 107:1–22.
14. Edgar BA, O'Farrell PH: The three post-blastoderm cell cycles of ***Drosophila*** embryogenesis are regulated in G2 by ***string***. *Cell* 1990, 62:469–480.
15. Foe VE, Odell GM, Edgar BA: Mitosis and morphogenesis in the ***Drosophila*** embryo: point and counter point. In *The development of* Drosophila melanogaster. Edited by Bate M, Martinez-Arias A. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1993:149–300.
16. Jürgens G, Hartenstein V: The terminal regions of the body pattern. In *The development of* Drosophila melanogaster. Edited by Bate M, Martinez-Arias A. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1993:687–746.
17. Hagman J, Belanger C, Travis A, Turck CW, Grosschedl R: Cloning and functional characterisation of early B-cell factor, a regulator of lymphocyte-specific gene expression. *Genes Dev* 1993, 7:760–773.
18. Wang MM, Reed RR: Molecular cloning of the olfactory neural transcription factor Olf-1 by genetic selection in yeast. *Nature* 1993, 364:121–126.
19. Finkelstein R, Boncinelli E: From fly head to mammalian forebrain: the story of ***otd*** and ***otx***. *Trends Genet* 1994, 10:310–315.
20. Payre F, Vincent A: Genomic targets of the sry β and sry δ zinc finger proteins and their respective DNA recognition sites. *EMBO J* 1991, 10:2533–2541.
21. Hagman J, Gutch MJ, Lin H, Grosschedl R: EBF contains a novel zinc coordination motif and multiple dimerization and transcriptional activation domains. *EMBO J* 1995, 14:2907–2916.
22. Campos-Ortega JA, Hartenstein: *The Embryonic Development of* Drosophila melanogaster. Berlin: Springer; 1985.
23. Jürgens G, Lehmann R, Schardin M, Nüsslein-Volhard C: Segmental organisation of the head in the embryo of ***Drosophila melanogaster***. *Wilhelm Roux's Arch Dev Biol* 1986, 195:359–377.
24. Martinez-Arias A, Lawrence PA: Parasegments and compartments in the ***Drosophila*** embryo. *Nature* 1985, 313:639–642.
25. Hauptmann G, Guster T: Two color whole mount in situ hybridisation to vertebrate and ***Drosophila*** embryos. *Trends Genet* 1994, 10:266.
26. Driever W: Maternal control of anterior development in the ***Drosophila*** embryo. In *The development of* Drosophila melanogaster. Edited by Bate M, Martinez-Arias A. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1993:301–325.
27. Grossniklaus U, Cadigan KM, Gehring WJ: Three maternal coordinate systems cooperate in the patterning of the ***Drosophila*** head. *Development* 1994, 120:3155–3171.
28. Mohler J, Mahaffey JW, Deutsch E, Vani K: Control of ***Drosophila*** head segment identity by the bZIP homeotic gene ***cnc***. *Development* 1995, 121:237–247.
29. Chadwick R, McGinnis W: Temporal and spatial distribution of transcripts from the ***Deformed*** gene of ***Drosophila***. *EMBO J* 1987, 3:779–789.
30. Mahaffey JW, Diederich RJ, Kaufman TC: Novel patterns of homeotic protein accumulation in the head of the ***Drosophila*** embryo. *Development* 1989, 105:167–174.
31. Häcker U, Kaufmann E, Hartmann C, Jürgens G., Knöchel W, Jäckle H: The ***Drosophila*** fork head domain protein crocodile is required for the establishment of head structures. *EMBO J* 1995, 14:5306–5317.
32. Nüsslein-Volhard C, Wieschaus E, Kluding H: Mutations affecting the pattern of the larval cuticle in ***Drosophila melanogaster***. I: zygotic loci on the second chromosome. *Roux's Arch Dev Biol* 1984, 193:267–282
33. Edgar BA, Lehman DA, O'Farrell PH: Transcriptional regulation of ***string (cdc25)***: a link between developmental programming and the cell cycle. *Development* 1994, 120:3131–3143.
34. Murre C, Schonleber Mc Caw P, Baltimore D: A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 1989, 56:777–783.
35. Jan YN, Jan LY: HLH proteins, fly neurogenesis and vertebrate myogenesis. *Cell* 1993, 75:827–830.
36. McGinnis W, Krumlauf R: Homeobox genes and axial patterning. *Cell* 1992, 68: 283–302.
37. Manak JR, Scott MP: A class act: conservation of homeodomain protein function. *Development Supplement* 1994, 61–71.
38. Rubenstein JLR, Martinez S, Shimamura K, Puelles L: The embryonic vertebrate forebrain: the prosomeric model. *Science* 1994, 266:578–580.
39. Gans C, Northcutt RG: Neural crest and the origin of vertebrate: a new head. *Science* 1983, 220:268–274.

40. Lin H, Grosschedl R: Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* 1995, 376:263–267.
41. Brown NH, Kafatos FC: Functional cDNA libraries from ***Drosophila*** embryos. *J Mol Biol* 1988, 203:425–437.
42. Henikoff S: Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods Enzymol* 1987, 155:156–165.
43. Vincent A, Kejzlarova-Lepesant J, Segalat L, Yanicostas C, Lepesant JA: ***sry h-1***, a new ***Drosophila melanogaster*** multifingered protein gene showing maternal and zygotic expression. *Mol Cell Biol* 1988, 8:4459–4468.
44. Thummel CS, Pirotta V: New pCaSpeR P-element vectors. *Drosophila Inf Service* 1992, 71:150.
45. Rubin G, Spradling A: Genetic transformation of ***Drosophila*** with transposable element vectors. *Science* 1982, 218:348–353.
46. Crozatier M, Kongsuwan K, Ferrer P, Merriam JR, Lengyel JA, Vincent A: Single amino-acid exchanges in separate domains of the ***Drosophila serendipity*** δ zinc finger protein cause embryonic and sex biased lethality. *Genetics* 1992, 131:905–916.
47. Tautz D, Pfeifle C: A nonradioactive ***in situ*** hybridization method for the localisation of specific RNAs in ***Drosophila*** embryos reveals a translational control of segmentation gene ***hunchback***. *Chromosoma* 1989, 98:81–85.
48. O'Connell PO, Rosbash M: Sequence, structure, and codon preference of the ***Drosophila*** ribosomal protein 49 gene. *Nucl Acids Res* 1984, 12:5495–5513.
49. Villares R, Cabrera CV: The achaete-scute gene complex of ***D. melanogaster***: conserved domains in a subset of genes required for neurogenesis and their homology to myc. *Cell* 1987, 50:415–424.
50. Thisse B, Stoetzel C, Gorostiza-Thisse C, Perrin-Schmitt F: Sequence of the ***twist*** gene and nuclear localisation of its protein in endomesodermal cells of early ***Drosophila*** embryos. *EMBO J* 1988, 7:2175–2183.
51. Ellis HM, Spann DR, Posakony JW: extramacrochaetae, a negative regulator of sensory organ development in ***Drosophila***, defines a new class of helix-loop-helix proteins. *Cell* 1990, 61:27–38.
52. Garrell J, Modolell J: The ***Drosophila extramacrochaetae*** locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop protein. *Cell* 1990, 61:39–48.
53. Riechmann V, Van Crüchten I, Sablitzk F: The expression pattern of Id4, a novel dominant negative helix–loop–helix protein, is distinct from Id1, Id2 and Id3. *Nucl Acids Res* 1994, 22:749–755.
54. Vinson CR, Garcia KC: Molecular model for DNA recognition by the family of basic-helix-loop-helix-zipper proteins. *The New Biologist* 1992, 4:396–403.
55. Costa M, Sweeton D, Wieschaus E: Gastrulation in ***Drosophila***: cellular mechanisms of morphogenetic movements. In *The development of* Drosophila melanogaster. Edited by Bate M, Martinez-Arias A. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1993:425–465.