The EMBO Journal vol.13 no.1 pp.168–179, 1994

spalt encodes an evolutionarily conserved zinc finger protein of novel structure which provides homeotic gene function in the head and tail region of the *Drosophila* embryo

Ronald P.Kühnlein, Götz Frommer, Markus Friedrich[1], Marcos Gonzalez-Gaitan, Alexander Weber[1], Juliane F.Wagner-Bernholz[2], Walter J.Gehring[2], Herbert Jäckle and Reinhard Schuh[3]

Max-Planck-Institut für biophysikalische Chemie, Abteilung Molekulare Entwicklungsbiologie, Am Fassberg, 37077 Göttingen, [1]Institut für Genetik und Mikrobiologie der Universität München, Maria-Ward-Str. 1a, 80638 München, Germany and [2]Biozentrum der Universität Basel, Abteilung Zellbiologie, Klingelbergstrasse 70, CH-4056 Basel, Switzerland
[3]Corresponding author

Communicated by H.Jäckle

The region specific homeotic gene *spalt* (*sal*) of *Drosophila melanogaster* promotes the specification of terminal pattern elements as opposed to segments in the trunk. Our results show that the previously reported *sal* transcription unit was misidentified. Based on P-element mediated germ line transformation and DNA sequence analysis of *sal* mutant alleles, we identified the transcription unit that carries *sal* function. *sal* is located close to the misidentified transcription unit, and it is expressed in similar temporal and spatial patterns during embryogenesis. The *sal* gene encodes a zinc finger protein of novel structure composed of three widely spaced 'double zinc finger' motifs of internally conserved sequences and a single zinc finger motif of different sequence. Antibodies produced against the *sal* protein show that *sal* is first expressed at the blastoderm stage and later in restricted areas of the embryonic nervous system as well as in the developing trachea. The antibodies detect *sal* homologous proteins in corresponding spatial and temporal patterns in the embryos of related insect species. Sequence analysis of the *sal* gene of *Drosophila virilis*, a species which is phylogenetically separated by ~60 million years, suggests that the *sal* function is conserved during evolution, consistent with its proposed role in head formation during arthropod evolution.

Key words: Drosophila embryogenesis/homeotic genes/*spalt* gene/transcription factors/zinc finger proteins

Introduction

Specification of segment identity in the trunk region of the *Drosophila melanogaster* embryo requires the activity of homeotic selector genes located within the Antennapedia (ANT-C) and the Bithorax (BX-C) complexes (Lewis, 1978; Kaufman *et al.*, 1980). Expression of the homeotic selector genes is initiated under the control of the segmentation gene cascade and spatially delimited by negative regulatory interactions between the different homeotic genes themselves, all of which encode homeodomain proteins likely to act as transcription factors (for reviews, see Gehring and Hiromi, 1986; Akam, 1987; Affolter *et al.*, 1990). The normal function of the *Antennapedia* (*Antp*) and BX-C genes depends on the activity of the gene *teashirt* (*tsh*) which is globally required for segmental identity throughout the entire trunk region (Röder *et al.*, 1992). At later stages of development the spatial expression domains of homeotic selector genes are maintained through the activity of members of the *Polycomb* (*Pc*) group of genes (Lewis, 1978; Jürgens, 1985).

An additional class of homeotic genes, the 'region specific homeotic genes', acts in the terminal regions of the embryo, specifying pattern elements in both the head and tail regions. *spalt* (*sal*) and *forkhead* (*fkh*), the two members of this class of homeotic genes, are located on different chromosomes outside the homeotic selector gene complexes (Jürgens, 1988; Jürgens and Weigel, 1988). In *fkh* mutants, ectodermal parts of the gut, i.e. the foregut and the hindgut, both develop as ectopic head structures. This suggests that the *fkh* gene promotes terminal as opposed to segmental development (Jürgens and Weigel, 1988). Mutations in the *sal* gene lead to incomplete transformations of pattern elements of the posterior head and the anterior tail towards the trunk, i.e. structures which are characteristic of the prothorax develop in the head, and structures of the eighth abdominal segment are formed in the tail region. These phenotypic effects within the head and the tail region of *sal* mutants seem to be very different, but double mutant analysis of *sal* and the homeotic selector gene *Abdominal-B* (*AbdB*) shows that *sal* activity promotes head as opposed to trunk development, i.e. *AbdB*/*sal* double mutants develop thoracic structures in place of the ectopic head structures found in the tail region of *AbdB* single mutant embryos (Jürgens, 1988). Furthermore, *sal* mutations cause inappropriate expression of the homeotic selector gene *Ultrabithorax* (*Ubx*) (Casanova, 1989) and hence *sal* may participate in the cross-regulatory interactions typical among other homeotic genes.

At the molecular level *fkh* has been shown to encode a DNA binding protein which is likely to act as a transcription factor with a conserved DNA binding motif, the '*forkhead*-domain' (Weigel *et al.*, 1989; Weigel and Jäckle, 1990). The gene, previously identified to carry *sal* function, encodes a small protein of 142 amino acids which lacks any known protein motif (Frei *et al.*, 1988). While *fkh*-related coding sequences and homeodomain proteins in particular have been identified in other insects as well as in vertebrates, the coding sequence of the previously identified *sal* gene was found to be conserved only in closely related *Drosophila* species (Reuter *et al.*, 1989). However, a basic genetic function which contributes to the separation of head and trunk segments should be conserved throughout insect evolution, since the basic separation of a primitive head from the segmented body region must have already occurred in myriapod-like ancestors of the recent insects (Jürgens, 1988).

© Oxford University Press

Based on this phylogenetic argument, the previously identified 142 amino acid protein was thought to have an accessory or modulating function for head development rather than representing the decisive gene product required to separate the terminal regions from the trunk (Reuter *et al.*, 1989).

Here we show that, in fact, *sal* function is not associated with the 142 amino acid protein. Based on *sal* mutant rescue by a transgene and *sal* mutant associated alterations of protein coding sequences, we present evidence that a zinc finger-type protein of novel structure provides *sal* function in *D.melanogaster*. The *sal* gene product and its expression pattern are conserved in other dipteran species.

Results

sal function has been mapped within 120 kb of DNA encompassing the chromosomal region 32F/33A on the left arm of the second chromosome (Frei *et al.*, 1988; Jürgens, 1988), and it has been assigned to a small transcription unit within a 15 kb genomic DNA fragment (Frei *et al.*, 1988). In order to identify molecular lesions associated with *sal* loss-of-function mutations, we analysed the various *sal* alleles in molecular detail. In all of the five *sal* loss-of-function alleles (Jürgens, 1988; this work, see Materials and methods), wild-type levels of transcripts in the correct spatial and temporal expression patterns were observed (data not shown). This suggested that the molecular lesions causing the loss-of-function mutations may reside within the coding sequences of the transcript. However, DNA sequence analysis revealed the wild-type coding sequence in all of the three *sal* alleles examined (data not shown). Thus, a different transcript from the previously identified one is likely to be essential for *sal* function. To search for additional transcribed DNA sequences we examined the region encompassing the 15 kb DNA fragment of the rescuing transgene by Northern blot analysis. No additional transcripts or different splicing forms of the previously identified transcript could be detected (data not shown). These results left severe doubts concerning the assignment of *sal* gene function to the previously identified transcription unit, consistent with the observation that the primary protein sequence encoded by this transcript is not conserved during insect evolution (Reuter *et al.*, 1989). For this reason, we repeated the P-element transformation experiments involving the 15 kb genomic DNA as reported previously (Frei *et al.*, 1988).

In the previous rescue experiments, the P-element construct containing 15 kb of DNA of the *sal* region was injected directly into embryos of the sal^{IIA55} *cn bw sp/CyO;* ry^{506}/ry^{506} genotype, and single eclosed males or females were mated with sal^{IIB57} *cn bw sp/CyO;*ry^{506}/ry^{506} flies. Two independent ry^{+} transgenic lines were analysed in more detail and they suggested a rescue of the embryonic *sal* phenotype due to the integrated DNA. To exclude an experimental artifact or error in this experimental design as a source of the 'rescuing activity', we altered the experimental design by injecting the P-element construct into ry^{506}/ry^{506} embryos. Two transgenic lines were established and the gene activity of the transgene was analysed in *sal* lack of function mutant background. No signs of rescuing activity coming from the transgene could be observed in these lines. These experiments strongly suggest that the previously identified *sal* transcript does not carry *sal* function as defined by the mutant phenotype. For reasons described below we refer to it as the *sal adjacent (sala)* transcript.

Identification of the sal transcription unit

In search of a transcription unit that encodes *sal* function we examined DNA fragments encompassing the 120 kb *sal* region by *in situ* hybridization to whole mount *Drosophila* embryos. Close to the chromosomal break point delimiting the *sal* region proximally, we found a transcript encoded by F4.5 DNA (Figure 1; see also Frei *et al.*, 1988) which is expressed in spatial and temporal patterns similar to those of *sala*. As shown in Figure 1Ba, F4.5 expression is found in three distinct regions of cellular blastoderm embryos. Transcripts are forming an anterior (60–70% of egg length) and a posterior (12–20% of egg length) stripe in positions corresponding to the precursors of the pattern elements which are affected in *sal* mutant embryos (Jürgens, 1988), and in a dorsally localized 'horse-shoe domain' in the presumptive pregnathal head region (80–86% of egg length).

A first hint that the F4.5 transcript may carry *sal* function is derived from examination of the *lacZ* enhancer-detection strain A405.1M2 (Bellen *et al.*, 1989). In this strain, the DNA of the enhancer trap construct resides within DNA sequences corresponding to clone F4.5 (Wagner-Bernholz *et al.*, 1991). Embryos containing this *lacZ* reporter gene show localized β-galactosidase expression in patterns that correspond to the patterns of F4.5 expression (Bellen *et al.*, 1989). Furthermore the A405.1M2 *lacZ* chromosome failed to complement *sal* lack of function mutations (Wagner-Bernholz *et al.*, 1991).

To test whether the *lacZ* insertion has caused the *sal* mutation we performed P-element 'jump-out experiments' (Cooley *et al.*, 1988). The removal of the P-element from its site of insertion resulted in a reversion of the *sal* allele to wild-type (data not shown), indicating that the *lacZ* insertion has caused the *sal* mutation which we refer to as sal^{A405}. In order to identify the F4.5 transcription unit as the one that carries *sal* function, we employed a P-element mediated germ line transformation and sequence analysis of DNA encoding the F4.5 transcript of *sal* mutant alleles. A 26 kb DNA fragment that contains the F4.5 transcription unit and 16 kb of non-transcribed sequences (Figure 1A) was used to generate a P-element construct, termed P[C20–*sal*26]. When inserted into the fly germ line, the P[C20–*sal*26] transgene rescues the allelic combination sal^{A405}/sal^{IIB57} that produces a weak *sal* phenotype in fertile flies. Furthermore, embryos homozygous for a *sal* loss-of-function mutation, such as sal^{IIB57} (Jürgens, 1988), develop a normal head region, and the tail phenotype is partially rescued in response to the P[C20–*sal*26] transgene (for details, see Figure 2). These results indicate that the P[C20–*sal*26] transgene contains *sal* function.

We also analysed F4.5 expression and the sequence of the F4.5 coding region of three different *sal* loss-of-function alleles of known genetic origin (see Materials and methods). In embryos homozygous for the alleles sal^{445} (Jürgens, 1988), sal^{16} and sal^{65} (this work, see Materials and methods), the F4.5 transcript is expressed in patterns and at levels indistinguishable from wild-type (data not shown). However, such embryos lack the expression of the corresponding protein as revealed by specific antibody stainings (see below). These observations suggest that each of the three independent *sal* mutations resides within the

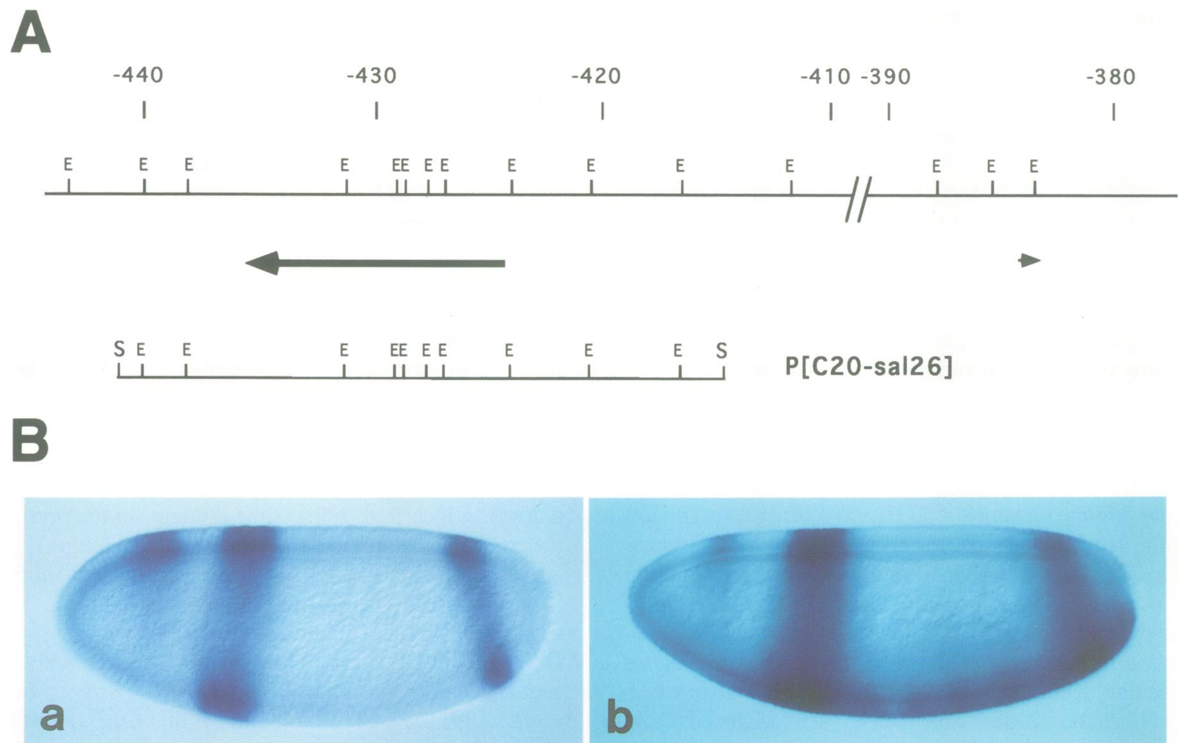


Fig. 1. Localization and expression of transcripts comprising the *sal* locus and genomic DNA used for P-element mediated transformation. (**A**) The scale showing DNA distance in kb (F4.5 walk clone encompasses DNA from −440 to −426) and the *Eco*RI. (E) restriction map of the chromosomal walk at 32F/33A described in Frei *et al.* (1985). The two arrows indicate the direction and the size of the F4.5 (large arrow) and the *sala* (small arrow) transcription unit. The lower part shows the 26 kb *Sal*I−*Sal*I genomic DNA fragment used for cloning into P-element vector Carnegie 20 (P[C20−*sal*26]). (**B**) Spatial expression of the F4.5 and the *sala* transcription unit in cellular blastoderm *Drosophila* embryos. Whole mount *in situ* hybridizations were performed with a digoxigenin-labelled probe of (**a**) c10 (F4.5 specific; see Figure 4a), and (**b**) cF9.1 (specific for *sala*; Frei *et al.*, 1988). Embryos are oriented with their anterior pole to the left and dorsal to the top.

coding region of the F4.5 transcript, thereby leading to protein products that failed to be recognized by the antibodies. We sequenced the DNA of F4.5 coding regions of the three *sal* mutants and the corresponding chromosomes in which the mutations were generated. Mutational changes in all three *sal* alleles were detected. The alleles sal^{16} and sal^{445} contain stop codons giving rise to truncated proteins, and the sal^{65} allele carries an 11 bp deletion which results in a frame shift (summarized in Figure 2e). These findings and the rescue of *sal* mutant alleles by the P[C20−*sal*26] transgene identify the F4.5 transcription unit as the one that carries *sal* gene function.

Structure and sequence of the sal gene

The structure of the *sal* transcription unit was determined by sequence analysis of five overlapping cDNAs and the corresponding 12.1 kb genomic DNA (Figures 3 and 4a). The *sal* transcription unit contains three introns and it codes for a single 6.1 kb long transcript matching the size of the embryonic poly(A)$^+$ *sal* RNA detected by Northern blot analysis (data not shown). The *sal* transcript contains a single open reading frame of 4065 bp (Figure 4a) which codes for a putative *sal* protein (SAL) of 1355 amino acids. SAL is characterized by seven C_2H_2 zinc finger motifs (Miller *et al.*, 1985), and by several regions rich in glutamines, alanines, prolines and serines, respectively (for details see Figure 4b).

The most prominent feature of SAL is the three widely separated, sequence related sets of two adjacent zinc finger motifs which we termed 'double zinc finger' (Figures 3 and 4b). The three double zinc fingers are connected by the evolutionarily conserved 'H/C-link' motif (Schuh *et al.*, 1986), and they show internal sequence similarity, i.e. 26 amino acids out of 49 are identical. In addition, a seventh sequence unrelated zinc finger motif is associated with the central double zinc finger (see Figures 3 and 4b). Each C-terminal zinc finger of the double zinc fingers contains a stretch of eight conserved amino acids of the sequence FTTKGNLK ('SAL-box'; Figure 4c) which is similar to the zinc finger sequences of the human transcription factor PRDII-BF1 (Fan and Maniatis, 1990) which contains two double zinc fingers similar to SAL (Figure 4d).

SAL expression in wild-type embryos

We examined the expression of SAL using antibodies generated against bacterially expressed *sal* protein (see Materials and methods). As shown in Figure 5, nuclear SAL antibody staining is detected after blastoderm formation throughout embryogenesis. SAL is first detected at the end of the syncytial blastoderm stage (late stage 4; stages according to Campos-Ortega and Hartenstein, 1985), forming a circumferential ring around the embryo which covers ~60−70% of egg length (EL; 'anterior SAL domain'; Figure 5a). During cellular blastoderm (stage 5) two additional expression domains can be observed; a circumferential ring in the posterior (12−20% EL; 'posterior SAL domain') and the 'horse shoe-shaped domain' in the anterior region (80−86% EL; Figure 5b) of the embryo. When the germ band is fully extended (stage 11), SAL accumulates in the neuroectoderm giving rise to a repetitive

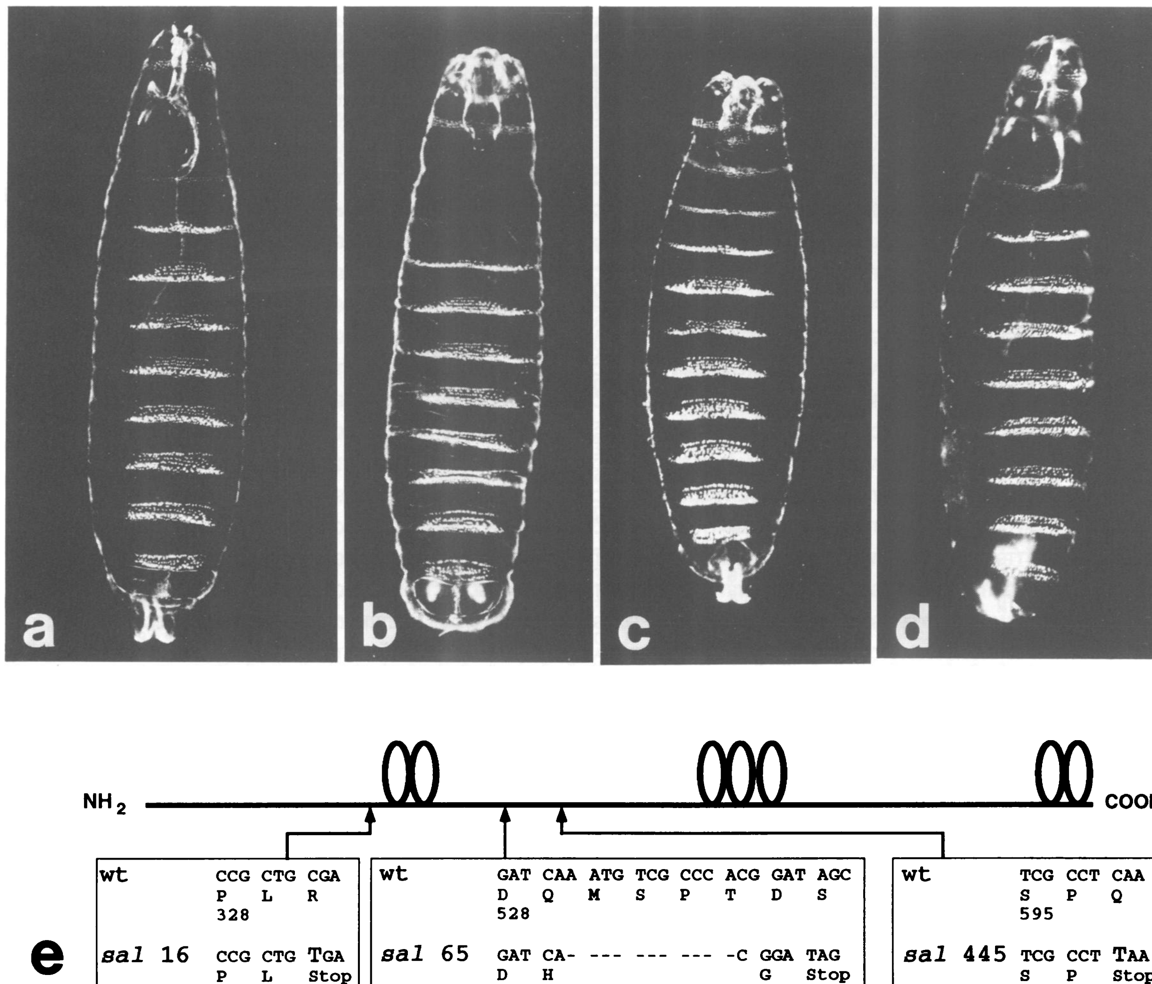


Fig. 2. Rescue of *sal* mutant embryos by germ line transformation and molecular lesions in the DNA of amorphic *sal* mutant alleles. Comparison of dark-field cuticle preparations of wild-type (**a**), amorphic sal^{IIB57}/sal^{IIB57} (**b**), hypomorphic sal^{IIB57}/sal^{A405} (**c**) and P-element transformed sal^{IIB57}/sal^{IIB57};P[C20−*sal*26] (**d**) larvae. (c) shows a hypomorphic sal^{IIB57}/sal^{A405} larva with an extreme *sal* mutant head similar to (b) and a wild-type tail similar to (a). In contrast to the amorphic phenotype (b) the head of the transformed larvae (d) is almost normal, whereas the tail of the transformed larvae (d) shows a somewhat shortened and reduced pair of Filzkörper compared with wild-type (a). Note that the P-element transformed sal^{IIB57}/sal^{A405};P[C20−*sal*26] embryos develop into viable and fertile flies, indicating a complete rescue up to adulthood of the otherwise lethal sal^{IIB57}/sal^{A405} (c) transheterozygotes by the P-element P[C20−*sal*26]. (**e**) The nucleotide sequences and the deduced amino acid sequences of wild-type and mutant *sal* alleles are compared. The numbers refer to wild-type SAL amino acid sequence (see Figure 3). Nucleotide changes in the mutant DNA are indicated in bold. Arrowheads facing the scheme of the *sal* protein show the positions of the stop codons in the allelic DNA. Ovals symbolize the location of the seven zinc fingers within the *sal* protein.

pattern in the central nervous system (Figure 5f and g). During stages 15−17 of embryogenesis, SAL is predominantly expressed in both the central nervous system and in the tracheal system (Figure 5g and h).

In order to localize the early SAL domains with respect to segment primordia, we used antibodies directed against the protein encoded by the segment polarity gene *engrailed* (*en*) to mark the anterior margins of each of the parasegments (Martinez-Arias and Lawrence, 1985). As defined by *en* expression, SAL expression in the anterior SAL domain spreads over parasegments 1−3 and fades to barely detectable levels in parasegment 4 (Figure 6). Thus, it covers the anlagen of the maxillary and the labial segments, as well as the posterior part of the mandibular segment and the anterior part of the first thoracic segment. The posterior SAL domain spans parasegments 14 and 15 as well as the primordium of the hindgut up to the Malpighian tubule anlagen. The posterior borders of the SAL domains are fuzzy while the anterior borders coincide cell-by-cell with parasegmental boundaries (Figure 6).

sal is conserved in higher Diptera

In order to see whether *sal* function is conserved during insect evolution at the molecular level, we analysed the SAL expression pattern in *Drosophila virilis* (*D.virilis*) and *Drosophila pseudoobscura* (*D.pseudoobscura*) embryos, and in embryos of the more distantly related dipteran species *Musca domestica* (*M.domestica*). We used the anti-SAL antibodies to examine whether SAL homologous protein is expressed in those embryos. As shown in Figure 7, the SAL antibody staining pattern in *D.virilis* and *D.pseudoobscura* corresponds both spatially and temporally to the expression pattern observed in *D.melanogaster* embryos. In *M.domestica,* however, SAL antibody staining corresponding to the *D.melanogaster* pattern is first detectable at the germ band extension stage (Figure 7e; see also Discus-

Fig. 3. Nucleotide and deduced amino acid sequences of the *sal* gene. The sequence of 12 164 bp of the genomic *sal* region (available under accession No. X75541), excluding DNA sequences of the first intron, is presented and numbered on the left side. The DNA sequence of the coding strand of a composite *sal* cDNA is indicated by upper case type. Introns and genomic sequences not represented in the cDNA are in lower case type. The predicted amino acid sequence is shown below the nucleotide sequence and numbered on the right side. The protein sequence shown is a conceptual translation of the longest open reading frame within the cDNA sequences. It begins at the fifth ATG of the cDNA and ends at a TGA triplet indicated by asterisks. Two putative polyadenylation signals at the 3′end of the transcription unit are double underlined and the poly(A) tail of a *sal* cDNA clone is indicated by an arrowhead. The P-insertion site of the *sal* mutant sal^{A405} is at nucleotide position 480.

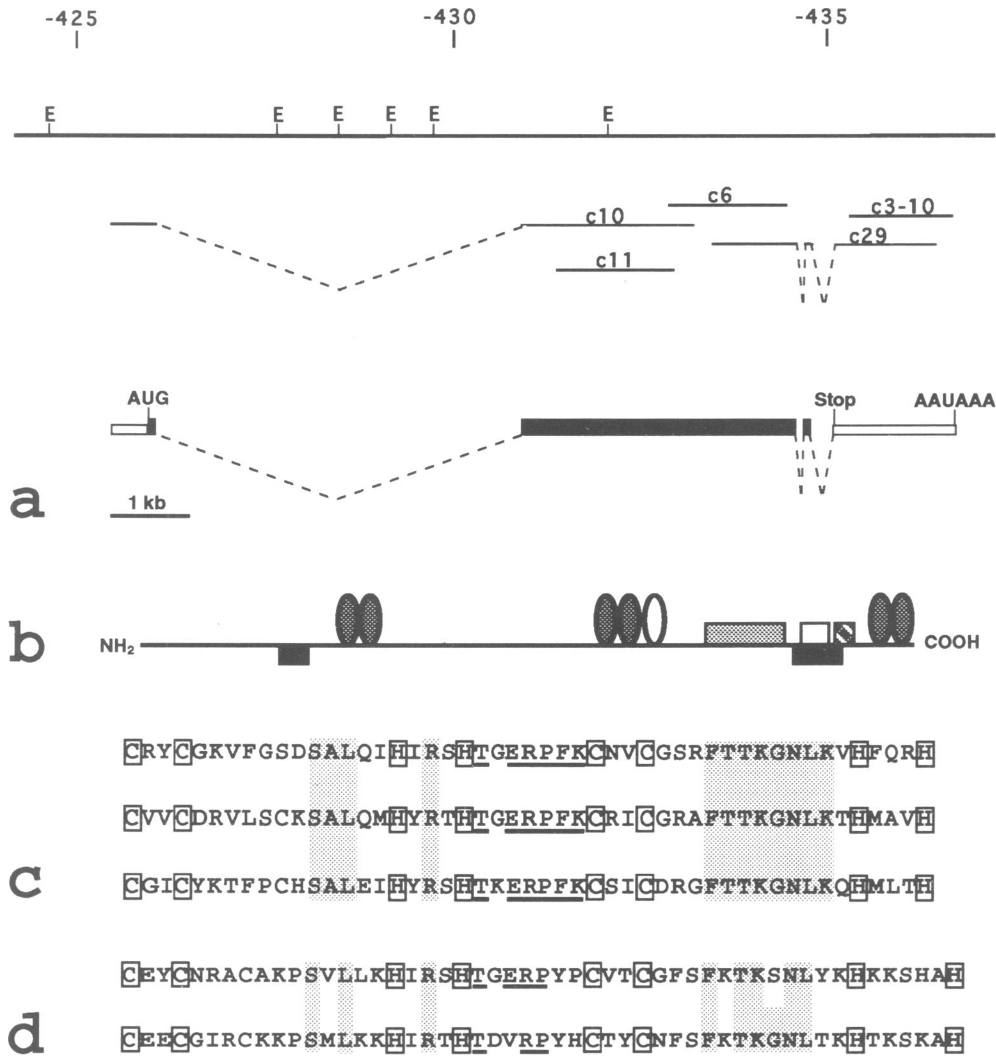


Fig. 4. Structural organization of the *sal* gene, the putative *sal* protein and its similarity with PRDII-BF1. (**a**) Molecular *Eco*RI (E) restriction map of the genomic region containing the *sal* gene (upper part) and the location of five cDNA clones. The scale refers to DNA distance (in kb) as described in Frei *et al.* (1985) (see also Figure 1). The composite molecular structure of the *sal* transcription unit is presented below. The translational start (AUG), the end of the open reading frame (Stop) and the poly(A) signal (AAUAAA) are indicated. Dotted lines: intronic sequences not present in the cDNA clones. Black bar: longest open reading frame (4065 bp) of the *sal* transcript. Open bar: untranslated region of the *sal* transcript. (**b**) Diagram showing structural features of the predicted *sal* protein. The seven ovals indicate the localization of the seven zinc finger motifs within the protein (filled ovals symbolize the double zinc fingers). Regions enriched for certain amino acids are shown as boxes with different shadings. Black boxes: regions with 38% (N-terminal) and 27% (C-terminal) glutamine residues, respectively. Open box: region with 53% alanine residues. Hatched box: region with 33% proline residues. Stippled box: region with 31% serine residues. (**c** and **d**) The invariant positions of the cysteines and histidines of SAL (c) and PRDII-BF1 (d) double zinc fingers are boxed. Identical H/C-link amino acids in the SAL double zinc fingers (c) are underlined whilst other identical amino acids are shown with dark background. Note the seven identical amino acids in the C-terminal zinc finger of the SAL double finger structures referred to as the 'SAL-box'. Amino acid positions of PRDII-BF1 double zinc fingers (d) shared by all SAL double zinc fingers are shown with dark background or, in the case of the H/C-link are underlined.

sion). These findings suggest that SAL corresponding protein is functionally conserved and required in the same anlagen as SAL in *D.melanogaster*.

In order to show the degree of molecular identity between SAL and the proposed SAL encoding gene of another *Drosophila* species, we cloned and sequenced the DNA of the *sal* homologue of *D.virilis*, a species that is sufficiently diverged from *D.melanogaster* to allow only functionally meaningful protein regions to be conserved (Kassis *et al.*, 1986; Treier *et al.*, 1989). In both *D.melanogaster* and *D.virilis*, the positions of the exon–intron boundaries of the *sal* transcription unit are conserved (data not shown). The putative protein sequences shown in Figure 7f indicate that the two proteins contain three zinc finger groups of almost complete sequence identity in the same relative positions. Sequences at each side of the three zinc finger groups show a higher degree of sequence similarity than the in-between regions. Within those, islands of 10–30 conserved amino acids are found. The longest detectable open reading frame of both genes has a common conserved initiation codon at the N-terminus, although this initiation codon is preceded by another in-frame initiation codon which adds 11 amino acids to the N-terminus of SAL from *D.melanogaster*. These results suggest that *sal* function is conserved both functionally and molecularly in *Drosophila* and probably also in other diptera.

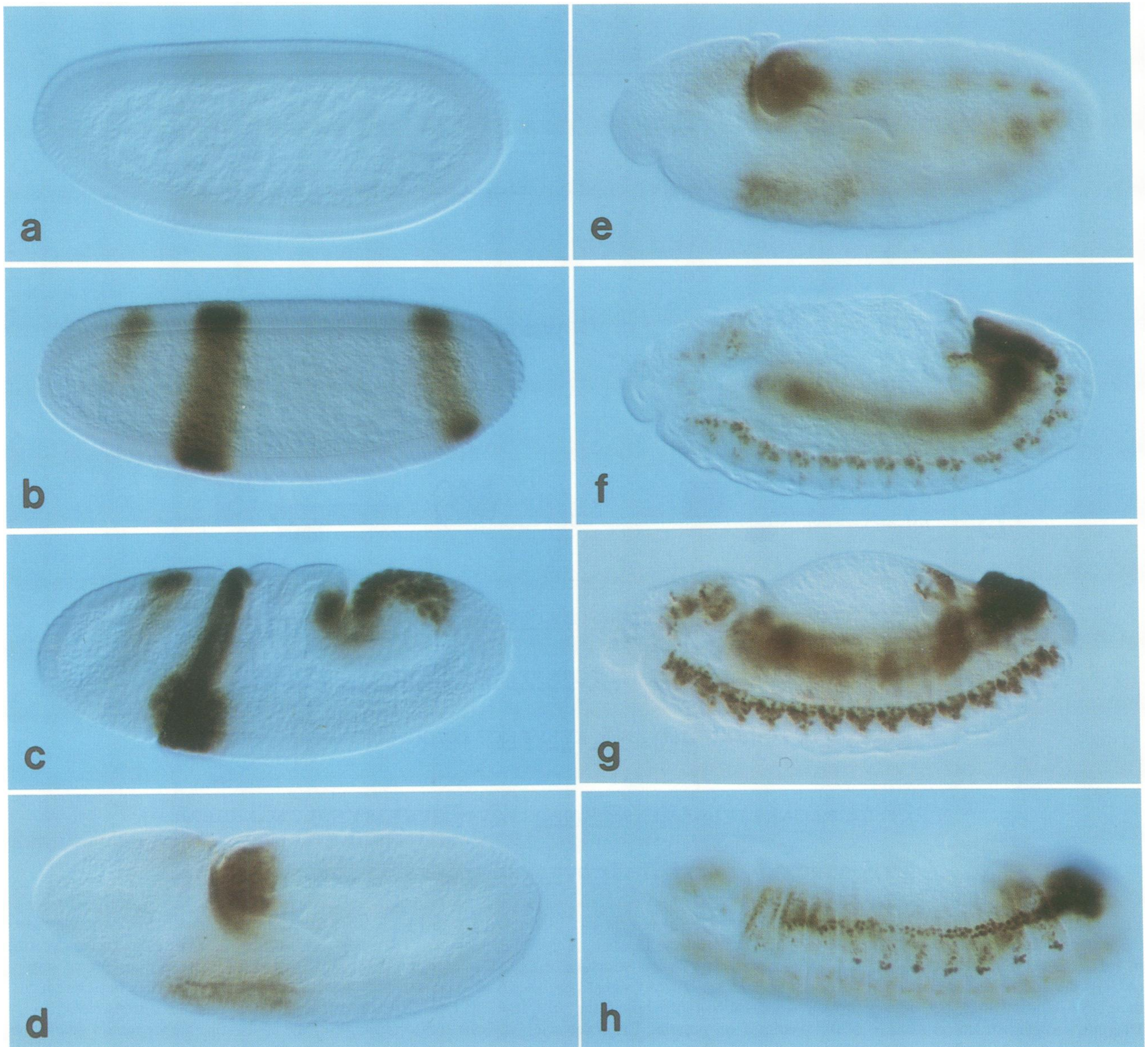


Fig. 5. *sal* protein expression during *Drosophila* embryonic development. Whole mount preparations of wild-type embryos were stained with anti-SAL antibodies. Stages are described according to Campos-Ortega and Hartenstein (1985). (**a**) Early stage 5, expression of the anterior SAL domain. (**b**) Late stage 5, cellular blastoderm. The posterior SAL domain and 'horse-shoe domain' become visible (see text). (**c**) Stage 8, germ band extension. The posterior domain moves cephalad during the phase of germ band elongation. (**d**) Stage 9, stomodeal plate formation. The anterior SAL domain starts to fade out, whilst the posterior domain persists. (**e**) Stage 10, fully extended germ band. (**f**) Stage 12, germ band retraction. Strong expression within the region of the developing posterior spiracles. Segmentally repeated SAL expression in restricted parts of the ventral cord and the procephalic neurogenic region. (**g**) Stage 14, beginning of head involution. SAL expression in the tracheal system (**h**) Stage 14. Focus on lateral epidermis. Staining in the oenocytes, bilateral groups of cells in abdominal segments 1–7 (Hartenstein and Jan, 1992) and parts of the tracheal system becomes visible.

Discussion

Our results show that the region specific homeotic gene *sal* encodes an evolutionarily conserved zinc finger protein. The identification of the *sal* gene is based on two independent lines of evidence. A transgene that contains a single transcription unit rescues *sal* mutant embryos, and molecular lesions were found in the sequence of all *sal* alleles analysed. These findings are in contrast to the previous assignment of the *sal* gene which, as is shown here, is based on an experimental artifact or an experimental error. With respect to its chromosomal location next to *sal*, we rename this gene as *sal adjacent* (*sala*).

SAL is expressed in the segment anlagen affected by sal mutant embryos

sal mutations affect posterior head and anterior tail segments. In the head region *sal* mutants cause partial transformation of maxillary and labial segments to develop prothoracic structures (Jürgens, 1988). In accordance with this mutant phenotype, SAL is expressed in parasegments 1–3 which include the primordia of both the maxillary and labial

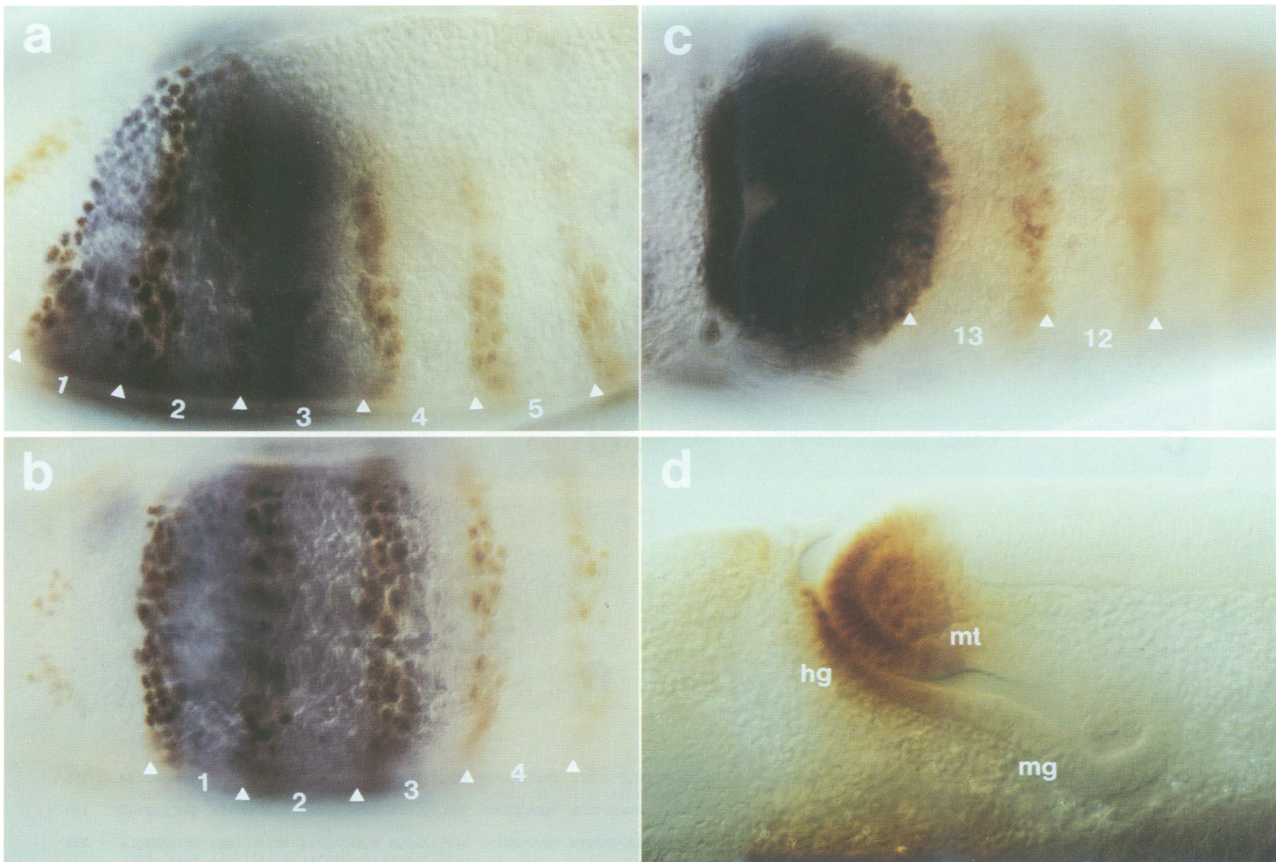


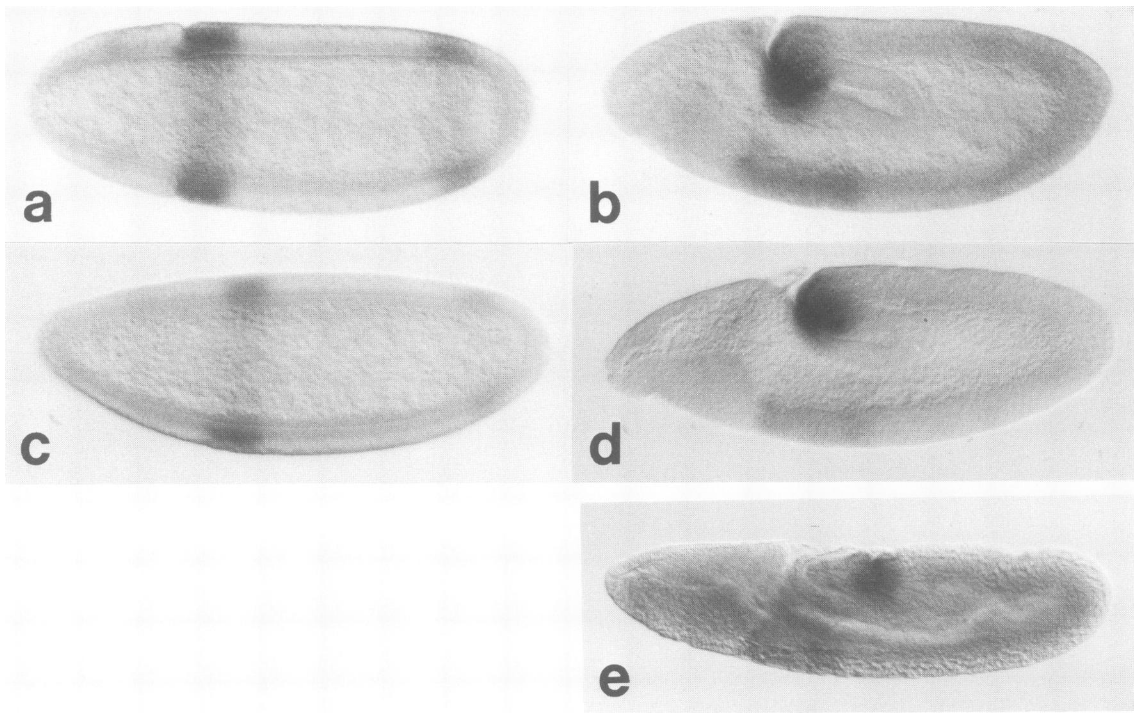
Fig. 6. *sal* protein expression with respect to parasegmental boundaries. Whole mount preparations of wild-type fully extended germ band embryos double stained with antibodies against *sal* (blue) and *engrailed* proteins (brown) (**a**–**c**); single staining against *sal* protein (brown) (**d**). (a and b) show the anterior SAL expression domain (a, lateral view: dorsal up, anterior left, b, ventral view). The mandibular *engrailed* stripe (in PS 1) marks the anterior boundary of the SAL expression; both limits coincide cell-by-cell. The SAL expression in the anterior prothoracic compartment (posterior PS 3) is weak and very weak expression is also detectable in PS 4. (c and d) show the posterior SAL expression domain (c: ventral view. d: lateral view; dorsal up, anterior left). The anterior limit of the posterior SAL expression domain coincides with *engrailed* expression in abdominal segment 8 (in PS 14). The posterior boundary coincides with the posterior tip of the hindgut, demarcated by the Malpighian tubules primordium that separates ectodermal hindgut from endodermal midgut (d). Arrowheads indicate parasegmental boundaries and numbers refer to parasegments. Abbreviations: PS, parasegment; hg, hindgut; mt, Malpighian tubules primordium; mg, midgut.

segments. However, very weak SAL expression is also detectable posterior to parasegment 3. Therefore, the parasegmental nature of the posterior boundary of the anterior SAL domain remains an open question. In the posterior domain, SAL expression is found in parasegments 14 and 15, which are homeotically transformed in the *sal* mutants (Jürgens, 1988), but it also fades into the primordium of the hindgut whose metameric nature is unclear (Jürgens and Weigel, 1988). Therefore early SAL expression is similar to the restricted parasegmental expression that has been observed with the region specific homeotic gene *fkh* and the homeotic selector genes (Akam, 1987; Ingham, 1988; Weigel *et al.*, 1989). However, the barely detectable SAL expression in parasegment 4 may also be of functional importance, since *Ubx* gene product (which is restricted to parasegments 5–13 in wild-type embryos) expands into parasegments 3 and 4 of *sal* mutants (Casanova, 1989). This suggests that low levels of SAL are able to repress *Ubx* in the wild-type embryo either by its direct interaction with *Ubx cis*-regulatory elements (see below) or indirectly through as yet unknown factors.

Widely spaced double zinc finger motifs in the sal protein

The *sal* protein (SAL) contains seven zinc finger motifs of the C_2H_2-type first identified in the transcription factor TFIIIA (Miller *et al.*, 1985). Their arrangement in three sets of highly conserved and widely spaced double zinc fingers suggests that they may have derived from an ancestral gene encoding a single double zinc finger motif, through sequence duplication or intragenic conversion. Among the double zinc fingers the most striking amino acid homology is found in two boxes, the H/C-link (Schuh *et al.*, 1986) and an array of eight identical amino acids, termed 'SAL-box', which are found in the C-terminal finger motif of the double zinc fingers. Interestingly, two pairs of widely spaced and conserved finger pairs have also been observed in a human transcription factor, PRDII-BF1 (Fan and Maniatis, 1990). In addition to the structural conservation, PRDII-BF1 and SAL show a significant degree of sequence similarity within the pairs of zinc fingers suggesting that widely spaced and sequence related 'double zinc fingers' may define a conserved subfamily of zinc finger proteins.

Zinc finger motifs are characteristic of a distinct class of nucleic acid binding proteins. Molecular modelling (Berg, 1988; Gibson *et al.*, 1988) and two-dimensional NMR (Lee *et al.*, 1989) have led to a proposal for the three-dimensional structure of C_2H_2-type zinc fingers: the Cys–Cys loop forms an antiparallel β-sheet followed by an α-helical region through the His–His loop which contacts in the major groove of DNA (Pavletich and Pabo, 1991). The 'SAL-box' extends into the proposed helical region. Thus, the three sets




```
MKNHLSNVLC AMRSDFKDNH QETINKMIQF GTVKYGIVKQ LKDRARSADK DIGSDQEENG GCSPLTTATT TASPSRSPEP EEEQPEEQST SEQSIPEQ_S TPDHQLENDI KSEAKSEIEP 119
           MRSDFKDNH QETINKMIQF GTVKYGIVKQ LKDPARSAEK DIASDQEDNG ACSPLTTANA SASAGNSPCP SRS_PQQHSE DERE_PEQVS EQELVPEVSA QSE__SEIGE 105

VEDNNNRVAM TKPSSEEREP NASGSMPSSP VAEASAEEAA TERTPEKEKE KDVEVDVEMP DEAPSSAVPS TEVTLPGGAG APVTLEAIQN MQMAIAQFAA KTIANGSNGA DNEAAMKQLA 239
_EIENNADET NADHNHNNNN NNKLVMTKPP V_EHEVEQNA NLNASMPNST TP_______P __ATN_AVIA GARAQQFGAT _PVTLEAIQN MQMAIAQFAA KTIANGASGT DNEAAMKQLA 212

FLQQTLFNLQ QQQLFQIQLI QQLQSQLALN QAKQEEDTEE DADQEQDQ_E Q_ETDTYEEE ERIADMELRQ KAEARMAEAK ARQHLINAGV PLRESSG_S_ _PAES__LKR __R_REHDHE 349
FLQQTLFNLQ QQQLFQLQLI QQLQSQLALN QVKQNDD_EA DEELEPEERE DGETDTYEEE ERIADMELRQ KAEARMAESK ARQHLINAGV PYANAPDPSH QPPHRCRLRR LKRKREEDA_ 330

SQPNRRTSLD NTHKADTAQD ALAKLKEMEN TPLPFGSDLA SSIITNHDDL PEPNSLDLLQ KRAQEVLDSA SQGILANSMA DDFAFGEKSG EGKGRNEPFF KHRCRYCGKV FGSDSALQIH 469
SAKSSGASAK IFGEQESSQD ALNKLKEMEN MPLPFGADLS SSIITNHDDL PEPNSLDLLQ KRTQEVLDSA SQGILANNMA DDFAFGDKSS DGKGRNEPFF KHRCRYCGKV FGSDSALQIH 450

IRSHTGERPF KCNVCGSRFT TKGNLKVHFQ RHAQKFPHVP MNATPIPEHM DKFHPPLLDQ MSPTDSSPNH SPAP____PP LG_SA_____ PA_SF__PPA FPGLQNLYRP PMEILKSLGA 576
IRSHTGERPF KCNVCGSRFT TKGNLKVHFQ RHAQKFPHVP MNATPIPEHM DKFHPPLLDQ MSP_DSSPTQ SPAPATGLPP PSTSTLTQMQ PSMSFASSPA FPGLPGIYRP PMELLKSLGA 569

_A___A__PH QYFPQE__L_ __________ ___PTDLRKP S_PQLD_EDE PQVKNE_PVE EKDQREEH_E QEM_AECSE_ PEPEPLPLEV RIKEERVEEQ EQVKQEDHRI _EPRR___TP 664
TAGSTAGLPH PFFPQMPGLG AALKHTHDQS QDMPTDLRKS SGPSSPHEEE DNIAARLPVK SELMEEEKTE HTMEAATRES AEMEPLPLEV RIKEERIDED QMHLQEGMQK PEPLTAYATP 689

SPSSEHRSPH HHRH_SHMGY PPVVQPIQPA ALMHPQSSP_ _______GSQS HLDHLPTPGQ LPP____RED FFAERFPLNF TTAKMLSPEH HSPVRSPAGG ALPPGVPPPP HHHPHHMARS 772
HPQQCLIPTT HAAAKSPRSL PLQCHARLSL WCSHPTTSNH ACAVLTGSQT HLDQLPTPDN VPPTMPQRED FFAERFPLNF TS_KTDD__H _SPIRSPAG_ __________ _HAHAHIPRS 793

PFFNPIKHEM AALLPRPHSN DNSWENFIEV SNTCETMKLK ELMKNKKISD PNQCVVCDRV LSCKSALQMH YRTHTGERPF KCRICGRAFT TKGNLKTHMA VHKIRPPMRN FHQCPVCHKK 892
PFFNPIKHEM AAFVPRPHSN DNSWENFIEV SNTSETMKLK ELMKNKKISD PNQCVVCDRV LSCKSALQMH YRTHTGERPF KCRICGRAFT TKGNLKTHMA VHKIRPPMRN FHQCPVCHKK 913

YSNALVLQQH IRLHTGEPTD LTPEQIQAAE IRDPPPSMMP GHFMNPFAAA AFHFGALPGG P_GGPPGPN_ __HGAHNGAL GSESSQGDMD DNMDCGEDYD _DDVSSEHLS NSNLEQEG_D 1006
YSNALVLQQH IRLHTGEPTD LTPEQIQAAE IRDPPPSMMP GHFMNPFAAA AFHFGAMPGG GAGGPPGATG MPGGPHNGTL GSESSQGDLD DNMDCGDGDD FDDISSEHLS NSNDPAATSD 1033

RSRSGDDFKS LLFEQKLRID ATGVVNTNPV _RPRSSAS__ _SHGHSVGST SAPTSPSVHA S_SQVIKRSS _SPARS__EA SQGALDLTPR AAPTS__SSS SRSPLPKEKP VSPPSLPRSP 1116
R_RSSDDFKS LLFEQKLRID PTGVVNINSH QRPHSAASNP NSIGSASASP SAPTSPSSQP KPSCSPVRSS CSPVRSVSET SQGALDLTPR ALPPPLASSS SRSPYRQLLS VRRRPLARSV 1152

SGSSHASAN_ __ILTSPLPP TVGIDCLPKG LQHHLQQQHQ HLMQQQAAVA AAAAAQHHHH Q_Q__MAALD QHQEQLRREA AEAQQKAA__ ___AAAAAAA AAAAAQRQTP PQARDQRQEG 1225
SSHRCVVPMV RALLSSQLPP SVGIDCLPPG LQHHLQQQHQ HLMQQQAAVA AAAAAQHHHH QMQQHAAALH QHQEHLRREA QEVQQKAAQE VQQKAAAAAA AAAAAQRQES PQPPPRSGES 1272

GPGAGPPPNP LMGARPPFGM FPNLPLFPPA TTQNMCNAMN QIAQSVMPAA PFNPLALSGV RGSTTCGICY KTFPCHSALE IHYRSHTKER PFKCSICDRG FTTKGNLKQH MLTHKIRDME 1345
SVGPPAQPNP LISARPPFGM FPNLPIFPPA TTQNMCNAMN QIAQSVMPAA PFNPLALSGV RGSTTCGICY KTFPCHSALE IHYRSHTKER PFKCNICDRG FTTKGNLKQH MLTHKIRDME 1392

QETFRNRAVK 1355
QETFRNRAVK 1402
```

f

Fig. 7. Early *sal* protein expression in different dipteran embryos and comparison of the putative *sal* proteins of *D.melanogaster* and *D.virilis*. Embryos of *D.pseudoobscura* (**a** and **b**), *D.virilis* (**c** and **d**) and *M.domestica* (**e**) are stained with anti-SAL antibodies. (a and c) cellular blastoderm stage; (b,d and e) germ band extension stage. (**f**) Sequence comparison of *sal* protein of *D.melanogaster* (upper sequence; numbered in bold type) and *D.virilis* (lower sequence; numbered in plain type). Vertical dashes represent identities; horizontal dashes represent gaps in the sequence. The zinc finger regions are shown with dark background. Alignment conditions using Mac Molly Tetra, Version 1.2 from Soft Gene GmbH: minimal window size 20; number of mismatches 5; gap penalty 5; mismatch penalty 4. DNA sequence of the *D.virilis sal* is available under accession No. Z27444).

of double zinc fingers in SAL may each recognize the same DNA target sites as shown for each of two double zinc fingers of PRDII-BF1 (Fan and Maniatis, 1990).

The role of sal

The C_2H_2-type of zinc fingers can be grouped in different functional classes. They might act as DNA binding transcriptional regulators and/or bind to RNA. Other zinc finger proteins are integral components of chromatin, or provide the nuclear transport of cytoplasmic components (reviewed in El-Baradi and Pieler, 1991). Zinc finger containing transcription factors have been shown to contain diagnostic domains such as proline- and glutamine-rich regions which are thought to function as activator sequences (Courey and Tjian, 1988; Mermod *et al.*, 1989), or alanine- and proline-rich regions required for repressor function (Licht *et al.*, 1990; Han and Manley, 1993). SAL, which accumulates in the nuclei, contains both types of sequence. Thus, it may function as a transcriptional activator or repressor of target gene expression.

A possible target gene of SAL is *tsh*, a unique homeotic gene which defines the ground state of the trunk (Röder *et al.*, 1992). In addition, *tsh* is essential for specifying the identity of the anterior prothorax by acting in concert with *Sex combs reduced* (*Scr*). In the absence of *sal*, the labial segment is partially transformed to anterior prothorax, although *Scr* expression in this segment is not altered in *sal* mutants (Casanova, 1989). However, *tsh* expression expands towards the anterior in such a way that *Scr* and *tsh* expression coexist in the labial anlagen which then gains anterior prothoracic identity (Röder *et al.*, 1992). Thus, *sal* activity inhibits *tsh* expression in the wild-type embryo, and thereby prevents trunk development. This observation is consistent with the hypothesis that SAL is a transcriptional repressor of the *tsh* gene and that its interaction with *tsh* prevents trunk development within the head and the tail regions.

Antp and *tsh* activities combine for mesothorax, the segment in which mesothoracic leg and wing imaginal discs normally form. Struhl (1981) observed that some cells in the mesothorax leg disc which lack the expression of *Antp* form patches of antenna-like cells in the adult and based on this finding he suggested that *Antp* normally represses the activity of 'head forming genes' in the leg disc. It has recently been shown that *sal* expression occurs in antennal but not in leg discs, and that *Antp* activity serves as a strong repressor of *sal* (Wagner-Bernholz *et al.*, 1991). Although we do not yet know whether *sal* plays a decisive role in antennal development, it will be interesting to determine whether *sal* may carry the function proposed by Struhl (1981). The early period of antennal disc expression of *sal* is at the time in development when ectopic expression of *Antp* is required to induce antenna-to-leg transformations. This SAL expression is entirely repressed by *Antp* activity and leg instead of antennal structures are formed (Wagner-Bernholz *et al.*, 1991). Since *sal* activity represses *tsh* expression, and *Antp* enhances *tsh* activity in the blastoderm (Röder *et al.*, 1992), it might be that *sal* activity represses *tsh* expression in the wild-type antennal discs, and ectopic expression of *Antp* might then act as an antagonist of *sal* i.e. repressing the *tsh* repressor and enhancing *tsh* activity at the same time. Obviously, these proposed interactions need to be elucidated by molecular means in order to establish the regulatory circuitry that leads to the homeotic antenna-to-leg transformation in the fly.

Conservation of SAL

The overall sequence conservation between SAL of *D.melanogaster* and its homologue in *D.virilis* is close to 70% amino acid identity, although the two species are ~60 million years of evolution apart from each other. This degree of overall sequence conservation is within the range described for several other developmentally important genes (Kassis *et al.*, 1986; Treier *et al.*, 1989; Michael *et al.*, 1990). The double zinc fingers are almost completely conserved in sequence, and a high degree of conservation is found within the glutamine-rich regions in front of the first and third double zinc finger. This implies that those regions, which might be required for DNA binding and transcriptional activation (Courey and Tjian, 1988), respectively, are essential for SAL function. In contrast to these conserved and diagnostic protein motifs, no sequence similarity with known protein modules has been detected for the conserved N-terminal region of SAL, and thus the significance of the conservation remains unclear. Contrary to the frequent occurrence of cryptic simplicity in other regulatory genes (Colot *et al.*, 1988; Treier *et al.*, 1989) the *sal* proteins lack such sequences. Thus, slippage-like processes may not be involved in the evolution of the *sal* proteins. Instead, the highly diverged sequences could be explained by insertions and deletions of blocks of 30–60 bp fragments within the *sal* transcription unit. However, those alterations do not affect the relative distance between the three double zinc finger motifs, suggesting evolutionary constraints concerning the spacing of these protein domains.

The argument that the two proteins carry corresponding biological functions during the development of the two *Drosophila* species is consistent with the finding that the SAL antigene is found in corresponding patterns in the blastoderm of both *D.melanogaster* and *D.virilis*, and during gastrulation of *M.domestica*, a dipteran species which is ~100 million years separated from *D.melanogaster* (Hennig, 1981). The relatively late appearance of SAL antigene in *M.domestica* embryos may be explained by a weak cross-reactivity which requires high amounts of the homologous protein to be visualized by the SAL antibody. Alternatively, SAL homologous expression might be delayed in *M.domestica* as has been observed with various segmentation genes that are expressed in the terminal regions of those embryos, which would imply that the mode of terminal development differs between *Musca* and *Drosophila* (Sommer and Tautz, 1991).

It had been proposed that *sal* activity plays a conserved role in head formation during arthropod development, an evolutionary event leading to the organizational level of myriapods through a process in which anterior trunk segments of annelid-like ancestors were integrated into a primitive head (Jürgens, 1988). As a first step towards a critical test of this hypothesis we have shown SAL conservation among higher dipteran species. Since SAL homologous sequences have already been identified in vertebrates such as *Xenopus laevis* (R.Stick, personal communication) and mouse (G.Schütz, personal communication), we expect SAL to be conserved throughout the animal kingdom in a manner preceded by the genes of the homeotic selector gene complexes.

Materials and methods

Isolation, sequencing and analysis of wild-type and mutant DNA

Isolation of *sal* cDNAs, preparation of DNA, Southern blot analysis and handling of DNA were done by standard methods (Sambrook *et al.*, 1989). DNA sequencing was performed using the USB Sequenase 2.0 Sequencing Kit based on the chain termination procedure (Sanger *et al.*, 1977). Single-stranded DNA templates were generated using M 13 vectors (Yanisch-Perron *et al.*, 1985). The sequences of the wild-type genomic DNA and cDNAs were determined on subcloned restriction fragments. Sequence analysis and comparison of the predicted *sal* protein were performed on the SwissProt database using the HUSAR program package (based on the GCG package) of the DKFZ, Heidelberg. To analyse the DNA from mutant *sal* alleles we amplified genomic *sal* DNA from single mutant embryos by PCR as described in Hülskamp (1991). An identified mutation was confirmed by DNA sequence analysis of at least one additional mutant embryo and the parental DNA for control.

Isolation of sal alleles

The two *sal* alleles, sal^{16} and sal^{65}, were induced by 30 mM EMS fed to *b pr cn* wx^{wxt} *bw* males. They were mated to *CyO* balancer females for 3 days. The F1 male progeny were individually crossed with sal^{IIB57} *cn bw sp/CyO* females. In 6800 lines two putative *sal* alleles were found by the lack of homozygous *cn, bw* (white eyed) F2 progeny. Both lines, sal^{16} and sal^{65}, were test-crossed with different *sal* alleles and checked for their embryonic phenotype and lethality. sal^{16} and sal^{65} do not complement sal^{IIB57}, sal^{IIA55}, sal^{445} (Jürgens 1988) and sal^{A405} (Bellen *et al.*, 1989; see also Results), and cuticle preparations of the various mutant combinations indicate that both sal^{16} and sal^{65} homozygous embryos show the *sal* lack of function phenotype described by Jürgens (1988).

Generation of antisera and antibody purification

The 10.1 cDNA coding for the amino acids 403–763 of the predicted *sal* protein was subcloned into pUR vector (Rüther and Müller-Hill, 1983) and pATH vector (Koerner *et al.*, 1991) to generate pUR–10.1 (lacZ–*sal* fusion protein) and pATH–10.1 (trpE–*sal* fusion protein). The purification of fusion proteins, generation of antisera in rabbits and the purification of 10.1 antibodies were performed according to Gaul *et al.* (1987) with minor modifications. In contrast to their procedure we used the trpE–*sal* fusion protein to generate antisera. For affinity purification of antibodies directed against the *sal* part of the trpE–*sal* fusion protein an Affigel 10/15 lacZ–*sal* fusion protein column was used. Purified antibodies were checked for activity against *sal* specific protein sequences by Western blot analysis and whole mount antibody staining of embryos.

DNA sequence analysis of strong *sal* mutants (sal^{445}, sal^{65} and sal^{16}) reveals that these embryos are only able to express truncated *sal* protein lacking major parts of the wild-type protein. Embryos derived from heterozygous parents of these *sal* mutants were stained with the affinity purified anti-*sal* antibodies and antibodies against *Kr* protein (Gaul *et al.*, 1987) as an internal control. One quarter of these embryos showed only *Kr* protein staining (data not shown), indicating that the anti-*sal* antibodies fail to recognize antigens in homozygous *sal* mutants. Therefore our purified antibodies detect the *sal* protein specifically.

Developmental expression analysis

Whole mount preparations of embryos were antibody stained using the VECTASTAIN Elite ABC-peroxidase system (Vector Laboratories) with DAB as substrate under the conditions described (Macdonald and Struhl, 1986). *In situ* hybridizations of whole mount embryos using digoxigenin-labelled DNA probes were done as published (Tautz and Pfeifle, 1989).

Germ-line transformation and genetic analysis

The 26 kb *SalI–SalI* fragment was subcloned from a genomic cosmid into *SalI*-digested Carnegie 20 vector (C20–*sal26*) (Rubin and Spradling, 1983). About 1000 embryos of the ry^{506}/ry^{506} stock were injected with the recombinant and the helper plasmid as published (Spradling, 1986). Two out of 400 GO flies produced ry^+ F1 flies indicating a transformation of two independent fly lines. The fly stocks P(C20–*sal26*)ry^+A (insertion mapped to the second chomosome) and P(C20–*sal26*)ry^+B (insertion mapped to the third chromosome) were established. Using P(C20–*sal26*)ry^+A and the jump-start technique (Cooley *et al.*, 1988; Robertson *et al.*, 1988) two additional fly strains could be obtained that had the P element inserted on the third chromosome. The following experiments were performed independently with the three P element lines inserted on the third chromosome.

To analyse the transformed chromosomes in a sal^- background we crossed sal^{IIB57}*cn bw sp/CyO*;ry^{506}/ry^{506} males with +/+;ry^{506} [P(C20–*sal26*)ry^+]/ry^{506} females. sal^{IIB57}*cn bw sp*/+ ry^{506}/ry^{506} [P(C20–*sal26*)ry^+] males were backcrossed with sal^{IIB57}*cn bw sp/CyO*;ry^{506}/ry^{506} females and from their progeny sal^{IIB57}*cn bw sp/CyO*;ry^{506}/ry^{506} [P(C20–*sal26*)ry^+] males and females were collected and crossed to establish fly lines. To analyse the *sal* mutant rescue capacity of P(C20–*sal26*) we crossed sal^{IIB57}*cn bw sp/CyO*;ry^{506}/ry^{506} [P(C20–*sal26*)ry^+] males with females carrying the hypomorphic *sal* allele *cn pr/CyO Pry*$^+$[sal^{A405}];ry^{506}/ry^{506}. This cross produced sal^{IIB57}*cn bw sp/CyO Pry*$^+$[sal^{A405}];ry^{506}/ry^{506} [P(C20–*sal26*)ry^+] adult flies showing a rescue to viability of the otherwise embryonic lethal sal^{A405}/sal^{IIB57} mutant combination. A stock of sal^{IIB57}*cn bw sp/CyO Pry*$^+$[sal^{A405}];ry^{506}/ry^{506} [P(C20–*sal26*)ry^+] was established and analysed in detail. The embryonic progeny of the stock were collected and embryos were stained for β-galactosidase activity (Bellen *et al.*, 1989) to distinguish between embryos carrying the *CyO Pry*$^+$[sal^{A405}] chromosome which expresses β-galactosidase from the sal^{IIB57}*cn bw sp* homozygous ones. Among the unstained embryos (homozygous sal^{IIB57} embryos) some show the amorphic *sal* phenotype while the majority develop a wild-type head but show a shortening of the normally stretched posterior spiracles. Those embryos die during first instar larval development. The same phenotype was observed among transheterozygous sal^{IIB57}/sal^{IIA55} (or sal^{445}, sal^{65} or sal^{16}) embryos. These rescue results were confirmed independently with the three P-element lines inserted on the third chromosome.

Screening and sequencing of the sal gene from D.virilis

About 80 000 plaques of a genomic *D.virilis* library were screened as described (Treier *et al.*, 1989), using the *sal* 10.1 cDNA. One positive clone was identified and shown to contain the *sal* gene of *D.virilis*. The gene was analysed by restriction analysis, hybridization and DNA sequencing. DNA sequence analysis was predominantly done of the coding sequences and the exon–intron boundaries.

Acknowledgements

We thank R.Kemler and D.Vestweber for helping with the antibody production and R.Sommer for providing the *M.domestica* embryos. We also thank D.Tautz for providing the *D.virilis* library and B.Hovemann for the cDNA library. Many thanks go to T.Berleth, S.Cohen, G.Jürgens and G.Schütz for critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (DFG grant Schu 683/1-2 and SFB 236).

References

Affolter,M., Schier,A. and Gehring,W.J. (1990) *Curr. Opin. Cell Biol.*, **2**, 485–495.
Akam,M. (1987) *Development*, **101**, 1–22.
Bellen,H.J., O'Kane,C., Wilson,C., Grossniklaus,U., Pearson,R.K. and Gehring,W.J. (1989) *Genes Dev.*, **3**, 1288–1300.
Berg,J.M. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 99–102.
Campos-Ortega,J. and Hartenstein,V. (1985) *The Embryonic Development of Drosophila melanogaster*. Springer-Verlag, Berlin.
Casanova,J. (1989) *Wilhelm Roux's Arch. Dev. Biol.*, **198**, 137–140.
Colot,H.V., Hall,J.C. and Rosbash,M. (1988) *EMBO J.*, **7**, 3929–3937.
Cooley,L., Kelley,R. and Spradling,A. (1988) *Science*, **239**, 104–210.
Courey,A.J. and Tjian,R. (1988) *Cell*, **55**, 887–898.
El-Baradi,T. and Pieler,T. (1991) *Mech. Dev.*, **35**, 155–169.
Fan,C.-M. and Maniatis,T. (1990) *Genes Dev.*, **4**, 29–42.
Frei,E., Baumgartner,S., Edström,J.-E. and Noll,M. (1985) *EMBO J.*, **4**, 979–987.
Frei,E., Schuh,R., Baumgartner,S., Burri,M., Noll,M., Jürgens,G., Seifert,E., Nauber,U. and Jäckle,H. (1988) *EMBO J.*, **7**, 197–204.
Gaul,U., Seifert,E., Schuh,R. and Jäckle,H. (1987) *Cell*, **50**, 639–647.
Gehring,W.J. and Hiromi,Y. (1986) *Annu. Rev. Genet.*, **20**, 147–173.
Gibson,T.J., Postma,J.P.M., Brown,R.S. and Argos,P. (1988) *Protein Engng*, **2**, 209–218.
Han,K. and Manley,J.L. (1993) *Genes Dev.*, **7**, 491–503.
Hartenstein,V. and Jan,Y.N. (1992) *Wilhelm Roux's Arch. Dev. Biol.*, **201**, 194–220.
Hennig,W. (1981) *Insect Phylogeny*. Wiley, Chichester, UK.
Hülskamp,M. (1991) PhD Thesis. University of Tübingen, Tübingen, Germany.

Ingham,P. (1988) *Nature*, **335**, 25–33.
Jürgens,G. (1985) *Nature*, **316**, 153–155.
Jürgens,G. (1988) *EMBO J.*, **7**, 189–196.
Jürgens,G. and Weigel,D. (1988) *Wilhelm Roux's Arch. Dev. Biol.*, **197**, 345–354.
Kassis,A.J., Poole,S.J., Wright,D.K. and O'Farell,P.H. (1986) *EMBO J.*, **5**, 3583–3589.
Kaufman,T.C., Lewis,R.A. and Wakimoto,B.T. (1980) *Genetics*, **94**, 115–133.
Koerner,T.J., Hill,J.E., Myers,A.M. and Tzagoloff,A. (1991) *Methods Enzymol.*, **194**, 477–490.
Lee,M.S., Gippert,G.P., Soman,K.W., Case,D.A. and Wright,P.E. (1989) *Science*, **245**, 635–637.
Lewis,E.B. (1978) *Nature*, **276**, 565–570.
Licht,J.D., Grossel,M.J., Figge,J. and Hansen,U.M. (1990) *Nature*, **346**, 76–79.
Macdonald,P.M. and Struhl,G. (1986) *Nature*, **324**, 537–545.
Martinez-Arias,A. and Lawrence,P.A. (1985) *Nature*, **313**, 639–642.
Mermod,N., O'Neill,E.A., Kelly,T.J. and Tjian,R. (1989) *Cell*, **58**, 741–753.
Michael,W.M., Bowtell,D.D.L. and Rubin.,G.M. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 5351–5353.
Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
Pavletich,N.P. and Pabo,C.O. (1991) *Science*, **252**, 809–817.
Reuter,D. ,Schuh,R. and Jäckle,H. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 5483–5486.
Robertson,H.M., Preston,C.R., Phillis,R.W., Johnson-Schlitz,D.M., Benz,W.K., Engels,W.R. (1988) *Genetics*, **118**, 461–470.
Röder,L., Vola,C. and Kerridge,S. (1992) *Development*, **115**, 1017–1033.
Rubin,G.M. and Spradling,A.C. (1982) *Science*, **218**, 348–353.
Rüther,U. and Müller-Hill,B. (1983) *EMBO J.*, **2**, 1792–1794.
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) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
Schuh,R., Aicher,W., Gaul,U., Cote,S., Preiss,A., Maier,D., Seifert,E., Nauber,U., Schröder,C.,Kemler,R. and Jäckle,H. (1986) *Cell*, **47**, 1025–1032.
Spradling,A.C. (1986) In Roberts,D.B. (ed.), *Drosophila: A Practical Approach*. IRL Press, Oxford, pp. 175–197.
Struhl,G. (1981) *Nature*, **292**, 335–338.
Sommer,R. and Tautz,D. (1991) *Development*, **113**, 419–430.
Tautz,D. and Pfeifle,C. (1989) *Cell*, **98**, 81–85.
Treier,M., Pfeifle,C. and Tautz,D. (1989) *EMBO J.*, **8**, 1517–1525.
Wagner-Bernholz,J.T., Wilson,C., Gibson,G., Schuh,R. and Gehring,W.J. (1991) *Genes Dev.*, **5**, 2467–2480.
Weigel,D. and Jäckle,H. (1990) *Cell*, **63**, 455–456.
Weigel,D., Jürgens,G., Küttner,F., Seifert,E. and Jäckle,H. (1989) *Cell*, **57**, 645–658.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received on August 16, 1993; revised on October 11, 1993