The EMBO Journal vol.12 no.2 pp.537-544, 1993

The expression pattern of a *Drosophila* homolog to the mouse transcription factor HNF-4 suggests a determinative role in gut formation

Weimin Zhong, Frances M.Sladek[1] and James E.Darnell,Jr

The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

[1]Present address: University of California, Riverside, CA 92521, USA

Communicated by F.Jacob

A *Drosophila* gene, HNF-4(D), was selected by cross-hybridization with a probe to rat HNF-4 (hepatocyte nuclear factor 4), a steroid hormone receptor superfamily member that plays an important role in liver-specific gene expression. The *Drosophila* gene matched the mouse gene in 60 out of 66 amino acids in the zinc finger DNA binding domain and in 140 out of 206 amino acids in the domain that specifies dimerization and ligand binding. HNF-4(D) is expressed in developing *Drosophila* embryos in mid-gut, fat bodies and malpighian tubules, a striking similarity to its limited expression in the adult intestine, liver and kidney of the mouse. Furthermore, *Drosophila* mutant that has a chromosome deletion spanning the HNF-4(D) locus fails to develop tissues where HNF-4(D) is expressed during late embryogenesis. These findings together with the earlier realization that the rat hepatocyte nuclear factor 3 (HNF-3) and *forkhead*, a *Drosophila* gene required for anterior and posterior gut formation, had virtually the identical DNA binding domain, lead us to speculate that a group of genes that participate in gut formation of invertebrates has survived in evolution to perform similar functions in mammals.

Key words: DNA binding domain/*Drosophila*/embryogenesis/HNF-4(D)/transcription factor

Introduction

In the development of specialized vertebrate cells such as the hepatocyte, the possibility of a regulated cascade of transcription factors extending from the fertilized egg to an adult differentiated cell implies the potential existence of a very large number of serially-dependent regulatory proteins. Nevertheless, in the face of a possibly protracted effort to understand hepatocyte development, a large effort has been made to identify and clone genes encoding transcription factors that participate in adult liver-specific transcription (C/EBP, HNF-1,-3,-4, LAP and DBP; for review see Lai and Darnell, 1991; Sladek and Darnell, 1992). The eventual aim of this work is to understand developmental decisions that allow different primitive endodermal cells to make the choices that result in becoming the secretory cells of the salivary gland, thyroid gland, pancreas and liver or the various epithelia of the gastro-intestinal tract and gall bladder. While these recently cloned transcription factors from hepatocytes (C/EBP, HNF-1,-3,-4, LAP, DBP) are only found in a limited number of adult cell types (Landschulz *et al.*, 1988; Frain *et al.*, 1989; Baumhueter *et al.*, 1990; Descombes *et al.*, 1990; Lai *et al.*, 1990; Mueller *et al.*, 1990; Sladek *et al.*, 1990), none of these proteins is strictly liver-specific. For example HNF-1 and -4 are found also in the kidney (Baumhueter *et al.*, 1990; Sladek *et al.*, 1990) and C/EBP is found in brain and fat cells, all tissues that are not of a endodermal origin (Birkenmeier *et al.*, 1989; Ruppert *et al.*, 1990; Xanthoupoulous *et al.*, 1989).

Recently it was realized that one family of these proteins, the hepatocyte nuclear factor 3 (HNF-3) family, has a highly conserved novel DNA binding domain that is also highly conserved (86 of 110 amino acids) in the *Drosophila* homeotic gene *forkhead* (Weigel and Jackle, 1990; Lai *et al.*, 1991). Since *forkhead* plays a critical role in the embryogenesis of the gut, malpighian tubule and salivary gland (Jurgens and Weigel, 1988; Weigel *et al.*, 1989a,b), it seems logical that HNF-3 could have important early embryonic functions in the mouse. Thus the cloned mammalian transcription factor genes that encode the adult liver factors may allow more rapid access to learning about mammalian endodermal choices than we could earlier have had reason to hope.

In this paper we describe the isolation of a *Drosophila* homolog to HNF-4, a steroid hormone receptor superfamily member that helps activate transcription of a wide array of genes in the liver (Costa *et al.*, 1989; Sladek *et al.*, 1990; Mietus-Snyder *et al.*, 1992). The identity in amino acid sequence between the presumed functional domains of HNF-4 and the *Drosophila* homolog is striking: 60 out of 66 amino acids in the zinc finger DNA binding domain and 140 out of 206 amino acids in the protein domain responsible for dimerization and ligand binding in the protein. The mRNA of the *Drosophila* HNF-4 counterpart is deposited in eggs uniformly (a maternal mRNA) and is only detected in the terminal regions during late nuclear divisions preceding cellularization, it then disappears before gastrulation and then reappears in the mid-gut, fat bodies and malpighian tubules in later organogenesis. Furthermore, a *Drosophila* mutant that has a chromosome deletion spanning the HNF-4(D) locus, which therefore does not express HNF-4(D) zygotically, fails to develop the tissues where HNF-4(D) is expressed during normal late embryogenesis. These results support an essential role of HNF-4 in early gut formation and in organogenesis that may have been preserved during evolution from invertebrates to vertebrates.

Results

A Drosophila HNF-4 homolog

To search for a *Drosophila* HNF-4 homolog, a *Drosophila* genomic DNA library was screened under moderately stringent conditions (see Materials and methods) with a labeled rat cDNA probe encoding the zinc finger domain of the rat HNF-4 [HNF-4(R)] sequence. Positive clones were selected and the cross-hybridizing segments identified, cloned

and sequenced. As the sequencing of the genomic DNA clones progressed, it became obvious that indeed a close sequence relationship existed between the probe and scattered regions of the *Drosophila* genomic DNA. Since many steroid thyroid hormone receptor genes have widely scattered exons, it seemed likely this was the case for the HNF-4-related *Drosophila* gene as well.

```
  1 MHADALASAYPAASQPHSPIGLALSPNGGGIGLSNSSNQSSENFALCNGN
 51 GNAGSAGGGSASSGSNNNNSMFSPYNNLNGSGSGTNSSQQQLQQQQQQQS
101 PTVCAICGDRATGKHYGASSCDGCKGFFRRSVRKNHQYTCRFARNCVVDK
151 DKRNQCRYCRLRKCFKAGMKKEAVQNERDRISCRRTSNDDPDPGNGLSVI
201 SLVKAENESRQSKAGAAMEPNINEDLSNKQFASINDVCESMKQQLLTLVE
251 WAKQIPAFNELQLDDQVALLRAHAGEHLLLGLSRPSMHLKDVLLLSNNCV
301 ITRHCPDPLVSPNLDISRIGARIIDELVTVMKDVGIDDTEFACIKALVFF
351 DPNAKGLNEPHRIKSLRHQILNNLEDYISDRQYESRGRFGEILLILPVLQ
401 SITWQMIEQIQFAKIFGVAHIDSLLQEMLLGGELADNPLPLSPPNQSNDY
451 QSPTHTGNMEGGNQVNSSLDSLATSGGPGSHSLDLEVQHIQALIEANSAD
501 DSFRAYAASTAAAAAAAVSSSSSAPASVAPASISPPLNSPKSQHQHQQHA
551 THQQQQESSYLDMPVKHYNGSRSGPLPTQHSPQRMHPYQRAVASPVEVSS
601 GGGGLGLRNPADITLNEYNRSEGSSAEELLRRTPLKIRAPEMLTAPAGYG
651 TEPCRMTLKQEPETGY
```

Fig. 1. Deduced amino acid sequence of *Drosophila* HNF-4 protein. Nucleotide sequence (not shown) of HNF-4(D) cDNA was based on o5-2-1 clone and verified from at least one other source. The probable initiator methionine is marked 1, the asterisks denote the cysteines that make up the two zinc fingers of the DNA binding domain.

A cDNA sequence containing the complete coding region was achieved by selection and sequencing of a *Drosophila* HNF-4 [HNF-4(D)] clone from an ovary cDNA library. The *Drosophila* cDNA is capable of encoding a 666 amino acid protein (Figure 1), longer than the 455 amino acids of the rat and 465 amino acids of the mouse proteins. Comparing the sequences of the *Drosophila* and rat (or mouse) sequence revealed a striking identity in the zinc finger (DNA binding) domain: 60 of 66 amino acids are identical. The 12 amino acids immediately after the zinc finger are also completely conserved (Figure 2B). This is a higher degree of identity than between any *Drosophila* steroid receptor family members (Figure 2A) or mammalian receptor family members, even those that bind the same DNA site such as the Arp-1 or glucocorticoid and mineralocorticoid receptors (Evans, 1988; Sladek *et al.*, 1990; Mietus-Snyder, 1992). In addition to the sequence identity, the genomic coding structure is the same in the two HNF-4 genes. One exon encodes five of the eight cysteines of the two zinc fingers and the following exon encodes the remaining three cystines (Figure 2B). In most steroid family members, each zinc finger (four cysteines) is encoded in a separate exon.

In addition to the DNA binding region, another section of the protein between amino acid 222 and 432 in HNF-4(D), and 163 and 369 in HNF-4(R) shows 68% amino acid identity (140 of 206 amino acids, Figure 2B). This region of the receptor proteins directs dimerization as well as ligand

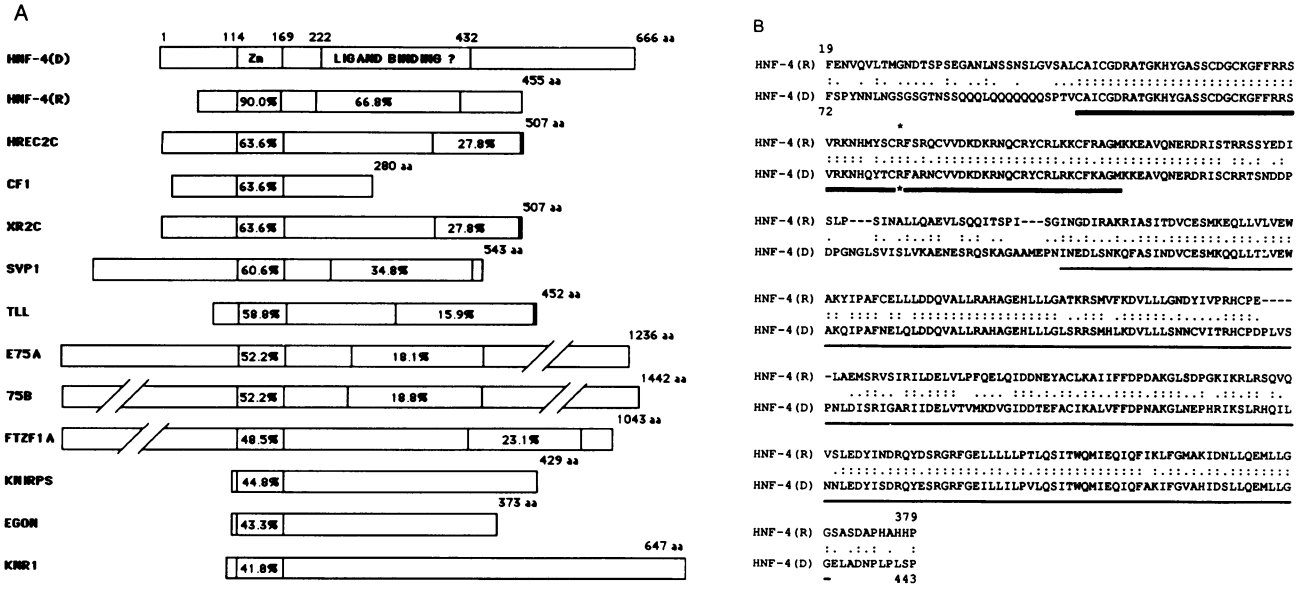

Fig. 2. Comparison of HNF-4 proteins and members of steroid hormone receptors in *Drosophila melanogaster*; (**A**) Sequence and structure similarity between HNF-4 proteins and members of steroid hormone receptor superfamily in *D.melanogaster*. The primary amino acid sequences of *Drosophila* HNF-4(D) were compared with HNF-4(R) and *Drosophila* steroid hormone receptors using the FASTA program (Pearson and Lipman, 1988). Percentages denote amino acid identity within the zinc finger (Zn) and ligand binding domains. The various genes include (HREC2C), a *Drosophila* steroid hormone receptor-like protein (Henrich *et al.*, 1990); (CF1) *Drosophila* chorioin factor CF1 (Shea *et al.*, 1990); (XR2C) *Drosophila* XR2C protein (Oro *et al.*, 1990); (SVP1) *Drosophila seven-up* protein type 1 (Mlodzik *et al.*, 1990); (TLL) *Drosophila tailless* protein (Pignoni *et al.*, 1990); (E75A) *Drosophila E75E* protein (Feigl *et al.*, 1989; Segraves and Hogness, 1990); (75B) *Drosophila* hypothetical 75B protein (Feigl *et al.*, 1989; Segraves and Hogness, 1990); (FTZF1A) *Drosophila ftz-f1* protein (Lavorgna *et al.*, 1991); (KNIRPS) *Drosophila knirps* protein (Nauber *et al.*, 1988); (EGON) *Drosophila embryonic gonad* protein (Rothe *et al.*, 1989); (KNR1) *Drosophila knirps-related* protein (Oro *et al.*, 1988). (**B**). Comparison of the putative *Drosophila* HNF-4 protein with rat HNF-4. Amino acid sequences of *Drosophila* HNF-4 [HNF-4(D)] and rat HNF-4. [HNF-4(R)] were compared using the FASTA program. (:) indicates amino acids that are identical; (•) indicates conservative amino acid changes. Sequences were aligned for maximum similarities by introducing gaps (hyphens). 60 out of 66 of the HNF-4(R) amino acids in the DNA binding domain (double underlined) are identical, while 140 out of 206 amino acids in the presumptive ligand binding domain (single underlined) are identical. Asterisks denote the splicing junction in the DNA binding domain that has been conserved between mouse and *Drosophila* (see text).

binding in those family members known to bind ligands (Evans, 1988). There is no known ligand for HNF-4. This high level of sequence identity contrasts with 20–40% identity between different members of the superfamily in this region (Figure 2B). Thus by sequence comparison and gene structure, HNF-4(R) and HNF-4(D) are very closely related.

HNF-4(D) is a site-specific DNA binding protein

The close similarity in amino acid sequence in the zinc finger region of HNF-4(R) and HNF-4(D) strongly suggests that HNF-4(D) is also a DNA binding protein. To test this possibility, we inserted the HNF-4(D) sequence into a transcription vector, produced the corresponding mRNA and translated it *in vitro*. The HNF-4(D) is 666 amino acids long compared with HNF-4(R), which is 455 amino acids long. The translation product of the HNF-4(D) mRNA, when mixed with a labeled deoxyoligonucleotide carrying an HNF-4 binding site, produced a specific gel retardation band that migrated more slowly than the characteristic HNF-4(R) band and was competed specifically by the unlabeled HNF-4 binding site. An antibody directed against the conserved region of the two proteins also recognized the HNF-4(D) translation product (Figure 3).

HNF-4(R), like all steroid thyroid family members, forms homodimers (Sladek *et al.*, 1990), but does not form heterodimers with several members of the superfamily that have been tested, presumably because their dimerization domains are too dissimilar (F.M.Sladek, unpublished observation; Mietus-Snyder *et al.*, 1992). However, when the HNF-4(R) and HNF-4(D) mRNAs were co-translated and then tested in the gel-shift assay, the translation products formed an intermediate gel-shift band indicative of heterodimer formation (Figure 3). This is further evidence of the close relationship between these two proteins, which has been maintained throughout a great evolutionary distance.

Chromosomal location of HNF-4(D)

To begin the analysis of possible functions for HNF-4(D), a digoxygenin-labeled probe was prepared and hybridized to chromosome preparations from the salivary glands of wild type third instar larvae. One single band of hybridization was found in the left arm of chromosome 2, in the general region of 29E (data not shown). A number of strains that have deletions spanning this region are known, all of which are lethal in homozygotes (Wustmann *et al.*, 1989). In the salivary glands of heterozygotes, where one half of the polytene chromosome has a deletion in the 29E region, resulting in split salivary gland chromosomes, hybridization occurred to only half of the polytene chromosome 2 (Figure 4). This confirms that the HNF-4 gene is deleted in one of the strains giving rise to the heterozygous flies.

The mRNAs of HNF-4(D)

A radiolabeled probe from a coding region of the HNF-4 gene was prepared for use in Northern blots to identify the size and time of appearance of mRNAs from the gene. Two mRNA bands of ~4.6 and 3.3 kb were observed in several different RNA samples. A strong signal was observed in the earliest sample (0–2 h), which contained mostly early embryos. The 2–4 h and 4–6 h samples had only a trace of each mRNA and later samples (8–12 h and 16–24 h) again contained significant amounts of mRNA, with the 8–12 h predominantly containing the 3.3 kb transcript and the 16–23 h predominantly containing the 4.6 kb species (Figure 5A).

In situ hybridization

In order to determine more carefully the time of mRNA presence and determine its localization within the fertilized egg, *in situ* hybridization to detect the HNF-4(D) mRNA was carried out. The strongest reaction was in stage 1 and 2 embryos where the mRNA was uniformly distributed. Furthermore, *in situ* hybridization of the ovaries indicated that this RNA was deposited in the egg by the nurse cells (data not shown). During cleavage stages, retention of the mRNA to the peripheral regions of the syncytial blastoderm was noted. Just before cellularization, the only detectable stain in the syncytial blastoderm was terminal, with the posterior end being stained more strongly than the anterior [Figure 5B (a–d)].

From ~2–3 h to 6 h after fertilization, there was no detectable HNF-4(D) mRNA. The mRNA then reappeared between 6 and 8 h, initially in the endodermal cells corresponding to the invaginating posterior midgut primordium and later also in the anterior midgut primordium [Figure 5B (e–f)]. The stain grew more intense and definitely conformed to the distribution of the dividing endodermal cells in the midgut, which form Y-shaped internal structures (Figure 6b; see Poulson, 1950). The cells of the foregut and hindgut contained little or no HNF-4(D) mRNA. Still later

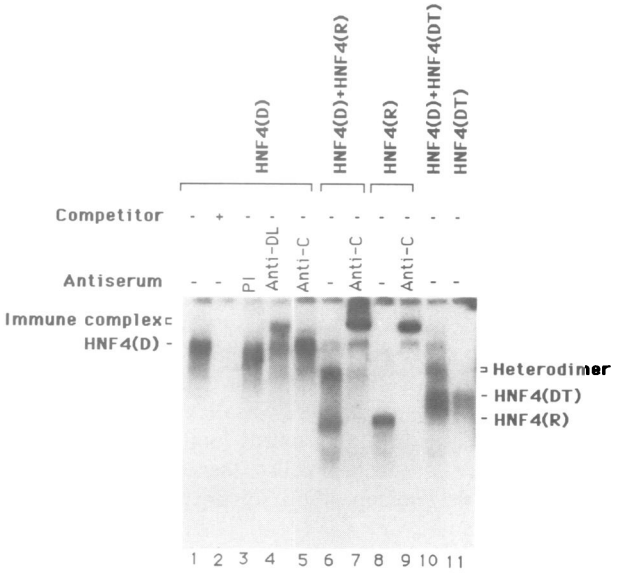

Fig. 3. *In vitro* synthesized *Drosophila* HNF-4 protein binds to a mammalian HNF-4 DNA recognition site and forms heterodimers with rat HNF-4. HNF-4 proteins (rat or *Drosophila* or truncated versions of these proteins) were prepared *in vitro* and used in gel-shift assays in the absence (–) or presence (+) of competing unlabeled oligonucleotide or various antisera to rat HNF-4 as indicated (see Materials and methods). Lanes 1–5, HNF-4(D) protein. PI is the preimmune serum. Anti-DL is a rabbit antibody against a portion of the rat HNF-4 DNA binding and ligand binding domains that are conserved in *Drosophila*. Anti-C is an antibody against the very C-terminus of the rat HNF-4 that is not conserved in *Drosophila*; lanes 6 and 7, *in vitro* transcribed HNF-4(D) RNA was mixed with HNF-4(R) RNA and then co-translated *in vitro*. An intermediary band was seen (lane 6) which can be supershifted by anti-C antibody (lane 7); lanes 8 and 9 HNF-4(R) protein; lanes 10 and 11, RNA encoding the truncated HNF-4(D) protein was either mixed with the RNA coding for the full length HNF-4(D) and co-translated (lane 10) or translated alone (lane 11). The intermediary band can be seen (lane 10).

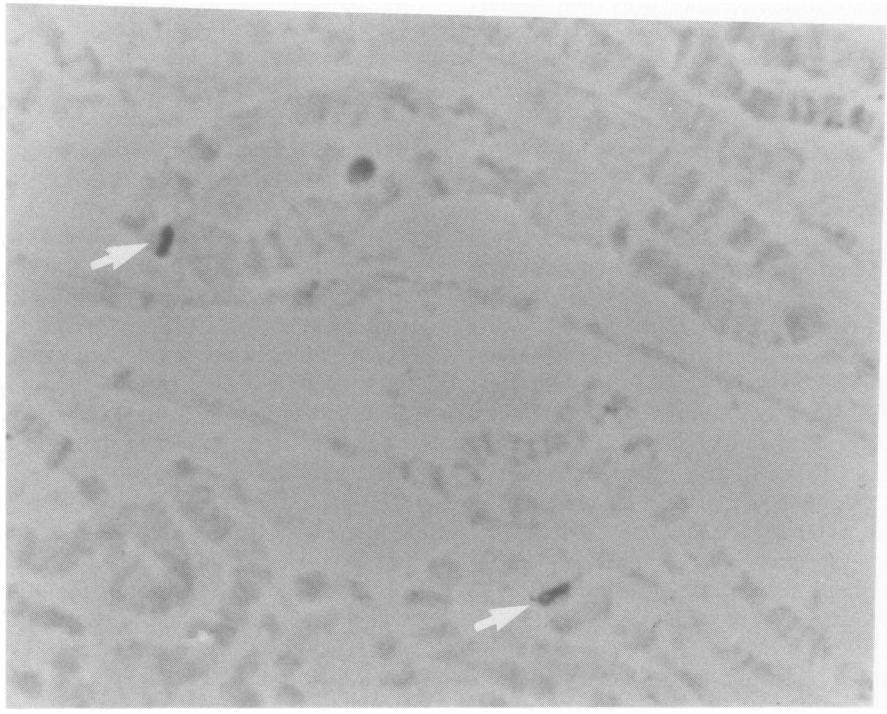

Fig. 4. Cytological analysis of HNF-4(D). *In situ* hybridization of the salivary gland polytene chromosomes from a third instar larva of the strain TE196X1 (Wustmann, 1989) that is heterozygous for a deficiency at 29E2-F1 to 30C2-4. Using a digoxygenin-UTP-labeled HNF-4(D) genomic DNA fragment (encodes a part of the ligand binding domain), hybridization was detected on half of the heterozygous chromosome around region 29E.

stages (stage 14/15) show a variety of tissues that contain HNF-4(D) mRNA. These include fat bodies, malpighian tubules, salivary glands and one cluster of cells on either side of each of the abdominal segments of the embryo, the nature of which is uncertain but may be related to the peripheral nervous system or endocrine system [Figure 5B(g−h), Figure 6(b−c)]. Note that the HNF-4(D) staining in the malpighian tubules was confined to the distal part of each tubule [Figure 5B (g−h)]. This distal region of the tubule contains dividing cells that are responsible for the elongation of the malpighian tubules. At the end of stage 15, when the fused midgut has contracted to form four loops, the most staining in the midgut was observed in the midgut caeca and in the first and fourth loops from which gut primordia (nests of imaginal cells rather than imaginal disks) arise in larvae (Poulson, 1950; Figure 6f).

Comparison of distribution of HNF-4(D) mRNA with forkhead mRNA

It was recognized recently that the *Drosophila* homeotic gene *forkhead* had significant amino acid identity within the DNA binding domain of three members of the HNF-3 family of mouse transcription factors (Weigle and Jackle, 1990; Lai *et al.*, 1991).

A panel of embryos stained for *forkhead* mRNA was assembled to compare with the distribution of HNF-4 mRNA (Figure 6). As described elsewhere (Weigle *et al.*, 1990), the initial expression of *forkhead* mRNA was in the terminal region at late synctial blastoderm stage. While we have not been able to compare exactly the same embryos for the time course of appearance in terminal cells of HNF-4(D) and *forkhead* mRNA, during late stages of gut development, the internal *forkhead* staining followed a different pattern from HNF-4(D). A sharp division is seen [Figure 6(a−e)] between foregut, hindgut and midgut with *forkhead* in the foregut and hindgut and HNF-4(D) in the midgut. *forkhead* mRNA was present in the salivary gland and the malpighian tubules, as was HNF-4(D), but not in the fat bodies where HNF-4(D) mRNA was prominent [Figure 6(a−e)].

Phenotypic analysis of the chromosome deletion mutant

One of the *Drosophila* mutant strains that have chromosome deletions spanning the HNF-4(D) locus (Figure 4) was used to search for an HNF-4(D) developmental function. While, as expected, the maternal expression of the HNF-4(D) was not affected in the embryos homozygous for the deletion, the zygotic expression after about stage 10 was expected to be abolished. *forkhead* mRNA staining was used as a marker to crudely assess the embryonic development in these homozygotes. Early mid-gut, malpighian tubule and salivary gland development appeared to be normal (Figure 7A and B). However, after stage 10, when the HNF-4(D) mRNA reappears in wild type embryos, there were clearly visible defects in midgut, malpighian tubule and salivary gland development in the mutants. Both the anterior and posterior midgut failed to further invaginate, the malpighian tubules failed to grow and the salivary gland invagination was arrested [Figure 7 (C−F)]. At around stage 16, the endodermal part of the midgut was clearly missing, the malpighian tubules were not formed and the salivary glands did not invaginate properly and were reduced significantly in size [Figure 7 (G and H)]. The foregut and hindgut development was much less affected.

Discussion

We have described here a previously undetected *Drosophila* gene that has a striking sequence similarity to the mammalian transcription factor termed HNF-4, a positively acting member of the steroid hormone receptor superfamily of transcription factors (Sladek *et al.*, 1990). Like HNF-4(R), the *Drosophila* protein can dimerize and bind to a DNA site defined as an HNF-4 site. Thus it seems highly likely that HNF-4(D) is also a positively acting transcription factor. The close sequence similarity of the rat and the *Drosophila* proteins raises the question of homologous function of the two proteins. The rat protein is a prominent constituent in adult liver, intestine and kidney nuclear extracts and can positively activate genes that are transcribed preferentially in these tissues. HNF-4(R) is not present in most other adult tissues (Sladek *et al.*, 1990). Even though the tissue distribution of HNF-4(D) in adult flies is unknown, in late *Drosophila* embryos, i.e. stage 15, the mRNA is found in the midgut, fat bodies and malpighian tubules (excretory organs), a startling recapitulation of its distribution in adult mouse intestine, liver and kidney. This highly conserved tissue distribution between mouse and *Drosophila* strongly suggests that HNF-4 may play a critical role during organogenesis of these tissues that has been conserved throughout animal evolution. The phenotype of *Drosophila* mutants with large chromosome deletions spanning the HNF-4(D) locus significantly substantiated this speculation.

Embryos homozygous for such deletions were drastically altered in midgut, malpighian tubule and salivary gland formation around stage 10, when the zygotic expression of HNF-4(D) started. In contrast, the development of both foregut and hindgut was much less affected. A point of particular interest was the localized expression of HNF-4(D) mRNA [Figure 6b and Figure 5B (g and h)] within the midgut and malpighian tubules. As tested by preliminary antibody staining experiments, HNF-4(D) protein shows a similar localization (data not shown). HNF-4(D) is highly expressed in the part of the midgut that corresponds to regions of the dividing endodermal cells that are responsible for the further invagination of the anterior and posterior midgut (Poulson, 1950). The rest of the midgut shows little, if any, HNF-4(D) expression. In the developing malpighian tubules, the HNF-4(D) protein appears to be localized in the very distal tip, the region in which cell division and tubule elongation occurs (Skaer, 1989). These locations are in accord with the arrest in the deletion strain of the endodermal midgut invagination and the cessation of the malpighian tubule elongation after stage 10, the time at which zygotic expression of HNF-4(D) begins in the wild type embryo. Because the deletions in the 29E region are large, we cannot be certain that the described phenotypes are due solely to HNF-4(D) loss. But such is possible and the phenotypes correlate well with the pattern of tissue distribution of HNF-4(D).

Do HNF-4(D) or (R) have any functions earlier in development than organogenesis? We have recently found that in cultured hepatoma cells, HNF-4(R) is a required positively acting factor in the expression of the genomic copy of another liver-enriched transcription factor, the hepatocyte nuclear factor 1 (HNF-1), a POU homeodomain protein that is also strongly expressed only in liver, intestine and kidney (Kuo *et al.*, 1992). Thus HNF-4 may be higher in a development hierarchy that leads to liver (endodermal) differentiation. Thus we were particularly interested in any very early role that HNF-4(D) might play as a guide to early effects in mammals.

HNF-4(D) has a very strong maternal mRNA signal that is uniformly distributed in the fertilized egg. Though we do not yet know the distribution of HNF-4(D) protein at this stage, we did observe the restriction of HNF-4(D) RNA in the terminal regions at the end of syncytial blastoderm preceding cellularization, with the expression at the posterior terminus being stronger than that of the anterior. The termini of the embryo contain the blastoderm anlagen of internal organs, including the midgut. Cells in the anterior region invaginate to become foregut and anterior midgut, while cells at the posterior end of the embryo contribute both to hindgut and to posterior midgut. While no detectable HNF-4(D) RNA was found during early gastrulation and midgut invagination, the mRNA reappears in the invaginating posterior and anterior midgut around stage 10. This expression pattern suggests that HNF-4(D) may be a component of the terminal organizer system, which gives rise to the unsegmented larval parts (acron and telson) and internal organs including midgut. Unlike the anterior–posterior or dorsal–ventral system, the maternal terminal patterning seems to require a locally activated signal transduction pathway. *torso*, a maternally encoded protein that by sequence analogy is a transmembrane tyrosine kinase, is a key component of this system. *torso* mutants lack the structures posterior to the seventh abdominal segment and internal organs including hindgut and posterior midgut. (Nusslein-Vollhard *et al.*, 1987; Klinger *et al.*, 1988; Strecker *et al.*, 1989). Two gap genes, *tailless* and *huckebein*, seem to execute the *torso*-initiated developmental pathway of the gut. *tailless* mutants fail to develop hindgut and the posterior midgut is reduced in size, while *huckebein* mutants lack the midgut anlage completely. Furthermore, *tailless* and *huckebein* double mutants are *torso* phenocopies and *torso* gain-of-function mutants are dependent on *tailless* and *huckebein* gene products (Pignoni *et al.*, 1990; Weigel *et al.*, 1990). However, it remains unknown what, if any, maternally contributed transcription factors exist in the hierarchy between *torso*, *tailless* and *huckebein*. *torso* is a tyrosine kinase cell surface receptor that is activated by its ligand in the embryonic poles and transmits its signal via the downstream gene *D-raf*, a serine-threonine kinase (Nishida *et al.*, 1988; Ambrosio *et al.*, 1989). It seems logical that some maternally provided transcription factors should be required to respond to the signals produced by *torso* or *D-raf*, which will in turn activate the two gap genes *tailless* and *huckebein*. It seems possible that the maternal expression of HNF-4(D) mRNA and its retention in the limited posterior position just before cellularization are related to *tailless* and/or *huckebein* expression. Specific mutations in the HNF-4 locus that would result in a maternal deprivation of HNF-4 mRNA are required to test this possibility.

Regardless of the resolution of the possible early function of HNF-4(D) and any analogous early function in mice, the findings to date raise the interest in HNF-4 as an important gene in the differentiation of various specialized cell types that arise from both endoderm and mesoderm. In analogy with the function of *Antennapedia* in determining the segment identity in flies and of *hox* proteins in vertebrae

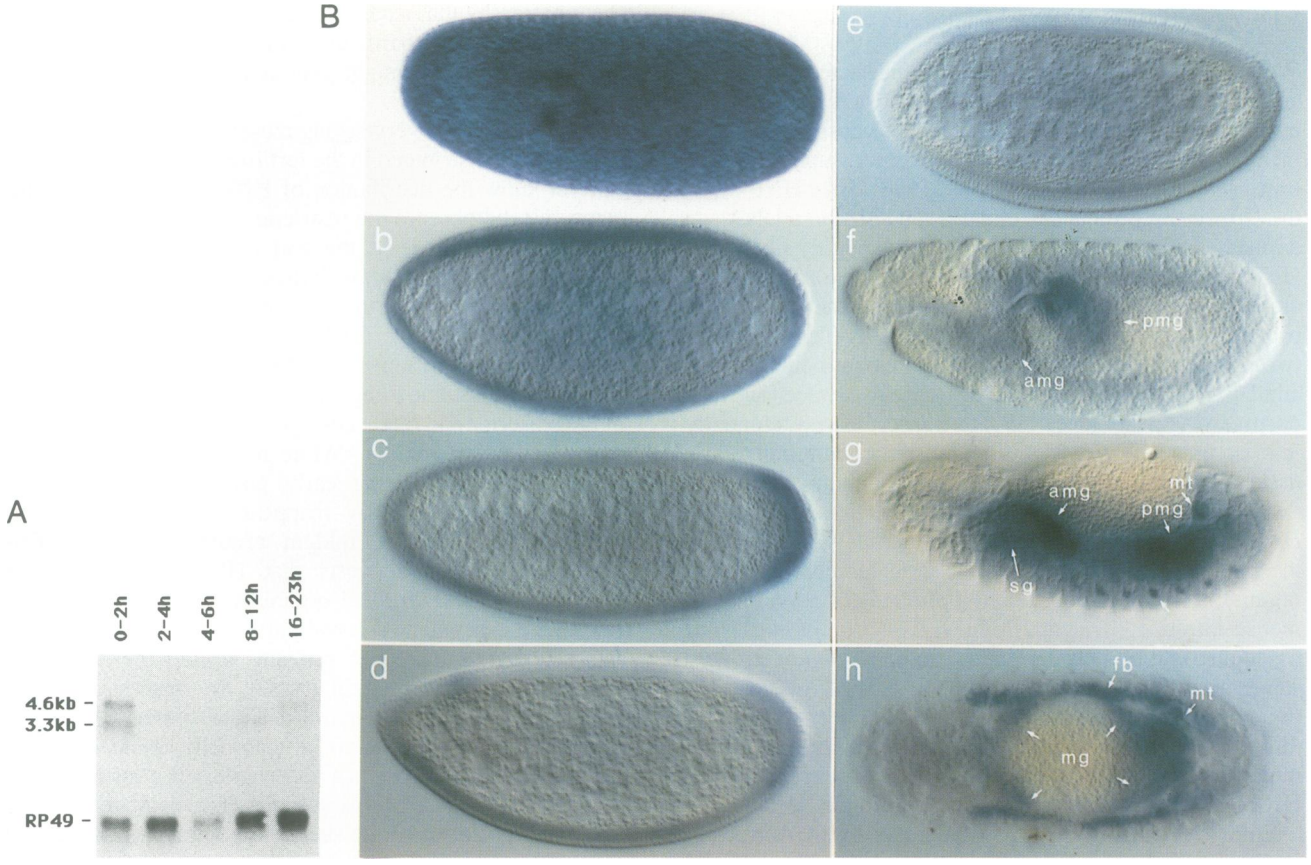

Fig. 5. Expression of *Drosophila* HNF-4 RNA during embryogenesis. (**A**). Northern blot analysis of total RNA from wild type embryos collected during different embryonic stages using an HNF-4(D) genomic DNA fragment containing part of the ligand binding domain as a probe. A probe for ribosomal protein RP49 served as a control. (**B**) *In situ* hybridization to whole mount wild type embryos. Embryos are stained with antisense HNF-4(D) RNA probe, labeled with digoxygenin-UTP. All views are lateral, with anterior to the left, unless otherwise indicated. (a) Stage 2. Staining of HNF-4(D) is uniform throughout the embryo. (b–e) Transition from late stage 3 to the end of stage of 5. Staining of HNF-4(D) can be visualized around the periphery of the embryo. (b, late stage 3/4), then is localized at the posterior terminus of the embryo (c and d, stage 4/5), and finally disappears at the end of cellulization, before gastrulation (e, stage 5). The unstained cells at the posterior terminus are the pole cells. (f) Stage 11. Staining of HNF-4(D) reappears in the developing posterior midgut primordium (pmg). (g–h) Stage 14, dorsal view (h). Staining of HNF-4(D) can be visualized in the midgut (mg), fat bodies (fb), malpighian tubules (mt), and salivary gland (sg, not seen here due to optical section).

formation in mice (Akam, 1989; Kessel and Gruss, 1990), we suggest that HNF-4(R) and HNF-4(D) conservation and the conservation between *forkhead* and the HNF-3 family suggests conservation throughout animal evolution of a network of genes responsible for gut formation.

Materials and methods

Isolation of genomic and cDNA clones

Rat HNF-4 5′ end probe containing most of the zinc finger DNA binding domain was labeled with ^{32}P by random priming (Feinberg and Vogelstein, 1983). The probe was used to screen a *Drosophila* λEMBL3 genomic DNA library as described in Maniatis *et al.* (1982) with modifications (Sladek *et al.*, 1990). Hybridization was at 42°C with 2×SSPE, 0.1% SDS, 5×Denhardt's solution, 0.1 mg/ml denatured salmon sperm DNA and 1 ng/ml of denatured probe. The highest stringency wash was 1×SSC, 0.1% SDS at 42°C. About 40 positive plaques were obtained out of a screen of 6×10^5 phages. One of them, D2, was fully analyzed by restriction mapping and hybridization with probes from both the zinc finger and ligand binding domains of the rat HNF-4. Several fragments that hybridized to the rat sequence were subcloned and sequenced using the dideoxy method (Sanger *et al.*, 1977). One of the fragments, a 0.5 kb *Bam*HI–*Eco*RI fragment (D2BE3) containing part of the ligand binding domain, was used to screen a *Drosophila* λZAPII ovary cDNA library as described above, hybridization and washing were carried out at 65°C. Five positive clones were obtained, one of which, o5-2-1, was fully sequenced by subcloning or sequencing with internal oligonucleotides. The sequence matched known genomic sequences. The other clones were partially sequenced, all matching o5-2-1 clone in the coding region, but were all truncated within the coding region.

Mobility shift assay and in vitro translation of HNF-4(D)

Drosophila HNF-4 cDNA clone o5-2-1 in pBluescript SK(−) was linearized with either *Xba*I (in the 3′ polylinker) or *Hpa*I (within the coding region, at position 1858) and transcribed *in vitro* with T7 polymerase. *In vitro* translation using rabbit reticulocyte lysate and [^{3}H]leucine was performed as recommended by the manufacturer (Promega). Rat HNF-4 (from clone pf7) was synthesized as described by Sladek *et al.* (1990). Gel mobility shift (DNA binding) assays (Fried and Crothers, 1981) were carried out in 15 μl reactions in shift buffer (20 mM HEPES pH 7.9, 40 mM KCl, 2 mM $MgCl_2$, 1 mM DTT and 0.5 mM EGTA pH 8.0) and contained 2 μg polyd(I-C), 0.5–2 μl *in vitro* translated protein mixture and 0.5 ng double-stranded APF-1 probe labeled with ^{32}P by the Klenow fragment (Sladek *et al.*, 1990). Reactions were incubated at room temperature for 30 min. Unlabeled APF-1 probe as specific competitor and antisera were added as indicated. 6 μl of shift reaction were loaded on a 6% polyacrylamide gel run in 25 mM Tris–borate and 0.25 mM EDTA at 4°C to separate protein bound DNA complexes from free probes.

In situ hybridization to polytene chromosomes

Polytene chromosomes were prepared from the salivary gland of late third instar larvae of either wild type or strain TE196x1 according to Pardue (1986) with minor modifications. Chromosomes were denatured in 0.07 N NaOH and hybridized with 0.5 kb segment of HNF-4(D) genomic DNA (D2BE3) labeled with digoxygenin-UTP by random priming (Boehringer Mannheim). Hybridization was carried out at 37°C in the presence of 3×SSC, 50% formamide and 0.1 mg/ml denatured salmon sperm DNA. Signal detection

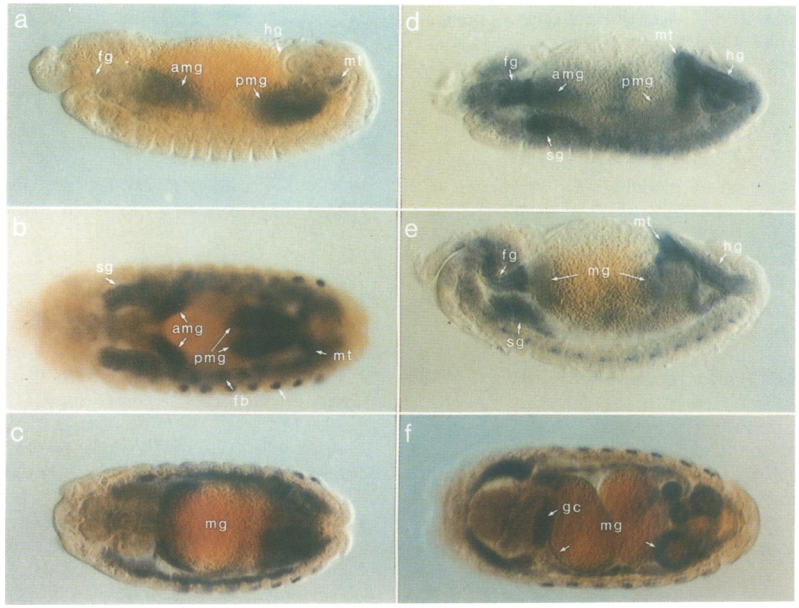

Fig. 6. Tissue distribution of HNF-4(D) RNA is distinct from that of *forkhead (fkh)* during late embryogenesis. Wild type embryos are stained either with digoxygenin-UTP labeled HNF-4(D) antisense RNA probe (a–c and f), or a *fkh* probe (d–e). (**a**) Stage 12/13. Both anterior (amg) and posterior (pmg) midgut primordium are stained by HNF-4(D) probe. (**b**) Stage 13/14, dorsal view. The anterior and posterior midgut are stained in Y-shaped configuration. HNF-4(D) staining in the fat bodies (fb), malpighian tubules (mt), salivary gland (sg) and clusters of ectodermal cells on either side of each of the abdominal segments (arrowhead) are also evident. Note the absence of the staining in the foregut (fg) and hindgut (hg). (**c**) Stage 15, dorsal view. The midgut (mg) has already formed a closed tube. (d–e) Stage 13/14(d) and 14/15(e). *fkh* staining in the foregut and hindgut, but not in the midgut. Also notice the absence of the staining in the fat bodies. Salivary glands and malpighian tubules show strong *fkh* staining. (**f**) Stage 16, dorsal view. Midgut convolutions are clearly visible. HNF-4(D) staining in the midgut is most visible in the first and fourth convolution (arrowhead) and midgut caeca(gc).

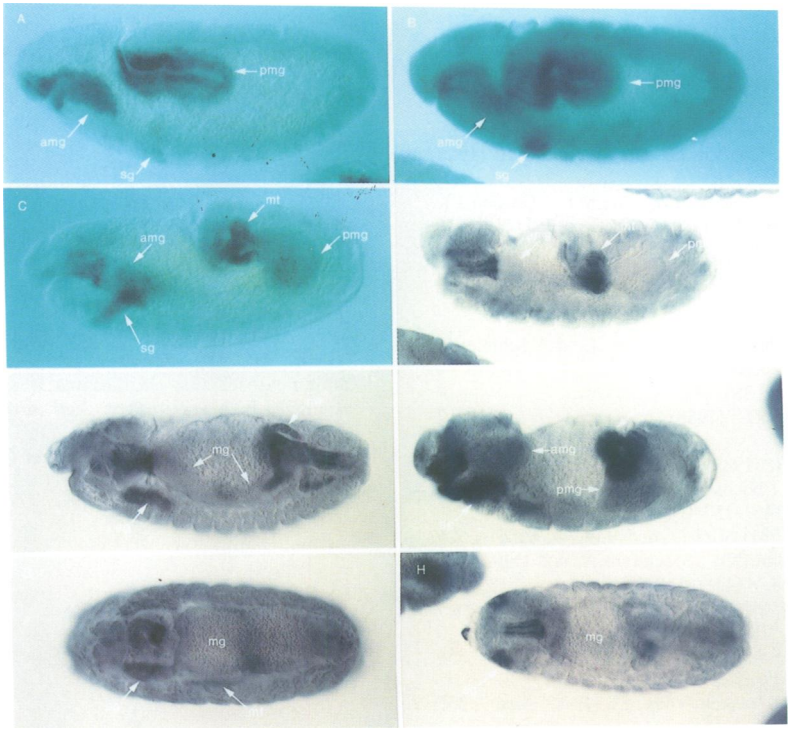

Fig. 7. Phenotypes of the chromosome deletion mutant TE196X1. Wild type embryos (A, C, E and G) and embryos homozygous for the chromosome deletion (B, D, F and H) are stained with digoxygenin-labeled *fkh* antisense RNA probe. (A and B) Stage 9, lateral view; (C and D) Stage 11, lateral view; (E and F) Stage 13/14, lateral view; (G and H) Stage 15/16, dorsal view.

was performed with alkaline phosphatase conjugated antidigoxygenin antibody as described elsewhere (Tautz and Pfeifle, 1989).

Northern blot analysis

Total RNA was extracted from different stages of wild type embryos according to Jowett (1986). RNA (5 μg/lane) was electrophoresed in a 1% agarose–formaldehyde gel as described in Sambrook *et al.* (1989). The RNA was transferred to Immobilon-N (Millipore) and probed according to the protocol suggested by the manufacturer. HNF-4(D) mRNA was detected with a 0.5 kb [^{32}P]dATP random-primed genomic fragment (D2BE3). The high stringency wash was with 0.1×SSC and 0.1% SDS at 65°C for 30 min. The filter was probed with the control ribosomal protein

(RP49) probe first and re-probed with HNF-4(D) probe without stripping. The autoradiography with the HNF-4 probe was exposed for 5 days with two intensifying screens at −70°C. Mouse rRNA (28S and 18S) was used as size markers.

In situ hybridization to whole mount embryos

Embryos were collected, fixed and hybridized with RNA probes according to Tautz and Pfeifle (1989) with minor modifications. Fragments of HNF-4(D) or *fkh* cDNA were cloned into transcription vector pBluescript KS(−) and transcribed in the presence of digoxygenin-UTP with either T3 or T7 RNA polymerase to generate antisense or sense RNA probes. Size reduction of the RNA probes was carried out in 60 mM Na_2CO_3 and 40 mM $NaHCO_3$ (pH 10.2) at 65°C for 40 min. Embryonic stages were judged according to Campos-Ortega and Hartenstein (1985).

Identifying TE196X1 embryos homozygous for the chromosome deletion

Embryos homozygous for the chromosome deletion were recognized either by the lack of HNF-4 RNA staining (after stage 11, from parents with Df/SM5) or by the lack of *lacZ* RNA staining (before stage 11, from parents with Df/SM6Bevelac). SM6Bevelac is the second chromosome balancer that carrys P[ry+, eve-lac8.0] which gives *even-skipped*-like expression of the *lacZ* reporter gene (provided by S.Beckendorf).

Acknowledgements

This work was supported by a grant from the National Institutes of Health (CA16006-18). We especially thank S.Dinardo and members of his lab, P.Gonczy and S.Dougan, for guidance with cytological studies and *in situ* hybridization techniques. We thank J.Tamkun for making available the genomic DNA library, T.Hazelrigg for providing the ovary cDNA library, T.Schubach for supplying the strain TE196X1, S.Beckendorf for making available the SM6Bevelac balancer and A.English for assistance with the manuscript.

References

Akam,M. (1989) *Cell*, **57**, 347−349.
Ambrosio,L., Mahowald,A.P. and Perrimon,N. (1989) *Nature*, **342**, 288−291.
Baumhueter,S., Mendel,D.B., Conley,P.B., Kuo,C.J., Turk,C., Graves,M.K., Edwards,C.A., Courtois,G. and Crabtree,G.R. (1990) *Genes Dev.*, **4**, 372−379.
Birkenmeier,E.H.B., Gwynn,B., Howard,S., Jerry,J., Gordon,J.I., Landschulz,W.H. and McKnight,S.L. (1989) *Genes Dev.*, **3**, 1146−1156.
Campos-Ortega,J.A. and Hartenstein,V. (1985) *The Embryonic Development of Drosophila melanogaster*. Springer-Verlag, Berlin.
Costa,R.H., Grayson,D.R. and Darnell,J.E.,Jr (1989) *Mol. Cell. Biol.*, **9**, 1415−1425.
Descomes,P., Chojkier,M., Lichtsteiner,S., Falvey,E. and Schibler,U. (1990) *Genes Dev.*, **4**, 1541−1551.
Evans,R.M. (1988) *Science*, **240**, 889−895.
Feigl,G., Gram,M. and Pongs,O. (1989) *Nucleic Acids Res.*, **17**, 7167−7178.
Feinberg,A.P. and Vogelstein,B. (1983) *Biochemistry*, **132**, 6−13.
Frain,M., Swart,G., Monaci,P., Nicosia,A., Stampfli,S., Frank,R. and Cortese,R. (1989) *Cell*, **59**, 145−157.
Fried,M. and Crothers,D.M. (1981) *Nucleic Acids Res.*, **9**, 6505−6525.
Henrich,V.C., Sliter,T.J., Lubahn,D.B., MacIntyre,A. and Gilbert,L.I. (1990) *Nucleic Acids Res.*, **18**, 4143−4148.
Jowett,T. (1986) In Roberts,D.B. (ed.), *Drosophila, A Practical Approach*. IRL Press, Oxford, pp. 279−280.
Jurgens,G. and Weigel,D. (1988) *Roux's Arch. Dev. Biol.*, **197**, 345−354.
Kessel,M. and Gruss,P. (1990) *Science*, **249**, 374−379.
Klinger,M., Erdely,M., Szabad,J., Nusslein-Volhard,C. (1988) *Nature*, **335**, 275−277.
Kuo,C.J., Conley,P.B., Chen,L., Sladek,F.M., Darnell,J.E.,Jr and Crabtree,G.R. (1992) *Nature*, **355**, 457−461.
Lai,E. and Darnell,J.E.,Jr (1991) *Trends Biochem. Sci.*, **16**, 427−430.
Lai,E., Prezioso,V.R., Smith,E., Litvin,O., Costa,R.H. and Darnell,J.E.,Jr (1990) *Genes Dev.*, **4**, 1427−1436.
Lai,E., Prezioso,V.R., Tao,W., Chen,W.S. and Darnell,J.E., Jr (1991) *Genes Dev.*, **5**, 416−427.
Landschulz,W.H., Johnson,P.F., Adashi,E.Y., Graves,B.J. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 786−800.
Lavorgna,G., Ueda,H., Clos,J. and Wu,C. (1991) *Science*, **252**, 848−851.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mietus-Snyder,M., Sladek,F.M., Ginsburg,G.S., Kuo,C.F., Darnell,J.E.,Jr and Karanthanasis,S.K. (1992) *Mol. Cell. Biol.*, **12**, 1708−1718.
Mlodzik,M., Hiromi,Y., Weber,U., Goodman,C.S. and Rubin,G.M. (1990) *Cell*, **60**, 211−224.
Mueller,C.R., Maire,P. and Schibler,U. (1990) *Cell*, **61**, 279−291.
Nauber,U., Pankratz,M.J., Kienlin,A., Seifert,E., Klemm,U. and Jäckle,H. (1988) *Nature*, **336**, 489−492.
Nishida,Y., Hata,M., Ayaki,H., Ryo,H., Yamagata,M., Shimizu,K. and Nishizuka,Y. (1988) *EMBO J.*, **7**, 775−781.
Nusslein-Volhard,C., Frohnhofer,H.G., Lehmann,R. (1987) *Science*, **238**, 1675−1681.
Oro,A.E., Ong,E.S., Margolis,J.S., Posakony,J.W., McKeown,M. and Evans,R.M. (1988) *Nature*, **336**, 493−496.
Oro,A.E., McKeown,M. and Evans,R.M. (1990) *Nature*, **347**, 298−301.
Pardue,M.L. (1986) In Roberts,D.B. (ed.), *Drosophila, A Practical Approach*. IRL Press, Oxford, pp. 111−137.
Pearson,W.R. and Lipman,D.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444−2448.
Pignoni,F., Baldarelli,R.M., Steingrimsson,E., Diaz,R.J., Patapoutian,A., Merriam,R. and Lengyel,J.A. (1990) *Cell*, **62**, 151−162.
Poulson,D.F. (1950) In Demerec,M. (ed.), *Biology of Drosophila*. Wiley, New York, pp. 168−174.
Rothe,M., Nauber,U. and Jäckle,H. (1989) *EMBO J.*, **8**, 3087−3094.
Ruppert,S., Boshart,M., Bosch,F.X., Schmid,W., Fournier,R.E.K. and Schutz,G. (1990) *Cell*, **61**, 895−904.
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.
Seagraves,W.A. and Hogness,D.S. (1990) *Genes Dev.*, **4**, 204−219.
Shea,M.J., King,D.L., Conboy,M.J., Mariani,B.D. and Kafatos,F.C. (1990) *Genes Dev.*, **4**, 1128−1140.
Skaer,H. (1989) *Nature*, **342**, 566−569.
Sladek,F.M. and Darnell,J.E. (1992) *Curr. Opin. Gene Dev.*, **2**, 256−259.
Sladek,F.M., Zhong,W., Lai,E. and Darnell,J.E.,Jr (1990) *Genes Dev.*, **4**, 2353−2365.
Strecker,T.R., Halsell,S.R., Fisher,W.W. and Lipshitz,H.D. (1989) *Science*, **243**, 1062−1065.
Tautz,D. and Pfeifle,C. (1989) *Chromosoma*, **98**, 81−85.
Weigel,D. and Jäckle,H. (1990) *Cell*, **63**, 455−456.
Weigel,D., Jurgens,G., Kuttner,F., Seifert,E. and Jäckle,H. (1989a) *Cell*, **57**, 645−658.
Weigel,D., Bellen,H.J., Jurgens,G. and Jäckle,H. (1989b) *Roux's Arch. Dev. Biol.*, **198**, 201−210.
Weigel,D., Jurgens,G., Klinger,M. and Jäckle,H. (1990) *Science*, **248**, 495−498.
Wustmann,G., Szidonya,J., Taubert,H. and Reuter,G. (1989) *Mol. Gen. Genet.*, **217**, 520−527.
Xanthopoulos,K.G., Mirkovitch,J., Decker,T., Kuo,C.F. and Darnell,J.E.,Jr (1989) *Proc. Natl. Acad. Sci, USA*, **86**, 4117−4121.

Received on June 12, 1992; revised on October 26, 1992