The EMBO Journal vol.10 no.8 pp.2259–2266, 1991

The *ovo* gene of *Drosophila* encodes a zinc finger protein required for female germ line development

Maryvonne Mével-Ninio[1,2], Régine Terracol[1] and Fotis C.Kafatos[2]

[1]Centre de Génétique Moléculaire du CNRS, 91190 Gif sur Yvette, France and [2]Department of Cellular and Developmental Biology, The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

Communicated by F.C.Kafatos

As defined by dominant and recessive *ovo* mutations, the *ovo* gene is required for development of the *Drosophila* female germ line, and does not exert any function in males or in somatic tissues. However, reversion of dominant *ovo* mutations can result in new phenotypes that are not related to the female germ line: the *svb* and *lzl* mutations affect cuticle and eye development, respectively. We have identified a 7.2 kb genomic fragment that rescues *ovo* mutations in transgenic *Drosophila* and thus contains all sequences necessary for ovo^+ function. This fragment has been sequenced almost in its entirety, defining the *ovo* locus at the molecular level. Multiple copies of the same fragment also rescue the *lzl* mutation. They do not rescue *svb* mutations, in agreement with genetic evidence that the *svb* function requires additional, more distal sequences. Nevertheless, a number of transposable element insertions that induce a *svb* phenotype interrupt the coding sequence of *ovo*. Taken together, the genetic and molecular data indicate the existence of a complex locus, where the *ovo* and *svb* functions depend on overlapping coding sequences but distinct regulatory elements. The data also suggest a model for the *lzl* phenotype. Expression of *ovo* at the RNA level is detectable at stage 8 of oogenesis in nurse cells and persists through the rest of oogenesis and in early embryogenesis. The *ovo* transcript encodes a protein of at least 1209 amino acids with four zinc fingers, suggesting that *ovo* might be a transcription factor required for female germ line maintenance and gametogenesis.

Key words: Drosophila/germ line/oogenesis/*ovo* locus/zinc fingers

Introduction

The *Drosophila* ovary is a model developmental system for the study of mechanisms involved in the maintenance of stem cells, the differentiation of cell types and the organization of positional information in the egg. During oogenesis, two or three oogonia at the distal end of the germaria divide and give rise to cystoblasts, which in turn undergo four incomplete cell divisions to produce a syncytium (or cyst) of 16 interconnected cystocytes. One of these cystocytes develops into the oocyte while the remaining 15 differentiate into nurse cells (King, 1970; Wieschaus and Szabad, 1979; Mahowald and Kambysellis, 1980). Hundreds of sex-specific genes are involved in these differentiation processes (King and Mohler, 1975; Gans *et al.*, 1975; Mohler, 1977). There is a class of female sterile mutations in which ovarian cystocytes undergo uncontrolled divisions with the result that ovaries become tumorous (Gateff, 1982). Among these, the mutations *otu* (King *et al.*, 1978) and *bag-of-marbles* (*bam*; McKearin and Spradling, 1990) affect early cystocyte divisions. By contrast, late cystocyte divisions are affected in some other 'tumor producing' mutations, including *fused* (*fu*; King, 1970), the alleles of *Sex-lethal*, $Sxl^{fs\#1}$ and $Sxl^{fs\#3}$ (Cline, 1978; Oliver *et al.*, 1990), *sans fille* (*snf*; Gans *et al.*, 1975; Oliver *et al.*, 1988; also named *liz*; Steinmann-Zwicky, 1988).

The *ovo* gene is of special interest because it is required exclusively in the female germ line. Homozygous null mutants result in female germ cell degeneration, beginning in late blastoderm or early gastrulation (Oliver *et al.*, 1987). Some mutant alleles of *ovo*, when homozygous, lead to the production of small ovarian tumors, as do some *trans*-heterozygous combinations of dominant and hypomorphic alleles. In these mutants, germ cells appear arrested in the differentiation process and closely resemble early spermatocytes (Oliver *et al.*, 1987, 1990). Recently, studies of several loci that produce 'tumorous' ovarian defects have suggested that these genes may interfere with the process of germ-line sex determination (Nöthiger *et al.*, 1989; Oliver *et al.*, 1990).

The *ovo* locus is also notable among genes involved in germ cell determination and differentiation, because of its very interesting dominant mutations. Three dominant alleles, ovo^{D1}, ovo^{D2} and ovo^{D3}, are known, all of which exhibit defective or arrested oogenesis (Busson *et al.*, 1983). Females heterozygous for ovo^{D1}, the strongest allele, fail to lay eggs. Irradiation of $ovo^{D1}/+$ first instar larvae can restore female fertility, by giving rise to homozygous wild-type germinal clones through mitotic recombination. Owing to this property, the ovo^{D1} mutation has been used extensively in investigating maternal effects of lethal mutations. Unfortunately, this has only been possible for genes which, like ovo^{D1} itself, are located on the X chromosome. Cloning of the locus would make it possible to extend germ line clonal analysis to autosomal mutations, by introducing copies of the ovo^{D1} allele in the autosomes through P-element mediated germ-line transformation.

Previously we reported the identification of genomic DNA sequences that belong to the *ovo* locus (Mével-Ninio *et al.*, 1989). We have now precisely located the gene, defining a chromosomal DNA fragment that rescues *ovo* mutations. As a first step towards understanding its function, we have examined expression of the locus in ovaries, and detected transcripts in nurse cells beginning at stage 8. These transcripts are apparently transferred to the oocyte when the nurse cells degenerate, and persist in the early embryo. We have also sequenced the locus, including a cDNA clone that

© Oxford University Press

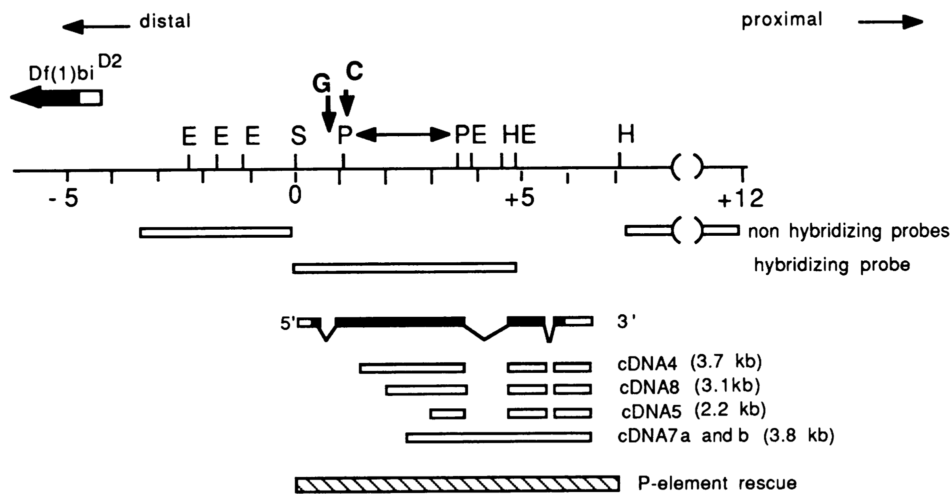

Fig. 1. Structure of the *ovo* gene. The upper restriction map and coordinates in kb were deduced from analysis of λ phage clones obtained in a chromosomal walk from an Oregon R library (Mével-Ninio *et al.*, 1989). Restriction sites are shown for *Eco*RI (E), *Hin*dIII (H), *Pst*I (P) and *Sal*I (S). Vertical arrows point to the sites of insertion of the *gypsy* (G) and *copia* (C) transposable elements in revertants of the ovo^{D1} and ovo^{D3} dominant mutations. The double-headed horizontal arrow indicates a *Pst*I fragment within which at least seven insertions associated with a *svb* phenotype have been mapped (Mével-Ninio *et al.*, 1989). The coordinates below the map are in kb from a *Sal*I site at position zero. The thick horizontal arrow gives the position of $Df(1)bi^{D2}$ which deletes sequences necessary for *svb* function but does not affect *ovo* wild-type function (Oliver *et al.*, 1987); the solid portion indicates the deleted material, and the open portion indicates the uncertainty of the breakpoint. Immediately below the map, three blocks indicate probes used to detect transcripts on RNA blots. The structure of the *ovo* transcript is shown below. ORF (solid blocks), second and third intron locations were identified by alignment of the sequence of cDNA4 with the genomic sequence. Open blocks in the 3′ and 5′ part of the gene show untranslated regions whose structure was inferred from genomic sequence analysis; this analysis also suggested the existence of the first intron. The fragment used for the P-element mediated transformation is shown at the bottom.

contains most of the protein coding region. The sequence predicts a zinc finger transcription factor which is apparently required in oogenesis.

Results

Identification of the ovo locus

The *ovo* locus is delimited distally by the breakpoint of deletion $Df(1)bi^{D2}$ which retains the wild-type function of *ovo* but does not complement the closely flanking mutation *shavenbaby* (*svb*; Wieschaus *et al.*, 1984; Figure 1). Proximally, the locus is defined less precisely by the non-complementing flanking mutation, *rugose*. We have cloned DNA from the locus by taking advantage of an insertion of the *gypsy* transposable element, which reverts the dominant mutation ovo^{D1} to a recessive loss-of-function mutation, ovo^{r}. A chromosomal walk was initiated, and 30 kb of DNA that extend proximally from the breakpoint of $Df(1)bi^{D2}$ were cloned (Mével-Ninio *et al.*, 1989). The *gypsy* insertion maps at 1.0 kb on this walk, counting from a reference *Sal*I site at position zero, and a *copia* insertion in an ovo^{D3} revertant maps at 1.2 kb (Figure 1). To localize the *ovo* gene we characterized transcripts within this walk. Poly(A)-containing RNA was isolated from ovaries, female carcasses and males and analyzed on Northern blots. No transcripts were detected with a 3.3 kb probe extending distally from the *Sal*I site (0), nor with probes proximal to a *Hin*dIII site (+7.2 to +12.0; see Figure 1). By contrast, two transcripts, 5 kb and 4.8 kb long, were detected (Figure 2) when we used as probe a 4.9 kb fragment extending between positions 0 and +4.9, i.e. in the region encompassing the integration sites of the *gypsy* and *copia* inserts that are responsible for reversions of ovo^{D1} and ovo^{D3}. The 5 kb transcript is much enriched in the ovaries, in comparison with males and female carcasses, as would be expected for the product of a gene required in the female germ line. The female carcasses and the males have comparable, lower levels of total transcripts (note the different loading evidenced by the rp49 standard in Figure 2). In these tissues, the 4.8 kb and 5 kb transcripts are roughly equal in amount.

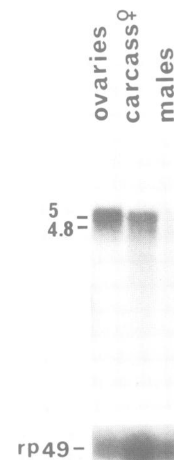

Fig. 2. Transcription of the *ovo* gene in adult flies. Poly(A)$^+$ RNAs (~10 μg) from ovaries, female carcasses and males were hybridized with a labeled 4.9 kb fragment, extending from position 0 to +4.9 (see Figure 1). The same filter was hybridized to a probe from the ribosomal protein 49 gene (O'Connell and Rosbash, 1984) as a control for loading.

For identifying the *ovo* gene unambiguously, we used P-element mediated germ line transformation to attempt a genetic rescue. A 7.2 kb genomic DNA fragment, from *Sal*I at position 0 to an artificial *Sal*I site in one of the phages of the walk (see Figure 1), was introduced into flies using the pW6 vector (Klemenz *et al.*, 1987). Four independent transformant lines were tested and found to complement fully

```
Sal I
[illegible] 150
[illegible] 300
[illegible] 450
[illegible] 600
[illegible] 750
[illegible] 900
[illegible] 1050
[illegible] 1200
 N F A A E L L Q R L T P N T A T T A Q N N I V N N L V N N S R A A T S V L A T K D C T 43

[illegible] 1350
I E N S P I S I P K N Q R A E D E E E Q E D Q E K E K P A E R E R E K S D E R T E Q V E K E E R V E 93

[illegible] 1500
R E E E E E D D E V D V G V E A P R P R F Y N T G V V L T Q A Q R K E Y P Q E P K D L S L T I A K S S 143
                                                                                     cDNA4 start
[illegible] 1650
P A S P H I H S D S E S D S D S D G G C K L I V D E K P P L P V I K P L S L R L R S T P P P A D Q R 193

[illegible] 1800
P S P P P P R D P A P A V R C S V I Q R A P Q S Q L P T S R A G F L L P P L D Q L G P E Q Q E P I D 243

[illegible] 1950
Y H V P K R R S P S Y D S D E E L N A R R L E R A R Q V R E A R R R S T I L A A R V L L A Q S Q R L 293

[illegible] 2100
N P R L V R S L P G I L A A A A G H G R N S S S S G A A G Q G F Q S S G F G S Q N S G S G S S S G 343
                                                      cDNA8 start
[illegible] 2250
N Q N A G S G A G S P G S G A G G G G G M G G G R D G R G N Y G P N S P P T G A L P P F Y E S L K S 393

[illegible] 2400
G Q Q S T A S N N T G Q S P G A N H S H F N A N P A N F L Q N A A A A A Y I M S A G S G G G G C T G 443

[illegible] 2550
N G G G G A S G P G G G P S A N S G G G G G G G G G G N G Y I N C G G V G G P N N S L D G N N L L N F 493
                                         cDNA7 start
[illegible] 2700
A S V S N Y N E S N S K F H N H H H H H Q H N N N N N N N G G Q T S M M G H P F Y G G N P S A Y G I 543

[illegible] 2850
I L K D E P D I E Y D E A K I D I G T F A Q N I I Q A T M G S S G Q F N S A A Y E D A I M S D L A S 593

[illegible] 3000
S G Q C P N G A V D P L Q F T A T L M L S S Q T D H L L E Q L S D A V D L S S F L Q R S C V D D E E 643
                                                cDNA5 start
[illegible] 3150
S T S P R Q D F E L V S T P S L T P D S V T P V E Q H N T N T T Q L D V L H E N L L T Q L T H N I V 693

[illegible] 3300
R G G S N Q Q Q Q H H Q Q H G V Q Q Q Q Q Q Q H S V Q Q Q Q Q H N V Q Q Q H G V Q Q Q H V Q Q Q P P 743

[illegible] 3450
P S Y Q H A T R G L M M Q Q Q P Q H G G Y Q Q Q A A I M S Q Q Q Q Q L L S Q Q Q Q Q S H H Q Q Q Q Q 793

[illegible] 3600
Q Q H A A A Y Q Q H N I Y A Q Q Q Q Q Q Q Q Q H H Q Q Q Q H H Q Q Q Q Q P Q P Q S H H S H H 843

[illegible] 3750
H G H G H D N S N M S L P S P T A A A A A A A A A A A A A A A A A A A A A H L Q R P M S S S S S S G 893

[illegible] 3900
G T N S S N S S G G S S N S P L L D A N A A A R A A A A L L D T K P L I Q S 931

[illegible] 4050
[illegible] 4200
[illegible] 4350
[illegible] 4500
[illegible] 4650
[illegible] 4800
L G L P P D L Q L E F V N G G H 947

[illegible] 4950
G I K N P L A V E N A H G G H H R I R N I D C I D D L S K H G H H S Q H Q Q Q Q G S P Q Q Q N M Q Q 997

[illegible] 5100
S V Q Q Q S V Q Q Q Q S L Q Q Q Q Q Q H H S N S S A S S N A S S H G S A E L A C M G S S G 1047

[illegible] 5250
A N E D S S S G N N K F V C R V C M K T F S L Q R L L N R H M K C H S D I K R Y L C T F C G K G F N 1097

[illegible] 5400
D T F D L K R H T R T H T G V R P Y K C N L C E K S F T Q R C S L E S H C Q K V H S V Q H Q Y A Y K 1147

[illegible] 5550
E R R A K                                                          M Y V C E E C G H T T C E 1165

[illegible] 5700
P E V H Y L H L K N N H P F S P A L L K F Y D K R H F K F T N S Q F A N N L L G Q L P M P V H N * 1213

[illegible] 5850
[illegible] 6000
[illegible] 6150
                                                                  cDNA4 end
[illegible] 6300
cDNA7 end
[illegible] 6450
[illegible] 6600
[illegible]
```

Fig. 3. The genomic DNA sequence and predicted amino acid sequence (below) of the *ovo* gene. Known intron sequences are indicated by lower-case letters. The termination codon is marked with an asterisk. The extent of the unspliced cDNA7 and the spliced cDNA4, as well as the starts of cDNA8 and cDNA5, are indicated. Note that the long ORF extends upstream of the cDNA4 start. The underlined AG at position 1070 is interpreted as the 3′ end of an intron (dashed vertical arrow); note that it is preceded by a polypyrimidine stretch and an in-frame UAA (underlined). The potential translation initiation codons in the 5′ region and polyadenylation signals in the 3′ region are underlined. The four zinc finger repeats are shown underlined, and the extent of the putative DNA binding domain is marked with solid vertical arrows.

the dominant sterility of $ovo^{D2}/+$ females. Among the three dominant ovo^{D} alleles, ovo^{D2} was chosen for this test because of its intermediate phenotype. The $ovo^{D2}/+$ females are fully sterile, while $ovo^{D2}/++$ are fertile; in contrast, some rare eggs of $ovo^{D3}/+$ develop normally, while $ovo^{D1}/+++$ females remain sterile. One of the

transformant lines was also tested and shown to complement the sterility of ovo^r/ovo^r females, homozygous for a null allele of *ovo*. This result proves that the 7.2 kb *Sal*I fragment contains all of the DNA required for the wild-type function of *ovo*.

We also attempted to rescue two mutations closely linked to *ovo*, *svb* and *lzl*, which appear frequently in revertants of the dominant mutations ovo^{D1} and ovo^{D3}. The *svb* mutation is observed in almost all revertants of ovo^{D1} obtained after γ-ray or EMS mutagenesis (Oliver *et al.*, 1987) and is also induced by insertion of *copia* or other unidentified transposable elements into a segment of DNA delimited by the *Pst*I sites at +1.1 and +3.6 (Mével-Ninio *et al.*, 1989). We found that the 7.2 kb *Sal*I DNA fragment does not rescue the *svb* mutation. This result is in agreement with genetic data which show that DNA sequences distal to the $Df(1)bi^{D2}$ breakpoint are necessary for the *svb* wild-type function (Oliver *et al.*, 1987).

The *lzl* mutation is only found among revertants of ovo^{D1}, and is caused by the insertion of a *gypsy* transposable element, in a defined orientation, into a site mapping at +1.0 (Figure 1; Mével-Ninio *et al.*, 1989). This mutation is also rescued by the 7.2 kb fragment, but in a dosage-sensitive manner: two copies of the fragment were necessary to restore an almost wild-type phenotype in hemizygous *lzl* males or heterozygous females (Figure 4). Intermediate mutant phenotypes were observed in heterozygous *lzl*/+ females containing only one copy of the 7.2 kb fragment. Possible interpretations of these results are considered in the Discussion.

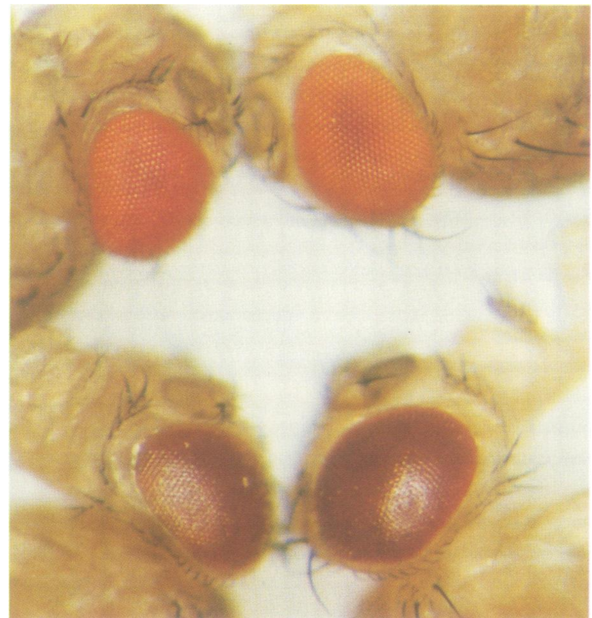

Fig. 4. Rescue of the *lzl* mutation in male flies. The fly on the bottom right shows the *lzl* eye mutation. The Oregon R fly on the top right has wild-type eyes. Flies on the left have respectively one copy (bottom) and two copies (top) of the pS-7.2 fragment in their genome.

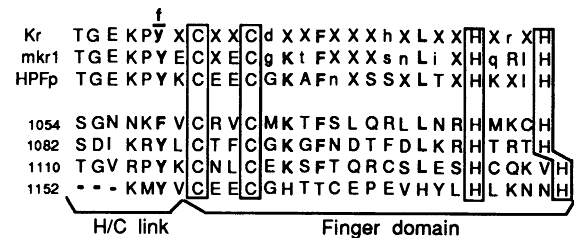

Fig. 5. Zinc fingers of *ovo*. Individual fingers are aligned below each other and compared to finger motif consensus sequences of the *Drosophila Kr* gene (Tautz *et al.*, 1987), the mouse *mkr1* gene (Chowdhury *et al.*, 1987) and human placental finger proteins (HPFp; Bellefroid *et al.*, 1989). The conserved cysteine and histidine residues are boxed. Other residues strongly conserved among finger proteins are indicated by bold type. Capital letters correspond to amino acids conserved over 75%, lower-case letters refer to amino acids conserved over 50%, and X indicates <50% conservation. Dots indicate sequence not shown.

cDNA and genomic sequencing

Five cDNAs were isolated from a directionally cloned 8–12 h embryonic library (Brown and Kafatos, 1988). Sequencing showed that four of the cDNAs have identical 3′ ends, while the fifth one, cDNA4, is 35 bp shorter. At their 5′ ends all cDNAs are different, except for cDNA7a and cDNA7b. The latter two clones are identical, according to their size (3.8 kb) and restriction patterns, which are the same as in genomic DNA. The remaining cDNAs, 4, 5 and 8, do not contain the *Eco*RI and the *Hin*dIII sites located at +3.9 and +4.5 respectively, and represent spliced transcripts.

We have sequenced on the antisense strand 6.65 kb of the genomic DNA fragment which rescues the sterility of ovo^r/ovo^r females. Sequencing was performed from the *Sal*I (0) site to beyond the 3′ end of the cDNAs. The two longest molecules, cDNA7 (unspliced) and cDNA4 (spliced), were also sequenced, on the sense strand. A comparison of genomic, cDNA7 and cDNA4 sequences indicated that the *ovo* gene has at least three exons and two introns (Figures 1 and 3). The splice donor and acceptor sites fit the invertebrate splice junction consensus (Shapiro and Senapathy, 1987). The genomic sequence predicts the existence of at least one additional intron: upstream of cDNA4 the ORF continues in-frame for 185 codons, until an AG dinucleotide is encountered at position 1070, preceded by an extensive pyrimidine tract (typical of intronic 3′ ends; Shapiro and Senapathy, 1987), and by an in-frame UAA terminator further upstream (Figure 3). The first in-frame AUG downstream of the AG dinucleotide is located 1090 nucleotides into the ORF, and therefore is probably not the translation initiator. Although we do not know the 5′ end of the putative intron, we note that one of the AUG triplets at positions 463, 782 and 795 could represent the start of translation. The first two of these triplets are the best candidates, since they are embedded in sequences that match the consensus for *Drosophila* translation initiation sites in four out of seven nucleotides (Cavener, 1987).

The protein coding portion of cDNA4 ends with a UAG termination codon at nucleotide 5695. No poly(A) tail is present at the 3′ ends of either cDNA4 or cDNA7, but both terminate in a domain of genomic DNA very rich in A repeats. These A repeats are likely to have served as primers in the synthesis of the cDNAs. Further downstream, two closely spaced polyadenylation signal sequences, AAUAAA, exist in the genomic DNA, starting at positions 6451 and 6475 respectively (Figure 3). From the AG at the end of the putative first intron, to 20 nucleotides downstream of the second AAUAAA, the mature mRNA would be 4.45 kb. With due allowance for a poly(A) tail and a putative first exon of unknown length, this is consistent with the transcript

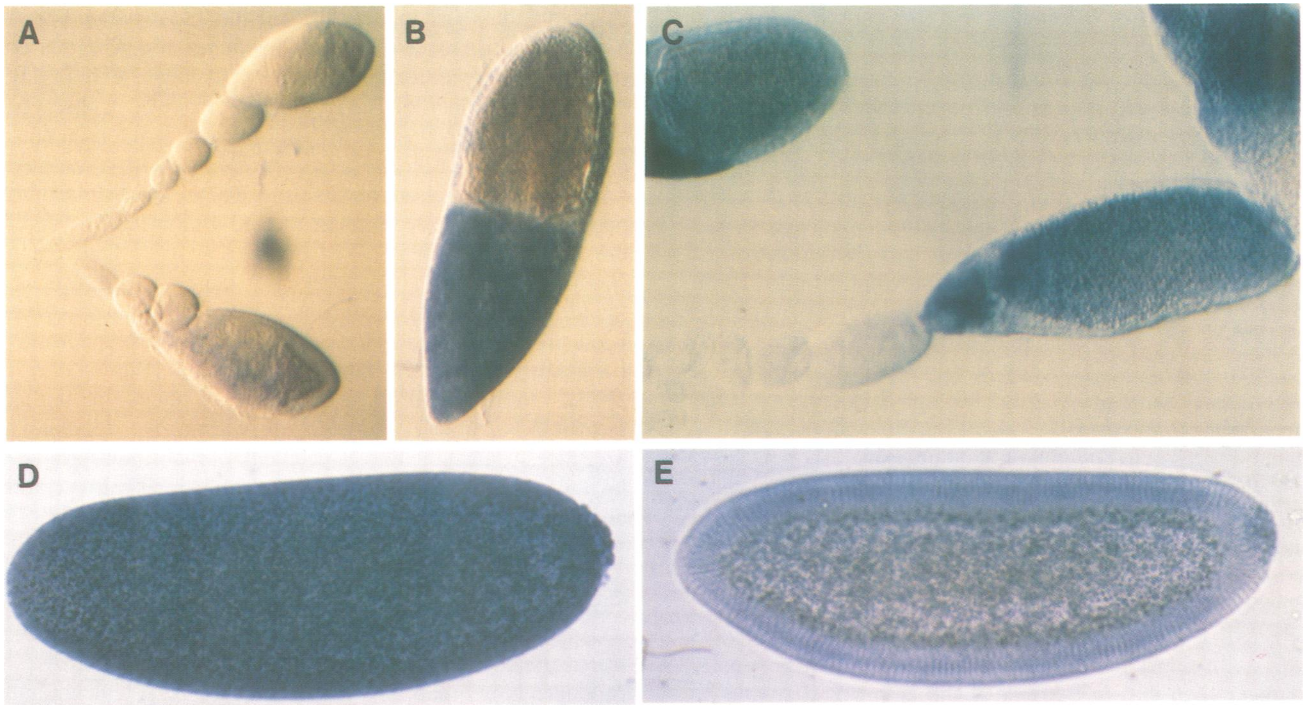

Fig. 6. Expression of *ovo*$^+$ in ovaries and early embryos. Embryos are shown in lateral view, posterior ends to the right. Egg chamber stages are defined according to King (1970). (**A**) The *ovo* transcript is clearly visible at stage 9; however, using a higher magnification, we could also detect very faint expression of *ovo*, visible as blue grains, at stage 8. (**B**) At stage 10, expression is abundant in the nurse cells, mostly in the cytoplasm. Weak staining is also visible in the follicular epithelial cells. (**C**) Transcript accumulates in the oocyte at stage 12. No staining is visible in the follicular epithelial cells. (**D**) At the syncytial blastoderm stage transcript appears uniformly distributed in the embryo. (**E**) At the cellular blastoderm stage, 2.5 h, the level of *ovo* transcript drops and appears higher in the pole cells. At early gastrulation stage, staining of pole cells disappears (data not shown). Three different cDNA fragments were digoxigenin labeled and used as probes: (1) a 2.2 kb fragment at the 5′ extremity of cDNA4; (2) a 450 bp fragment devoid of homopolymer repeats; (3) a 1.4 kb fragment at the 3′ extremity of cDNA4 including the zinc finger coding region. The pictures shown were obtained with the 2.2 kb probe, but the same staining pattern was obtained whatever probe we used.

sizes (5 kb and 4.8 kb) detected on Northern blots. The cDNA7 sequence was found to be identical to the genomic sequence. Unspliced cDNAs, like cDNA7, are not uncommon in this library (Brown *et al.*, 1989). We presume that cDNA7 resulted from priming of pre-mRNA in one of the oligo(A) domains near the end of the *ovo* transcript.

The ovo protein has zinc fingers

Analysis of the derived amino acid sequence of the ovo protein predicts four zinc finger repeats (underlined in Figure 3), near the carboxy terminus. These zinc fingers are similar to those found in the *Drosophila Krüppel* (*Kr*) gene product (Rosenberg *et al.*, 1986) and in the putative chorion regulator, CF2 (Shea *et al.*, 1990). A search in the NBRF protein database revealed no strong similarity with the ovo protein, except for the zinc finger domain, which is also similar to the corresponding domain of the mouse protein mkr1 (Chowdhury *et al.*, 1987), as well as that of a family of proteins found in human placenta (Bellefroid *et al.*, 1989). Figure 5 shows an alignment of the fingers of the ovo protein with consensus sequences of the fingers of the latter two proteins and with the Kr consensus. The fingers are tandem repeats of the Kr consensus sequence $X_7CX_2CX_3FX_5LX_2HX_3H$, where X is any amino acid. The position of the highly conserved cysteine and histidine residues is strictly maintained, except that the third and fourth repeats contain four amino acids between the two histidines instead of the three commonly found. In Kr and Kr-related proteins, a stretch of seven amino acids, called the H/C link, separates the fingers and conforms to a TGEKPYX consensus sequence. In ovo, fingers two and three partially match that consensus. A short intron of 95 bp separates the coding sequence of the first three fingers from that of the fourth. This last finger deviates from the rest in that it contains a threonine and a histidine residue instead of the normally conserved phenylalanine and leucine residues, respectively, of the Kr consensus. The link separating this last finger from the other three is 17 rather than seven amino acids long.

The *ovo* gene encodes a protein of at least 1209 amino acid residues, with 130 880 molecular weight, an isoelectric point of 7.0, and numerous charged residues (22%) very typical of a *Drosophila* protein (21.8% on average). A notable feature of this protein is its extremely hydrophilic profile. One highly charged region occurs between residues 58 and 103. This region contains predominantly acidic residues, that is glutamic acid (46%) and aspartic acid (13%). Another feature of the ovo protein is the presence of numerous homopolymeric runs of amino acids, particularly polyglutamines but also polyalanines, polyglycines and a polyasparagine and polyhistidine run. The significance of such repetitiveness is unknown, although polyglutamine runs are frequently found in eukaryotic transcription factors (Mitchell and Tjian, 1989). Recently, another homopolymer-rich structure has been described for the product of the *Drosophila mastermind* locus (Smoller *et al.*, 1990).

ovo is expressed in nurse cells

The degree of sterility of females carrying one of the dominant *ovo*^{D} mutations, *ovo*^{D1}, *ovo*^{D2} or *ovo*^{D3}, decreases

from ovo^{D1} to ovo^{D3}. Females heterozygous for ovo^{D1} do not lay any eggs; egg chambers form, but degenerate before vitellogenic stages. By contrast, $ovo^{D3}/+$ females lay eggs, although these are flaccid and degenerate. Similarly, the recessive null mutation, ovo^{D1rS1}, results in a lack of egg chamber formation (Oliver *et al.*, 1987). These data suggest that *ovo* encodes a function necessary for normal development of the oocyte. To test this prediction, we characterized expression of *ovo* transcripts by *in situ* hybridization of digoxigenin-labeled DNA probes (Tautz and Pfeifle, 1989) to wild-type ovaries and early embryos (Figure 6). The *ovo* transcripts are not detectable in the ovarian stem cells, oogonia or early cysts. They become visible at stage 8 of oogenesis (Figure 6). At stage 10 high levels of *ovo* RNA are present in nurse cells, mostly in the cytoplasm. Weak staining is also visible in the follicular epithelial cells. Some *ovo* RNA is retained in the nuclei of degenerating nurse cells, but the majority is apparently transferred to the oocyte and persists in the early embryo, where it is very abundant and homogeneously distributed. During cellularization of the blastoderm, the *ovo* RNA level decreases strongly. It is still detected throughout the cortical region of the embryo, as well as in the pole cells, the staining of which is slightly stronger than that of the rest of the embryo. It should be noted that the staining of the pole cells is cytoplasmic and is present in every embryo of the blastoderm stage. Thus, no staining difference was detected in the pole cells according to the sex of the embryos.

Discussion

The ovo locus is complex

We have shown that the *ovo* locus is included within a 7.2 kb *Sal*I fragment (pS-7.2) which is able to restore completely normal female germ line development and fertility both in females heterozygous for the dominant ovo^{D2} mutation and in ovo^r/ovo^r females, homozygous for a loss of function allele. We have not yet attempted rescue of ovo^{D1}, the most dominant ovo^D allele; since it has been observed that $ovo^{D1}/+++$ females remain sterile, we predict that at least three doses of the pS-7.2 insert will be necessary to restore fertility. Within the pS-7.2 fragment, the downstream end of the *ovo* gene is now reasonably well delimited by the 3′ extremities of the cDNAs and the nearby AAUAAA motifs. Further experiments are needed to determine the beginning of the *ovo* transcription unit as well as the site of initiation of translation. Nevertheless, we now know that these sites are contained within pS-7.2 and should be within a 1.1 kb region bounded by the *Sal*I extremity of this fragment and the beginning of the coding sequence shown in Figure 3. As discussed in Results, the sequence of this 1.1 kb region predicts that it contains at least one intron. This region most likely also includes regulatory sequences necessary for *ovo* expresssion.

Within the pS-7.2 fragment, the ovo^D mutations have not yet been molecularly characterized. Since their genomic restriction patterns are normal (Mével-Ninio *et al.*, 1989) and they were obtained by EMS mutagenesis, they might represent point mutations, and sequencing will be necessary to determine which part of the molecule is altered.

The pS-7.2 fragment rescues the *lzl* mutation in addition to *ovo*. This was a quite unexpected result. Genetic analysis had shown that *ovo* and *lzl* are very close, but it was not suspected that these loci overlapped and that *ovo* could exert a function in eye development. Females homozygous or males hemizygous for null alleles of *ovo* never exhibit any mutant eye phenotype. The *lzl* mutations have been observed exclusively in revertants of the dominant ovo^{D1} mutation, caused by insertion of a *gypsy* element in a defined orientation. The present results show that the *gypsy* insertions take place in a DNA segment located within the 5′ part of the *ovo* gene, between *Sal*I (0) and *Pst*I (+1.1). This region should contain both regulatory and transcribed sequences, including one or more introns. The ovo^{D1} mutation has been proposed to code for an antimorphic product which competes with the wild-type product of *ovo* (Busson *et al.*, 1983). One possibility to explain the *lzl* phenotype is that insertions of the *gypsy* element within *ovo* regulatory sequences change the tissue specificity of the transcripts. The level of the ovo^{D1} transcripts in the female germ line could decrease, resulting in reversion of the dominant female sterile mutation, while at the same time ovo^{D1} RNA could be expressed in the eyes, thus accounting for the semi-dominant characteristic of the *lzl* phenotype. Several copies of the wild-type pS-7.2 fragment would be necessary to neutralize, by an unknown mechanism, the toxic effect of the ovo^{D1} product. Now that the *ovo* locus has been cloned and characterized, it is possible to test these hypotheses by examining the sequence of the ovo^{D1} allele and the tissue distribution of transcripts in the wild-type, ovo^{D1} and its *lzl* revertants.

The *svb* mutation was isolated in a search for mutants affecting the pattern of the larval cuticle (Wieschaus *et al.*, 1984). It has an effect on the number and size of denticle belt setae in embryos and larvae and is partially lethal. The few *svb* males that survive show a phenotype characterized by lightly curved bristles and unspread wings. Our attempt to rescue the *svb* mutation with the pS-7.2 fragment was unsuccessful. This result is not surprising, since it has been shown that sequences within the $Df(1)bi^{D2}$ deletion (located distally to pS-7.2; Figure 1) are necessary for the *svb* wild-type function, while they are not necessary for the wild-type function of *ovo* (Oliver *et al.*, 1987). We had previously mapped at least seven insertions of transposable elements associated with the *svb* mutant phenotype within a 2.5 kb segment, delimited by the two *Pst*I sites at +1092 and +3655 (Figure 3; see Mével-Ninio *et al.*, 1989). The present analysis indicates that this fragment encompasses exclusively coding sequences of *ovo*. Thus, *ovo* and *svb* are overlapping loci which apparently share coding sequences but differ in at least some regulatory elements. The complex *ovo*−*svb* locus would be implicated in two distinct functions: an *ovo* function required for female germ line development, and a somatic function responsible for the *svb* embryonic, larval and adult phenotypes.

Of the two adult transcripts encompassed by the pS-7.2 fragment, the 5 kb transcript is differentially enriched in the ovary and is likely to serve the *ovo* germ line function. The 4.8 kb transcript, which is enriched relative to the 5 kb species in males and female carcasses, might serve the *svb* function, since *svb* mutants exhibit an adult phenotype visible in wings and bristles. The two transcripts might result from the use of two different initiation sites, followed by alternative splicing, giving rise to two overlapping final products. Again, these hypotheses could be tested by further characterization of the locus, especially localization of the

5′ exons which may distinguish the two transcripts, and use of exon-specific probes for examining the tissue distribution of the two transcripts in more detail.

The functions of ovo

Previous work had indicated that zygotic *ovo* function is required at multiple times during development. The ovo^- female germ line begins to degenerate in late blastoderm/early gastrula (Oliver *et al.*, 1987). Thereafter, the ovo^+ function may be constitutive in the female germ line, since elimination of the ovo^D alleles in $ovo^D/+$ individuals by mitotic recombination, at any time from first larval instar to adult emergence, can rescue the mutant phenotype (Perrimon, 1984). The gene is also required for gametogenesis in adult females. While both the dominant antimorphic alleles and the loss-of-function alleles can lead to arrested or defective oogenesis, the phenotypes vary according to the combination of alleles used (Oliver *et al.*, 1990). In female flies *trans*-heterozygous for ovo^{D2} or ovo^{D3} on the one hand, and partial loss of function alleles (such as ovo^{rM1}, ovo^{rM2} or ovo^{D1rG23}) on the other, abnormal cysts develop which contain an excess number of germ-line cells that did not continue to differentiate and formed tumorous egg chambers. Early disruption of the oogenesis differentiation process in such mutants could result in uncontrolled cell divisions and explain the tumorous egg chamber phenotype observed. In view of these results, it seems paradoxical that no *ovo* transcripts were detected in early oogenic stages. However, we have no information about the perdurance of ovo protein produced during even earlier developmental stages. Antibodies specific for the ovo protein can be generated now that the locus has been cloned and sequenced, and they should prove valuable for understanding the developmental dynamics and mechanism of action of this intriguing gene.

A notable feature of the expression of the locus is the production of large amounts of transcripts in the nurse cells of late vitellogenic stages. These transcripts are apparently transferred to the oocyte and persist in the early embryo, suggesting a maternal function for *ovo*. Due to the nature of the *ovo* mutation, maternal effects could not be studied with either recessive or dominant alleles. However, a dominant maternal effect has been observed in a partial revertant of the ovo^{D3} allele, which shows low fertility. Progeny of females heterozygous for this allele exhibit segmentation defects in the second abdominal segment (M.Gans, unpublished data). It will be interesting to test whether this maternal function is included in the pS-7.2 fragment.

The high levels of *ovo* RNA observed in early embryos decline rapidly, but persist in pole cells of either sex up to the stage of early gastrulation. According to the observations of Oliver *et al.* (1987), the *ovo* product is required zygotically at late blastoderm and early gastrula stages for female pole cell maintenance. The early action of *ovo* on female pole cell maintenance could be mediated by a sex-specific signal such as the X/A ratio.

The most striking feature of the ovo protein is the presence of a tandem array of four zinc finger repeats. In *Drosophila melanogaster*, zinc finger proteins are among the products of genes involved in developmental control, such as the 'gap' genes *Kr* and *hunchback* (Rosenberg *et al.*, 1986; Tautz *et al.*, 1987), the *snail* and *glass* genes which are involved in dorso-ventral development and photoreceptor cell development, respectively (Boulay *et al.*, 1987; Moses *et al.*, 1989), and the gene which encodes the putative chorion regulatory factor, CF2 (Shea *et al.*, 1990).

Two zinc finger proteins, TFIIIA and SpI (Kadonaga *et al.*, 1987), are known to be activators of transcription *in vitro*. Beside the presence of zinc fingers, the ovo protein shares features in common with the SpI DNA binding protein and other transcription factors. These include glutamine-rich regions which, in the case of SpI, have been reported to be transcriptional activation domains. In SpI, the glutamine-rich regions are in close proximity to serine-threonine-rich segments, as is also the case at the carboxy-terminal end of the ovo sequence (residues 981 − 1054). Furthermore, the ovo protein shows near the N-terminus a prominent 'acid patch' of the type frequently found in transcription factors (Ptashne, 1988). In summary, the structure of this complex locus suggests that it encodes a DNA binding protein with transcriptional regulatory activity, and that differentially regulated production of this protein (or possible variants) in the female germ line and in somatic embryonic and post-embryonic cells accounts for the diverse phenotypes associated with mutations in this locus.

Materials and methods

Drosophila stocks

Flies were maintained on the standard Gif medium (Gans *et al.*, 1975) at 23°C. The strains ovo^{D2} and ovo^{D1rS37} or $ovo^{D1rS124}$ were used for testing the rescue of the female sterility and of the *lzl* phenotype, respectively. We had previously introduced the *white* (*w*) mutation in the two latter strains by recombination. These strains as well as the dominant ovo^{D1} mutation have been described previously (Mével-Ninio *et al.*, 1989). Rescue of *svb* was tested on the strain svb^{YD39}, which is marked with *w* and balanced with *FM7* (Wieschaus *et al.*, 1984).

Nucleic acids

DNA fragment isolation, ligation and subcloning of DNA into plasmid vectors were carried out as described in Maniatis *et al.* (1982). $Poly(A)^+$ RNA was purified from adult males, female carcasses or hand dissected ovaries from *D.melanogaster* (Oregon R strain), and electrophoresed on 1.4% agarose gels according to Mariani *et al.* (1988). Gel purified DNA probes were labeled with the Klenow fragment of *Escherichia coli* DNA polymerase I. Blotting and hybridization conditions were as described by Mariani *et al.* (1988).

Isolation of cDNA clones

The 8 − 12 h cDNA library of Brown and Kafatos (1988) was screened by a high-density filter screening protocol (Brown and Kafatos, 1988), with a 4.9 kb (0; +4.9) genomic DNA fragment. From 60 000 colonies tested, five clones were isolated and analyzed. The longest ones, cDNA4 and cDNA7, were chosen for sequencing.

Germ line transformation

A 7.2 kb *Sal*I fragment (see Figure 1) was cloned into the pW6 vector (Klemenz *et al.*, 1987), which utilizes wild-type *w* expression as a transformation marker, and was injected into the w^{1118}; *P(rosy$^+$Δ2-3) Sb/Tm6 Ubx* host line (Robertson *et al.*, 1988) under standard conditions (Rubin and Spradling, 1982; Spradling and Rubin, 1982). F1 progeny expressing eye color were mated to balancer stocks to establish lines. Insertions from independent transformed lines were transferred by appropriate crosses into the strains ovo^{D2}, *w* ovo^{D1rS37}, *w* $ovo^{D1rS124}$ or *w* svb^{YD39}. One copy of the insertion restored full fertility into $ovo^{D2}/+$ and ovo^r/ovo^r females.

DNA sequencing and sequence analysis

Genomic and cDNA fragments were subcloned into the pBluescript KS+ vector (Stratagene). Nested deletions were obtained using exonuclease III (Pharmacia system) and sequenced using the dideoxy-chain termination method (Sanger *et al.*, 1977) as modified for use with the Pharmacia

sequencing kit. The sense strands of cDNAs and the corresponding antisense strands of genomic DNAs were sequenced. Genomic sequences extending 5′ and 3′ of the cDNA extremities, and intron sequences, were determined from the sense strand of genomic subclones, using nested deletions. Six sites showed differences between the genomic and cDNA sequences. These included two silent substitutions (A at base 5601 and T at base 5888 in the genomic DNA were G and C, respectively, in cDNA4 and cDNA7), and three amino acid changes (histidine 980 and leucine 1184 in genomic DNA and cDNA4 were replaced by a glutamine and a proline respectively in cDNA7; tyrosine 1170 in genomic DNA and cDNA7 was replaced by an asparagine in cDNA4). These last three changes might represent error of the reverse transcriptase, since they were only found in one of the cDNA, the other being identical to the genomic DNA. The last site showed variation in the number of alanine residues within a homopolymer run: region 860–880 encompassed a four residue deletion within cDNA4 and cDNA7. This change and the silent substitutions probably result from polymorphism, as the cDNA and genomic templates were derived from different strains: cDNA, *dp cn bw* (Brown and Kafatos, 1988); genomic, Oregon R. Polymorphism could also explain the three amino acid replacements, if the *dp cn bw* strain is not isogenic. DNA sequences were compiled using the UWGCG: University of Wisconsin Genetic Corporation Group software (Devereux *et al.*, 1984). The deduced amino acid sequence was compared with the NBRF database using the program FASTA (Pearson and Lipman, 1984).

Whole-mount in situ hybridizations

Whole-mount *in situ* hybridizations were performed according to Tautz and Pfeifle (1989), with some modifications provided by U.Waldorff (Biozentrum, Basel). Embryos were fixed for 25 min at room temperature in a two phase mixture consisting of heptane and 10% formaldehyde, phosphate-buffered saline (PBS), 50 mM EGTA. After devitellinization in methanol, embryos were stocked at −70°C in 100% ethanol. They were fixed a second time, before use, in 5% formaldehyde, PBT (PBS plus 0.1% Tween 20), for 20 min at room temperature. Hybridization and washes were performed at 50°C. Ovaries were dissected in PBS and fixed twice: first, in 4% paraformaldehyde, PBS, on ice; second, in the same medium plus 0.5% Triton X-100, at room temperature. Hybridizations were at 45°C and washes at 48°C. The cDNA4 restriction fragment probes were prepared using the digoxigenin DNA labeling kit from Boehringer Mannheim, following the manufacturer's instructions.

Acknowledgements

We thank B.Bouchon-Limbourg, and the rotating students L.Cohen and P.Lapie for production of transformed lines. We thank N.Brown for supplying cDNA libraries and for many helpful suggestions and discussions, E.Wieschaus for sending us the svb^{YD39} strain, and F.Michel and N. Randsholt for help with the computer work and for critical reading of the manuscript. For their encouragement and support throughout this work, we thank M.Gans and all other colleagues at Gif. Technical assistance by H.LeFur, curator of the stocks in Gif, is greatly appreciated. This work was supported by the Centre National de la Recherche Scientifique, and by grants from La Ligue Nationale Française contre le cancer (M.M.-N. and R.T.) and from the National Science Foundation (F.C.K. and M.M.-N.).

References

Bellefroid,E.J., Lecocq,P.J., Benhida,A., Poncelet,D.A., Belayew,A. and Martial,J.A. (1989) *DNA*, **8**, 377–387.
Bouley,J.L., Dennefeld,C. and Alberga,A. (1987) *Nature*, **330**, 395–398.
Brown,N.H. and Kafatos,F.C. (1988) *J. Mol. Biol.*, **203**, 425–437.
Brown,N.H., King,D.L., Wilcox,M. and Kafatos,F.C. (1989) *Cell*, **59**, 185–195.
Busson,D., Gans,M., Komitopoulou,K. and Masson,M. (1983) *Genetics*, **105**, 309–325.
Cavener,D.R. (1987) *Nucleic Acids Res.*, **15**, 1353–1361.
Chowdhury,K., Deutsch,U. and Gruss,P. (1987) *Cell*, **48**, 771–778.
Cline,T.W. (1978) *Genetics*, **90**, 683–698.
Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
Gans,M., Audit,C. and Masson,M. (1975) *Genetics*, **81**, 683–704.
Gateff,E. (1982) *Adv. Cancer Res.*, **37**, 33–74.
Kadonaga,J.T., Carner,C.R., Masiarz,F.R. and Tjian,R. (1987) *Cell*, **51**, 1079–1090.
King,R.C. (1970) In King,R.C. (ed.), *Ovarian Development in Drosophila melanogaster*. Academic Press, New York, pp. 33–98.
King,R.C. and Mohler,J.D. (1975) In King,R.C. (ed.), *Handbook of Genetics*. Plenum Press, New York, Vol. 3, pp. 757–791.
King,R.C., Bahns,M., Horowitz,R. and Larramendi,P. (1978) *Int. J. Insect Morphol. Embryol.*, **7**, 359–378.
Klemenz,R., Weber,U. and Gehring,W. (1987) *Nucleic Acids Res.*, **15**, 3947–3959.
Mahowald,A.P. and Kambysellis,M.P. (1980) In Ashburner,M. and Wright,T.R.F. (eds), *The Genetics and Biology of Drosophila*. Academic Press, New York, Vol. 2d, pp. 141–224.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mariani,B.D., Lingappa,J.R. and Kafatos,F.C. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 3029–3033.
McKearin,D.M. and Spradling,A.C. (1990) *Genes Dev.*, **4**, 2242–2251.
Mével-Ninio,M., Mariol,M.C. and Gans,M. (1989) *EMBO J.*, **8**, 1549–1558.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Mohler,J.D. (1977) *Genetics*, **85**, 259–272.
Moses,K., Ellis,M.C. and Rubin,G.M. (1989) *Nature*, **340**, 531–536.
Nöthiger,R., Jonglez,M., Leuthold,M., Meier-Gerschwiler,P. and Weber,T. (1989) *Development*, **107**, 505–518.
O'Connell,P. and Rosbash,M. (1984) *Nucleic Acids Res.*, **12**, 5495–5514.
Oliver,B., Perrimon,N. and Mahowald,A.P. (1987) *Genes Dev.*, **1**, 913–923.
Oliver,B., Perrimon,N. and Mahowald,A.P. (1988) *Genetics*, **120**, 159–171.
Oliver,B., Pauli,D. and Mahowald,A.P. (1990) *Genetics*, **125**, 535–550.
Pearson,W.R. and Lipman,D.J. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
Perrimon,N. (1984) *Genetics*, **108**, 927–939.
Ptashne,M. (1988) *Nature*, **335**, 683–689.
Robertson,H.M., Preston,C.R., Phillis,R.W., Johnson-Schlitz,D.M., Benz,W.K. and Engels,W.R. (1988) *Genetics*, **118**, 461–470.
Rosenberg,U.B., Schröder,C., Preiss,A., Kienlin,A., Côté,S., Riede,I. and Jäckle,H. (1986) *Nature*, **319**, 336–339.
Rubin,G. and Spradling,A. (1982) *Science*, **218**, 348–353.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
Shapiro,M.B. and Senapathy,P. (1987) *Nucleic Acids Res.*, **15**, 7155–7174.
Shea,M.J., King,D.L., Conboy,M.J., Mariani,B.D. and Kafatos,F.C. (1990) *Genes Dev.*, **4**, 1128–1140.
Smoller,D., Friedel,C., Schmid,A., Bettler,D., Lam,L. and Yedvobnick,B. (1990) *Genes Dev.*, **4**, 1688–1700.
Spradling,A. and Rubin,G. (1982) *Science*, **218**, 341–347.
Steinmann-Zwicky,M. (1988) *EMBO J.*, **7**, 3889–3898.
Tautz,D. and Pfeifle,C. (1989) *Chromosoma*, **98**, 81–85.
Tautz,D., Lehmann,R., Schürch,H., Schuh,R., Seifert,E., Kienlin,A., Jones,K. and Jäckle,H. (1987) *Nature*, **327**, 383–389;
Wieschaus,E. and Szabad,J. (1979) *Dev. Biol.*, **68**, 29–46.
Wieschaus,E., Nüsslein-Wolhard,C. and Jürgens,G. (1984) *Wilhelm Roux's Arch. Dev. Biol.*, **193**, 296–307.

Received on April 4, 1991; revised on May 2, 1991