Development 115, 75-88 (1992)

The *Drosophila orb* gene is predicted to encode sex-specific germline RNA-binding proteins and has localized transcripts in ovaries and early embryos

VALERIE LANTZ[1], LINDA AMBROSIO[2] and PAUL SCHEDL[1]

[1]Department of Molecular Biology, Moffett Laboratory, Princeton University, Princeton, NJ 08544, USA
[2]Department of Zoology and Genetics, 339 Science 2, Iowa State University, Ames, IA 50010, USA

Summary

We report the identification of a new gene, *orb*, which appears to be expressed only in the germline and encodes ovarian- and testis-specific transcripts. The predicted proteins contain two regions with similarity to the RRM family of RNA-binding proteins but differ at their amino termini. In testes, *orb* RNA accumulates in the primary spermatocytes and at the caudal ends of the spermatid bundles. In ovaries, *orb* transcripts display an unusual spatial pattern of accumulation in the oocyte. Preferential accumulation in the oocyte of *orb* RNA is first detected in region 2 of the germarium and is dependent on *Bicaudal-D* and *egalitarian*. While in stage 7 egg chambers *orb* RNA is localized posteriorly in the oocyte, during stages 8-10 it is localized at the anterior of the oocyte, asymmetrically along the dorsal-ventral axis. In embryos the transcripts accumulate at the posterior end and are included in the pole cells. This pattern of localization and the similarity to RNA-binding proteins suggest that the *orb* gene product may mediate the localization of maternal RNAs during oogenesis and early embryogenesis.

Key words: *orb*, RNA localization, RNA-binding protein, germline-specific, *Drosophila* oogenesis, spermatogenesis.

Introduction

The proper spatial distribution of macromolecules in the cell is critical to many biological processes. Proteins, in particular, must be correctly targeted to the subcellular compartment where their activity is needed (Hartl and Neupert, 1990; Landry and Gierasch, 1991; Silver, 1991). Although the precise mechanism for protein localization differs for each subcellular compartment, studies on a variety of systems have suggested that a similar sorting strategy is often employed. Signal sequences within the protein are used to specify its subcellular address. These sorting signals are then recognized by factors that mediate the localization of the protein to the correct compartment.

This is not, however, the only strategy used for protein sorting. Another mechanism for directing proteins to the appropriate subcellular location is mRNA targeting. For example, in cultured fibroblasts, actin, vimentin and tubulin mRNAs are highly concentrated in regions of the cell where their protein products are required (Lawrence and Singer, 1986). Similarly, in the developing rat brain the mRNA encoding MAP2, a dendrite-specific microtubule-associated protein, accumulates in neuronal dendrites but not in the cell bodies (Davis et al., 1987; Garner et al., 1988).

mRNA targeting may be particularly important as a mechanism for generating the correct spatial distribution of positional information in oocytes and early embryos. Several localized mRNAs have been identified in *Xenopus laevis* oocytes (Rebagliati et al., 1985). For example, Vg1 RNA is restricted to the vegetal cortex of *Xenopus* oocytes (Melton, 1987; Weeks and Melton, 1987). The asymmetric distribution of Vg1 mRNA and protein and homology to TGF-β has led to the suggestion that the Vg1 product may be involved in mesodermal induction in the developing *Xenopus* embryo (Dale et al., 1989; Tannahill and Melton, 1989). In *Drosophila,* the importance of RNA localization is most clearly demonstrated by the maternal effect locus *bicoid* which is required for the formation of head and thoracic structures (reviewed by St. Johnston and Nüsslein-Volhard, 1992). *bicoid* protein is distributed in an anterior-to-posterior concentration gradient which determines cell fate in the anterior half of the embryo. This protein gradient is dependent upon *bicoid* mRNA localized to the anterior pole during oogenesis. That this localized *bicoid* mRNA is essential for establishing the anterior-to-posterior protein gradient is demonstrated by mutations in *swallow* and *exuperantia* which perturb the distribution of *bicoid* mRNA and cause defects along the anterior-posterior axis. Formation of posterior structures in the early *Drosophila* embryo also involves the localization of maternal mRNAs encoded by genes such as *oskar* and *nanos* (reviewed by St. Johnston and Nüsslein-Volhard, 1992).

In this paper we describe the isolation and characterization of the *orb* (oo18 RNA-binding) gene which encodes female- and male-specific protein products containing two putative RNA-binding domains. Expression of *orb* transcripts appears to be restricted to the germline, and both the ovarian- and the testis-specific transcripts display unusual spatial patterns of accumulation. The pattern of localization of *orb* RNA in ovaries suggests that *orb* may have a role early in oocyte differentiation and later in establishing the correct spatial distribution of positional information in the oocyte.

Materials and methods

Isolation of the orb genomic recombinant, genomic and cDNA libraries, general methods

The *orb* genomic phage recombinant was isolated in a screen for genes expressed during oogenesis (Ambrosio and Schedl, 1984). Description of genomic phage recombinant libraries is presented elsewhere (Steward et al., 1984; Riggleman et al., 1989). cDNAs corresponding to the 4.7 kb transcript and the 3.2 kb transcript were isolated from a 0-3 hour embryonic library and an adult male library, respectively (Poole et al., 1985) Cloning techniques were used according to Maniatis et al. (1982). Standard Southern and northern analysis was used (Southern, 1975; Bhat et al., 1988; Salz et al., 1989). Random primed probes were used for Southern and northern experiments (Feinberg and Vogelstein, 1983).

In situ hybridization

For in situ hybridization 1 μm tissue sections of Oregon R ovaries were prepared and hybridized with a ^{35}S-labeled antisense riboprobe; control slides were hybridized with a sense strand riboprobe (Melton et al., 1984; Parks and Spradling, 1987). Whole-mount in situ hybridization to ovaries and embryos was performed according to Tautz and Pfeifle (1989) with the modifications of Suter and Steward (1991). Cryostat sections and hybridization with a ^{3}H-nick translated probe are described in Ambrosio and Schedl (1984). For whole-mount in situ hybridization to testes, 4- to 5-day-old males were dissected in Ringer's and then fixed in (10:1:30) 4% paraformaldehyde: dimethylsulfoxide. heptane for 20 minutes; the rest of the procedure was performed according to Tautz and Pfeifle (1989). PCR-generated single-strand antisense cDNA probe I (Saiki et al., 1988; Kreitman and Landweber, 1989) and random primed cDNA probe I or II were used for the whole-mount experiments (Feinberg and Vogelstein, 1983) (Fig. 1B). PCR sense probes were used as controls.

Sequence analysis

For DNA sequence analysis of cDNA clones, nested deletions were generated by treatment with exonuclease III and S1 nuclease (Henikoff, 1987) using the Bluescript vector system (Stratagene). Single-stranded DNA was sequenced by the method of Sanger et al. (1977) using the Sequenase system (US Biochemical). The GenBank database search was done with the TFASTA program (Lipman and Pearson, 1985). The computer programs used for the analysis of nucleic acid and protein sequences were from Staden (Staden, 1986) and GCG (Devereux et al , 1984). Polymorphisms between the predicted ovarian and testis proteins are as follows: aa 96, Ala to Val (GCC to GTC); aa 101, Gly to Arg (GGT to CGT); aa 634 Pro to Leu, (CCT to CTT).

Fly stocks

The wild-type strain Oregon R is described in Lindsley and Grell (1968). The *Bic-D*^{PA66} and *Bic-D*^{R26} recessive female sterile alleles were described in Steward et al. (1987), Suter et al. (1989), and Schüpbach and Wieschaus (1991). The *egl*^{WU50} allele is described in Schupbach and Wieschaus (1991). The *tudor*^{1} allele is described in Boswell and Mahowald (1985).

Results

Isolation of orb

In the course of previous studies on gene expression during oogenesis in *Drosophila*, we isolated a series of genomic phage recombinants which hybridize to poly(A)$^{+}$ RNA prepared from ovaries (Ambrosio and Schedl, 1984). A number of these recombinants were used as probes to examine the accumulation of RNA in cryostat tissue sections of *Drosophila* ovaries. One of these genomic clones, oo18, hybridized to RNA that is expressed very early in oogenesis and is localized in different regions of the oocyte during the course of oogenesis (see below). The oo18 genomic phage recombinant hybridizes to a single band on polytene chromosomes at position 94E11-13. Northern analysis using this phage or subcloned fragments as probes (A-F, Fig. 1A) detects only a single 4.7 kb transcript in poly(A)$^{+}$ RNA from adult females and 0-3 hour embryos. We have named the gene *orb* for oo18 RNA-binding (see below).

The orb ovary/early embryo-specific transcript

In order to characterize the 4.7 kb *orb* RNA, cDNAs were isolated from a 0-3 hour embryonic cDNA library using genomic fragment E as a probe (Fig. 1A). The structure of the transcription unit for the 4.7 kb *orb* RNA was determined by aligning the cDNA and genomic restriction maps, cross-hybridization experiments, and sequence analysis of a near full-length cDNA. The organization of the 4.7 kb *orb* transcription unit is diagrammed in Fig. 1B, while the sequence of the cDNA (D5) is presented in Fig. 2A.

The D5 *orb* cDNA is 4,737 nucleotides in length and contains a single long open reading frame (ORF). The first ATG codon in this ORF is located at nucleotides 783-785 giving a predicted protein-coding region of 2,745 nucleotides. This ATG is preceded by AACA which differs by one nucleotide from the *Drosophila* translational consensus sequence (C/A, A,A,C/A, Cavener, 1987). Translation could initiate at this ATG or at a second in-frame ATG (which has a better match to the *Drosophila* translational consensus sequence) located eighteen codons downstream. Upstream of these two ATGs, there are eight additional ATG codons which define several small ORFs with a maximum size of 54 codons (see Fig. 2A). In all but one case, the ATGs are preceded by sequences that poorly match the *Drosophila* translational consensus. Conceivably, these small ORFs may have some role in regulating the translation of the *orb* mRNA as has been found for some eukaryotic messages (Kozak, 1984; Hinnebush, 1988; Kozak, 1989).

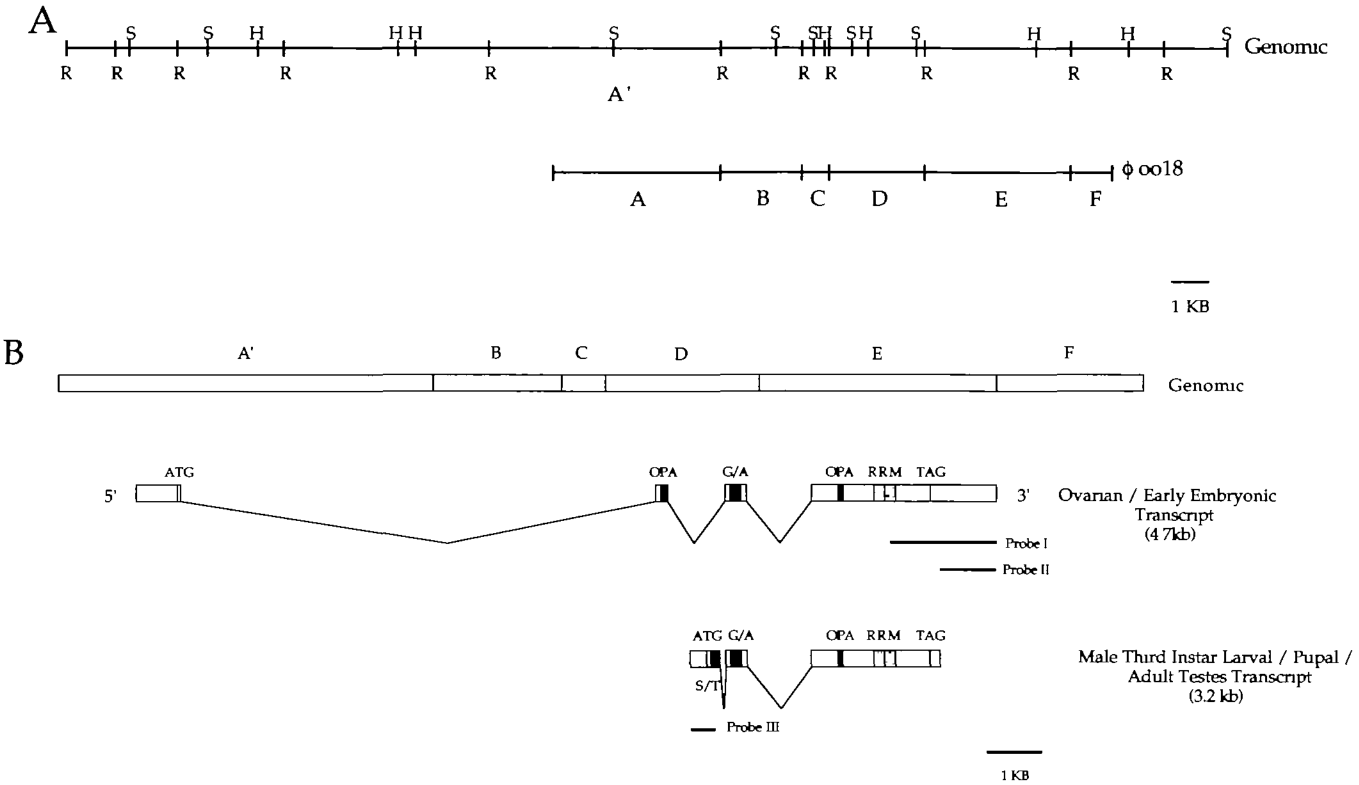


Fig. 1. Genomic organization of the *orb* region and structure of the *orb* transcripts. (A) At the top is a restriction map of the *orb* genomic region (R) *Eco*RI; (S) *Sal*I; (H) *Hind*III. Below is an *Eco*RI restriction map of the original oo18 genomic phage. Subcloned *Eco*RI restriction fragments (A', A-F) used to define the *orb* transcription unit are indicated (B) Structure of *orb* 4 7 kb ovarian/embryonic and 3.2 kb testis transcripts. Below the *orb* genomic region (*Eco*RI restriction fragments A'-F are indicated) are schematic diagrams of the approximate intron-exon structure of the *orb* transcripts. Also indicated are the RNA-binding domains (RRM), the opa or M repeat regions (OPA), the glycine/alanine-rich region (G/A), the serine/threonine-rich region (S/T), the initiation codon (ATG), and the termination codon (TAG). Location of probes I-III is also shown.

At the end of this ORF is a relatively long, 1207 nucleotide, 3′ untranslated region which is followed by six adenine residues. Since two other cDNAs (C36 and E4) end at the same position with a variable number of A residues (see Fig. 2A), this may correspond to the 3′ terminus of the 4.7 kb *orb* mRNA. Consistent with this suggestion, there is a possible polyadenylation signal, AAUAUA, fourteen nucleotides 5′ of these A residues. Even though this sequence differs from the eukaryotic consensus sequence for polyadenylation, AAUAAA (Proudfoot and Brownlee, 1982), it is thought to serve as the probable polyadenylation signal for transcripts encoded by the *Drosophila bicoid* and *cyclin A* genes (Berleth et al., 1988; Lehner and O'Farrell, 1989).

The predicted orb protein shows similarity to RNA-binding proteins

The 2,745 base ORF of the 4.7 kb ovarian/early embryonic transcript encodes a 915 amino acid protein of 99 × 10^3 M_r with a pI of 8.0. Comparison of the predicted *orb* amino acid sequence with the GenBank database revealed that the *orb* protein displays sequence similarities to the RRM family of RNA-binding proteins. Two regions of the predicted *orb* protein (amino acids 577-652 and 689-766, underlined in Fig. 2A) show similarity with the RNA recognition motif (RRM) or the ribonucleoprotein particle consensus sequence (RNP-CS) (Dreyfuss et al., 1988; Mattaj, 1989; Kenan et al., 1991). This family of RNA or single-stranded nucleic acid binding proteins includes proteins involved in binding the poly(A) tail of cytoplasmic mRNAs, splicing factors and hnRNP proteins. Fig. 3 shows a sequence comparison between the two *orb* putative RNA-binding domains and a number of other proteins in this class. Each region of similarity extends over approximately 80 amino acids, and there is 20-25% identity between each of the *orb* RRM domains and the RRM domains of other members of this family. Two peptides, RNP2 and RNP1, are highly conserved among all the proteins in this family and are present in each of the RNA-binding domains of the predicted *orb* protein (Bandziulis et al., 1989). The most distinctive feature of RRM proteins is the conserved aromatic amino acids (Tyr or Phe) in the RNP2 and RNP1 regions (Kenan et al., 1991). These conserved aromatic residues are present in the first *orb* RRM domain, at residue 579 in RNP2 and at residues 619 and 621 in RNP1. In addition, a basic residue located in RNP1 has

```
A
CGGAAATTCGAGTTAATGTGGTGCCATACCACAGTCGCAAAGATTTTCAACAATCAAAATAGTTTAATAGTAAAAAATGCGTCCGTGAGTGAGTAGTTTGGCGAAACAGCATCGGAAGGA  120

GAAACGGAAAACGGCGAAGGGCGTGCAAAGCAGAAGTAACGGTTCCTGGCGGCGTGTGTGCGAGTGTGTGTAAAAATGTGCTTGGTGGACGACAAAAGAAACGGGATGCGATTGCAAGTG  240

CGGAATTAAAATAAAATTCACACGAATTTATGCGCCATCCGCGTGTTTTTGCGAGTCCTGGCCTTTGGAAATTTGCATCGTTGACGAGGCGGAATGAATCGATGACTATTTGCTGCTGAT  360

ACGGCCACATCTGCCAGCTGTCTTGGTTCGCCTTCGCCCCTCTCGCTCGCTCAGTCGCGTTATCCTGTGGAGCTCCAATAAACAACAACAAAAACCAGATTCCTTGCCCGAGGCTCGGCA  480

TCCAATTCAAACTCCAACCAGATCCAACACGAACTTGGACTCGAACCGCAGCGAACTCCAACCGAAAAGCCTGCGCAGGAGTGAAAGAGAGAGGGCGAGAGAGACGGTGCGCGAACGGAC  600

TCTCCGCCGGCAGTGGGAGTGGGAAAGGGAAGGTACGAGAGGCGGAAGTGGTGACATCTGTTTTGCCGCTTTGTCAGTTGGCGGATGCTGCTCCTGCCGAACACGAACGAACGCGAACGA  720

                                                                                     -----------------------------------
                                                                                     M P L L Q Q Y D T P D C S G S G G N M R   20
GCACGAACGAACAAGGAGCGCAACGTGAACAAGTGTGGAAATTGGAATTTTTATTCAAAACAATGCCTCTGCTGCAACAGTACGACACCCCCGATTGCTCCGGTTCTGGTGGCAACATGC  840

------------------------------------------------------------------------------------------------------------------------
 A L S G G S T T E L L Q K H S I S S Y L D H H H Q Q Q Q Q Q Q H H L Q L Q Q H Q   60
GAGCCCTGAGCGGAGGCTCCACAACGGAGTTGCTGCAGAAACACTCGATAAGCTCCTACCTGGATCACCACCATCAGCAGCAGCAGCAACAGCAGCATCACTTGCAACTGCAGCAACACC  960

-------------------------------------------------------------
 Q Q H S L L E R C N D D G L I S F I N D P I T L N D L L G L C G A S T A N E V V   100
AACAGCAGCACAGTCTGTTGGAGAGATGCAACGACGATGGATTGATATCTTTTATAAACGATCCAATCACACTGAACGATCTACTCGGCTTGTGCGGTGCCAGCACTGCCAATGAAGTGG  1080

 G T G Q T P S T S A P I L G A G G G G R A N G V T A G A A T A T G V G V G A G G   140
TCGGTACTGGCCAGACCCCCTCGACATCAGCGCCGATACTTGGAGCAGGAGGAGGCGGAAGAGCAAATGGAGTTACAGCCGGCGCAGCAACGGCAACAGGAGTGGGAGTTGGAGCTGGAG  1200

 T L P G P G V P S I Q G G G G G G V V G Q Q T N A S C N T S A A N P S A S F G G   180
GGACATTACCTGGGCCAGGAGTTCCATCGATCCAGGGAGGTGGTGGAGGAGGAGTCGTGGGACAACAAACCAATGCCTCGTGCAACACCTCAGCGGCTAACCCATCAGCCAGCTTCGGTG  1320

 N G S S S D V N N L L L A S A A A A A A A A G D G A Q L S A N A A A Y A P L T   220
GCAACGGCAGCAGCAGCGACGTCAATAATCTGCTGCTCGCCAGTGCTGCTGCCGCTGCCGCCGCGGCTGCCGGTGATGGCGCTCAACTGTCCGCCAATGCGGCAGCCTACGCTCCGCTGA  1440

 P S S T H S S A S P G T K S N F D Y F Q F E N V A Q S N P L K A F Q R T N I S F   260
CGCCCAGCTCCACTCATTCATCGGCCTCGCCGGGCACAAAGTCCAATTTTGACTACTTCCAGTTTGAAAATGTTGCACAAAGTAACCCACTTAAGGCTTTCCAAAGAACAAACATCAGCT  1560
                                                                                            5'12A1a

 D C S A P L S P S T P T S I Y N R S F H S S P L V S D S S N S S S G I G L S M D   300
TCGATTGCTCTGCACCCTTATCGCCAAGTACGCCTACATCTATTTACAATCGCTCCTTTCACAGTTCACCTTTAGTCAGCGACAGCAGCAACTCGTCCAGCGGCATTGGACTCTCCATGG  1680

 S I N M F Y N Q Q Q Q Q Q Q P E Q Q G Y T S L G N S M G S G L G L S L A N A S T   340
ACTCCATCAACATGTTCTACAACCAGCAACAACAACAGCAACAGCCGGAGCAGCAAGGTTACACCAGTTTGGGCAACTCCATGGGCAGCGGATTGGGACTTAGTTTGGCAAACGCTTCCA  1800

 R S N S P E S Q N S S N S T T E Q N L L D M I N L L S V N S N K I P H Q Q Q Q Q   380
CGCGCTCGAACTCGCCAGAGAGTCAGAACAGCAGCAACTCAACGACTGAACAAAATCTACTGGACATGATTAATTTGCTGTCTGTAAATAGCAACAAAATTCCACATCAACAACAGCAAC  1920
                                                                                                   3'9A1

 Q Q Q Q Q Q Q Q Q Q Q Q Q N Q Q Q L Q V Q Q Q H Q L Q Q Q F V N L N R N Y E Q Q   420
AACAACAGCAGCAGCAACAACAACAGCAGCAGCAACAACAGAATCAGCAGCAATTGCAGGTTCAGCAACAACATCAGTTGCAGCAGCAGTTTGTTAACCTAAACAGGAACTACGAGCAAC  2040

 I S A N L G S Q Q H G F E H N G V G V G A S S S G N E N C F S Q Y N L E N I T S   460
AGATATCTGCCAATTTGGGTAGCCAACAACATGGGTTCGAGCACAACGGCGTGGGTGTGGGTGCTTCTAGTAGCGGAAATGAGAACTGCTTTAGCCAGTATAATCTAGAAAACATCACCA  2160

 V D M E L A K L Q N L Q R I N T L K L L H A Q A Q Q M P L I N Q L L Q S Y A G N   500
GTGTGGACATGGAGCTGGCCAAACTGCAGAATTTGCAACGCATCAACACGCTAAAGCTCCTGCATGCCCAAGCTCAACAAATGCCGCTGATCAACCAGCTCCTGCAGAGCTATGCCGGCA  2280

 A I G S V G G S N L G N L M S A G G S S L M T E M A G N V G G I I T T N D G H L   540
ACGCCATAGGCAGCGTCGGTGGAAGCAATCTTGGTAACTTGATGAGCGCAGGCGGATCATCACTGATGACAGAAATGGCGGGAAACGTTGGCGGCATTATAACCACCAACGATGGCCACC  2400

 D R V A K F Y K S S A A L C D A T C T W S G H L P P R S H R M L N Y S P K V F L   580
TGGATCGCGTAGCTAAGTTCTATAAGAGTTCGGCTGCCCTGTGCGATGCCACATGCACATGGAGTGGACACCTGCCGCCGCGTTCGCACCGTATGCTCAACTATTCTCCCAAGGTCTTCC  2520

 G G I P W D I S E Q S L I Q I F K P F G S I K V E W P G K E Q Q A A Q P K G Y V   620
TCGGCGGGATTCCCTGGGATATTAGTGAGCAGTCGCTCATCCAGATCTTCAAGCCATTTGGATCTATTAAAGTGGAGTGGCCCGGCAAGGAGCAGCAGGCGGCTCAGCCCAAGGGTTATG  2640

 Y I I F E S D K Q V K A L P S A C V L Q V D D S H C G R N Y F F K I S S R R I K   660
TTTACATAATCTTTGAATCGGACAAGCAGGTCAAGGCATTACCTTCGGCCTGTGTGCTTCAGGTGGATGATTCTCACTGTGGTAGGAACTACTTCTTCAAAATCTCTTCGCGGCGTATTA  2760
```

Fig. 2. For legend see p 80

```
S K D V E V I P W I I A D S N F V R S S S Q K L D P T K T V F V G A L H G K L T   700
AGTCCAAGGATGTTGAAGTCATTCCTTGGATTATCGCTGACTCCAATTTTGTGCGATCCAGCTCCCAGAAACTTGACCCAACGAAAACCGTGTTTGTGGGCGCACTGCATGGAAAACTGA   2880

A E G L G N I M D D L F D G V L Y A G I D T D K Y K Y P I G S G R V T F S N F R   740
CTGCCGAGGGCTTGGGAAATATAATGGATGATCTTTTCGACGGCGTGCTGTATGCTGGTATAGATACGGACAAGTACAAGTACCCGATCGGATCGGGACGTGTGACATTTAGCAACTTTC   3000

S Y M K A V S A A F I E I R T T K F T K K V Q V D P Y L E D A L C S I C G V Q H   780
GCTCCTATATGAAAGCTGTTTCGGCCGCCTTTATCGAGATTAGGACCACGAAGTTCACCAAGAAGGTGCAGGTGGATCCCTACTTGGAGGATGCCCTATGTTCCATATGCGGTGTGCAGC   3120

G P Y Y C R E L S C F R Y F C R S C W Q W Q H S C D I V K N H K P L T R N S K S   820
ACGGTCCCTACTATTGTAGGGAATTATCGTGCTTCCGATACTTCTGCCGCAGCTGCTGGCAATGGCAGCACAGCTGTGACATCGTCAAAAATCACAAGCCCTTGACTCGCAACTCCAAGT   3240

Q S L V G I G P S S S N V S L P F S G Q R S I R D N R M G N G Q H Q Q H Q Q H Q   860
CGCAGAGCCTGGTTGGCATCGGACCATCTTCGTCGAATGTTTCGTTACCCTTCTCTGGCCAACGAAGTATCAGGGACAACAGAATGGGGAACGGTCAGCATCAACAGCACCAGCAGCATC   3360

Y Q Q Q K H R Q L Q E Y S Q P H S L N V M G N S G A A N A A A T S M V T L Q Q R   900
AGTACCAACAGCAGAAACACCGTCAGCTGCAAGAATACTCGCAGCCCCACAGTCTCAACGTGATGGGAAACTCAGGAGCTGCCAATGCCGCTGCTACATCAATGGTAACTCTTCAGCAGC   3480

Q I H K V R I Q R Q Q H Q A I *                                                  915
GGCAAATTCACAAGGTGCGAATACAGCGTCAACAGCATCAGGCGATCTAGACTGTTACGGCTTTTATCCACACCGTTTTAACGGATGTTTCCGCAATATAATGTGTGAGACTTTGGACTT   3600

GTAGGCGACGTAGAAATATGGCGGAGATAGTTTTCTTGCAGCCGATGGAGACCCGGGCTCATCATCGTCATTATCATTGTAAACACAGTTTTATTGATGTGTATATAATATCGGTACAAT   3720

ATGAAGCTTACACTTTGAGTTTCATTTAAGATAATTATTGTTAGACGACTCCCCCAAAAACGAACACCCTCGAATCGAAACAGAAGAGAGACCCCCATTCATACTATTTTTTGATAATCA   3840
   3'12A1   3'5B1 3'4A1
    3'2A1
TGCTTACATATCAGCATTTGGAGCTGGCATTCGAATGCTAAATGAATGATACATGAAATGAGCAATTTCATTTACTCATACCCCCTCATACAAATACTACTCAAGTTACCTTTTGGCTAA   3960

GAAAGCAGTTATTAATTTCTAATTTGTATGTTTAATTTTAAGGAAACGAAACCGCGCGGCAAATGGCTACTTGAAATGTCTCGACCAATTTGCCGCGCTGCGAGATTAGAAACACCATTT   4080

TTGTTATGTTTTGTGCTATTACTGAGGAATTTTATGTAGTTTCTTTTTACATAGCCAAGCCCCCGACTCGAGTTAATATGATATTATATTATTTGGATTGTCCGCTAAGCGTTTATCAGG   4200

AATTTCAATTTTTAAGAAAACATTTTAAAAATTGTAAATTCGTTTAACTCACCAGTCTCCCTTTTTTGTTTTTCATTGAATTGTTACACACATGAGTTATTATAAATATACTGTTTAGTT   4320

TATTTTTTGTTTTGGATTAGCTTTAAGCATTATCCTTGTGAACATTAACGCGATGCCTGATTGATTGTTGAACTATGTTTTAAGCATTTATCATTCTTTGGCTTTTTTGGCTGTTCCAGT   4440

TTTATAGCTCATGGGCAATAAGCATATAATTTATGTCTACTTTATTTTTCATGTATTTTGAGAAAGGATCTCTCTAGCGCTTATCTATAGAAAACACACATATGTATGTATACAAAAATA   4560

TTACATAGTGTACATTACAAGGCTTTATATAATTTAAAACTTGATAAGTTTGTAAACCTACACAGAATAGGAAAAAAATACTTAAGTCTATATCTTAAACACGAATTGATCACTAAACGA   4680

TAAAATAACGCACACACACCTATTCACCAACCAAACAAATATACAAAATCGCTATTGAAAAAAn                                                             4737
                                                               3'D5
                                                               3'C36
                                                                3'E4
```

```
B ATAGTTTTTTACTTTTAGAAGTCGGCACACATAATTCGTAGACACTGTCATTATTGTTATTAACTTAATTTAATCATACACATCCTTTTTCAAGTCGGACTGCACCCCCGCCGTCCGCAG   120

CCATAAATTTTCATTTTCATTTCCATTTCAAATTGTCGCAACGGTGCCAGCAAGAAGCGCTAATTAAATTTTACAGTACCAATCAAAATTTTAAATAAAATAGGCCCCAAATTGTTTCCG   240

                                                                                   M L G V E K P Y V V E P A M   14
TTCAATAAGAAAGCGCAGACGCAGTAGTTGCAGTCCGTATCGAAGGCCAGAAAGGATTCCATCAGGATCAGTGCACACATGTTGGGAGTCGAGAAGCCATATGTGGTAGAGCCAGCTATG   360

A V A Q D S F E F G S N G D G S S T N S H T S N A S S K D L G R L T S G D G A Q   54
GCCGTAGCGCAGGACTCGTTTGAATTTGGCAGCAACGGCGATGGGAGCTCCACCAACAGCCACACAAGCAACGCCAGCTCCAAGGACCTTGGCAGGCTGACTAGTGGCGACGGAGCCCAG   480

N L A Q D S S A N S I D Q D S T D N T N S D C N                                   78
AACTTGGCCCAAGACAGCAGTGCCAACAGCATTGACCAGGACAGCACGGACAACACAAACAGCGACTGCAAT                                                 531
```

Fig. 2. For legend see p. 80

Fig. 2. Sequence of the *orb* transcripts and the predicted *orb* proteins. (A) Sequence of *orb* ovarian/embryonic cDNA D5. The eight ATG codons upstream of the large ORF and the first two ATGs in the ORF are in boldface type. The putative RNA-binding domains are underlined (aa 577-652 and 689-766) The glycine/alanine-rich region and the opa or glutamine-rich regions are italicized. The probable polyadenylation signal of the ovarian mRNA is double underlined. The 80 amino acids in the ovarian protein which differ from the testis protein are overlined. The position of the last nucleotide of the cDNAs is located above the last digit in the cDNA name; ovarian cDNAs: D5, C36 and E4; testis cDNAs: 12A1, 4A1, 2A1 and 5B1. The 5′ and 3′ ends of the 12A1 cDNA are also indicated. An AT-rich sequence upstream of the polyadenylation sites of the testis cDNAs is underlined. (B) Sequence of first 531 nucleotides of the 3.2 kb testis transcript and predicted amino-terminal protein sequence. Only the first 531 nucleotides of the 9A1 cDNA including the amino-terminal amino acid sequence which differs from the ovarian protein are shown. The rest of the nucleotide and protein sequence is essentially identical to the ovarian protein (see Materials and methods for a list of polymorphisms).

been reported to be important for RNA-binding activity of the snRNP protein U1-A and is present in the majority of RRM proteins including the first RRM domain of the *orb* protein. The second *orb* RNA-binding domain is an atypical RRM. As indicated in Fig 3, the second *orb* RRM has the conserved aromatic amino acid in RNP2; however, in RNP1 the two conserved aromatic residues are replaced by Ser at position 731 and an Arg at position 733. A Ser residue is also found at the equivalent position in the La protein which has been shown to bind RNA in vitro and in vivo (Reddy et al., 1983; Matthews and Francoeur, 1984; Chambers and Keene, 1985; Chambers et al., 1988). Similarly, the first RRM of the yeast protein, PRP-24, which is involved in U4/U6 snRNA base pairing (Kenan et al., 1991), has an Arg residue in RNP1 corresponding to Arg 773 of the predicted *orb* protein.

The three-dimensional structure of the RNA-binding domain of the U1-A protein has been determined by NMR analysis and X-ray crystallography (Hoffman et al., 1990; Nagai et al., 1990). These studies suggest that the RNA-binding domain of U1-A consists of two alpha helices and a four-stranded antiparallel beta sheet. Alignment of the RNA-binding domains of a number of RRM proteins with U1-A reveals strong conservation of residues that correspond to the hydrophobic core positions of U1-A which are thought to be important for formation of the alpha helical and beta sheet secondary structure (Kenan et al., 1991). The conserved arrangement of hydrophobic residues is also apparent in both *orb* RNA-binding domains (Fig. 3). From these comparisons it appears that the *orb* protein is a new member of the RRM family of RNA-binding proteins.

A number of RNA-binding proteins in this family possess regions outside of the RNA-binding domain which are abundant in a particular amino acid(s) (Bandziulis et al., 1989). These 'auxiliary domains' have been suggested to be involved in protein-protein interactions or in affecting the specificity or affinity of binding of the protein to its target RNA (Bandziulis et al., 1989). The putative *orb* ovarian protein also contains such amino acid sequences (see Figs 1B and 2A). There are glutamine-rich regions located in the amino terminal half of the protein which resemble the opa or M repeat (Wharton et al., 1985; Courey et al., 1989). A glycine/alanine-rich region of about 100 amino acids is located between the two opa repeats. Glycine-rich regions have been suggested to be flexible regions which may act as hinges between different domains of a protein (Creighton, 1984; Haynes et al., 1987).

orb encodes female- and male-specific transcripts

The developmental profile of the *orb* transcript(s) was determined using a unique cDNA fragment (probe I in Fig. 1B) to probe northern blots. These experiments revealed two findings of interest. First, the accumulation of *orb* RNA is restricted to specific developmental stages. Second, the *orb* gene encodes female- and male-specific transcripts. As shown in Fig. 4, the 4.7 kb *orb* transcript is present in adult females, ovaries and 0-2 hour embryos, but is not detected at other stages of development or in males. Two smaller *orb* transcripts of 3.2 kb and 2.0 kb are found only in males and first appear in male late third instar larvae. A high level of *orb* RNA is detected by in situ hybridization in the larval gonad of males, while only a very low level is evident in the larval gonad of females (K. Bhat, personal communication). They are also found in pupae and adult testes but are not detected at any other stages of development. Since the *orb* transcripts are found in female and male gonads, it was of interest to determine if their accumulation is dependent upon a functional germline. The progeny of homozygous *tudor* females do not form pole cells; consequently, adult F_1 progeny do not have the germline components of the ovary or testes but apparently do have normal somatic components (Boswell and Mahowald, 1985). Poly$(A)^+$ RNA from male and female F_1 progeny of *tudor* females was hybridized with *orb* probe I. As shown in Fig. 4B, none of the *orb* transcripts are detected in these animals suggesting that *orb* expression requires a functional germline.

orb RNA is localized in ovaries and early embryos

We used whole-mount and tissue section in situ hybridization to examine the pattern of accumulation of *orb* RNA during oogenesis and early embryogenesis. The results of this analysis are summarized below.

Germarium

The *Drosophila* ovary is composed of about 15 ovarioles containing egg chambers at different stages of development (reviewed by King, 1970; Mahowald and Kambysellis, 1980). At the distal end of each ovariole is a structure called the germarium. Region 1 of the germarium contains the stem cells, cystoblasts (a product of the division of a stem cell), and 2-, 4- and 8-cell cysts which arise from the cystoblasts by a series of mitotic divisions with incomplete cytokinesis. As shown

			Beta-1 (RNP2: + * + +)	Loop-1	Alpha-1 (+ ++ +)	Loop-2	Beta-2 (+ + +)	Loop-3
Fly	oo18#1	577	K V F L G G	I P W D - I S	E Q S L I Q I F K P F G	S I - - K	V E W P G K	E Q Q A
	oo18#2	689	T V F V G A	L H G K - L T	A E G L G N I M D D L F	D G V L Y	A G I D T D	K Y K -
	SXL#1		N L I V N Y	L P Q D - M T	D R E L Y A L F R A I G	P I - N T	C R I M R D	Y K T -
	SXL#2		N L Y V T N	L P R T - I T	D D Q L D T I F G K Y G	S I - V Q	K N I L R D	K L T -
Human	U1-A#1		T I Y I N N	L N E K I K K	D E L K K S L Y A I F S	Q F - G Q	I L D I L V	S R S -
	U1-A#2		I L F L T N	L P E E - T N	E L M L S M L F N O F P	G F - - K	E V R - L V	P G R -
	La RNP		S V Y I K G	F P T D - A T	L D D I K E W L E D K G	O V - L N	I Q - M R R	T L H -
	U1 snRNP 70K		T L F V A R	V N Y D - T T	E S K L R R E F E V Y G	P I - K R	I H M V Y S	K R S -
	RNPC		R V F I G N	L N T L V V K	K S D V E A I F S K Y G	K I - - -	V G - C S V	H - - -
Yeast	PRP24#1		T V L V K N	L P K S - Y N	Q N K V Y K Y F K H C G	P I - I H	V D V A D S	L - - -
	PRP24#2		T L W M T N	F P P S - Y T	Q R N I R D L L O D I N	V V - A L	S I R L P S	L R F -
	PRP24#3		E I M I R N	L S T E L L D	E N L L R E S F E G F G	S I - E K	I N I P A G	O K E H
	PABP#1		S L Y V G D	L E P S - V S	E A H L Y D I F S P I G	S V - S S	I R V C R D	A I T K
	PABP#2		N I F I K N	L H P D - I D	N K A L Y D T F S V F G	D I - L S	S K I A T D	E N - G
	PABP#3		N L Y V K N	I N S E - T T	D E Q F Q E L F A K F G	P I - V S	A S L E K D	A D - G
	PABP#4		N L F V K N	L D D S - V D	D E K L E E E F A P Y G	T I - T S	A K V M R T	E N - G
Rat	HDP#1		K L F I G G	L S F E - T T	D E S L R S H F E Q W G	T L - T D	C V V M R D	P N T -
	HDP#2		K I F V G G	I K E D - T E	E H H L R D Y F E Q Y G	K I - E V	I E I M T D	R G S -
Hamster	C23		T L F V K G	L S E D - T T	E E T L K E S F E G S V	R A - R I	V T D R E T	G S S -

			Beta-3 (RNP1: * + * + * + +)		Alpha-2 (+ ++ +)	Loop-5	Beta-4 (+ + +)	
Fly	oo18#1	A Q P K	G Y V Y I I F	E S	D K Q V K A L P S A C	V L Q V D D S H C - -	G R N Y F F	652
	oo18#2	- Y P I	G S G R V T F	S N	F R S Y M K A V S A A	F I E I R T T K F T K	K V Q V D P	766
	SXL#1	G Y S F	G Y A F V D F	T S	E M D S Q R A I K V L	N G I T V R N K R - -	L K V S Y A	
	SXL#2	G R P R	G V A F V R Y	N K	R E E A Q E A I S A L	N N V I P E G G S Q P	L S V R L A	
Human	U1-A#1	L K M R	G Q A F V I F	K E	V S S A T N A L R S M	Q G F P F Y D K P - -	M R I Q Y A	
	U1-A#2	- - - H	D I A F V E F	D N	E V Q A G A A R D A L	Q G F K I T Q N N - A	M K I S F A	
	La RNP	K A F K	G S I F V V F	D S	I E S A K K F V E T P	G Q K Y K E T D - - -	L L I L F K	
	U1 snRNP 70K	G K P R	G Y A F I E Y	E H	E R D M H S A Y K H A	D G K K I D G R R - -	V L V D V E	
	RNPC	- - - K	G F A F V Q Y	V N	E R N A R A A V A G E	D G R M I A G Q V - -	L D I N L A	
Yeast	PRP24#1	K K N F	R F A R I E F	A R	Y D G A L A A I T K -	T H K V V G Q N E - -	I I V S H L	
	PRP24#2	N T S R	R F A Y I D V	T S	K E D A R Y C V E K L	N G L K I E G Y T - -	L V T K V S	
	PRP24#3	S F N N	C C A F M V F	E N	K D S A E R A L Q M -	N R S L L G N R E - -	I S V S L A	
	PABP#1	T S - L	G Y A Y V N F	N D	H E A G R K A I E Q L	N Y T P I K G R L - -	C R I M - -	
	PABP#2	K S - K	G F G F V H F	E E	E G A A K E A I D A L	N G M L L N G Q E - -	I Y V A P H	
	PABP#3	K L - K	G F G F V N Y	E K	H E D A V K A V E A L	N D S E L N G E K - -	L Y V G R A	
	PABP#4	K S - K	G F G F V C F	S T	P E E A T K A I T E K	N Q Q I V A G K P - -	L Y V A I A	
Rat	HDP#1	K R S R	G F G F V T Y	A T	V E E V D A A M N A R	P H K V D G R V V - E	P K R A V S	
	HDP#2	G K K R	G F A F V T F	D D	H D S V D K I V I Q K	Y H T V N G H N C - E	V R K A L S	
Hamster	C23	- - - K	G F G F V D F	N S	E E D A K A A K E A M	E D G E I D G N K - -	V T L D W A	

Fig. 3. Alignment of the putative *orb* RNA-binding domains with domains from other RNA-binding proteins. Sequences are aligned in relation to the human U1-A RNA-binding domains (Kenan et al., 1991). The predicted location of the alpha helices, beta sheets and loop regions are indicated. Amino acids which compose the hydrophobic core are designated by a plus (+). The conserved peptides RNP1 and RNP2 are indicated (Bandziulis et al., 1989). The conserved basic and aromatic amino acids are indicated by asterisks. *Sxl* is a *Drosophila* sex determination switch gene (Bell et al., 1988); U1-A and U1-70K are snRNPs which associate with U1 snRNA (Sillekens et al., 1988; Query et al., 1989; Spritz et al., 1987; Theissen et al., 1986), La autoantigen binds the U-rich 3′ ends of nascent RNA polymerase III transcripts (Chambers and Keene, 1985; Chambers et al., 1988); RNPC is an hnRNPC protein (Swanson et al., 1987); PRP24 is a yeast protein involved in U4/U6 base pairing (Shannon and Guthrie, 1991; Kenan et al., 1991); PABP is yeast poly(A)-binding protein (Adam et al , 1986, Sachs et al., 1986), HDP is rat helix destabilizing protein (Cobianchi et al., 1986); C23 is nucleolin, a nucleolar specific phosphoprotein (Lapeyre et al., 1987).

in the whole-mount ovariole in Fig. 5A and 5B, little or no *orb* RNA is detected in germarial region 1. Region 2 contains as many as eight to ten 16-cell cysts. The 16 cells in each of these cysts are interconnected by one to four cytoplasmic bridges or ring canals. One of the two cells that has four ring canals becomes the oocyte while the remaining cells develop into nurse cells. *orb* RNA is first detected in this region of the germarium; however, as shown in Fig. 5A only a subset of the cells (6-7) accumulate a high level of RNA. From the number and distribution of cells containing high levels of RNA, we suspect that these cells correspond to the presumptive oocytes in each of the 16-cell cysts. A much lower level of *orb* RNA is found in the other cells of the cysts which are likely to correspond to presumptive nurse cells.

Stage 1-14 egg chambers

Oogenesis in *Drosophila melanogaster* has been divided into 14 stages based on the morphology of the egg chambers (reviewed by King, 1970; Mahowald and

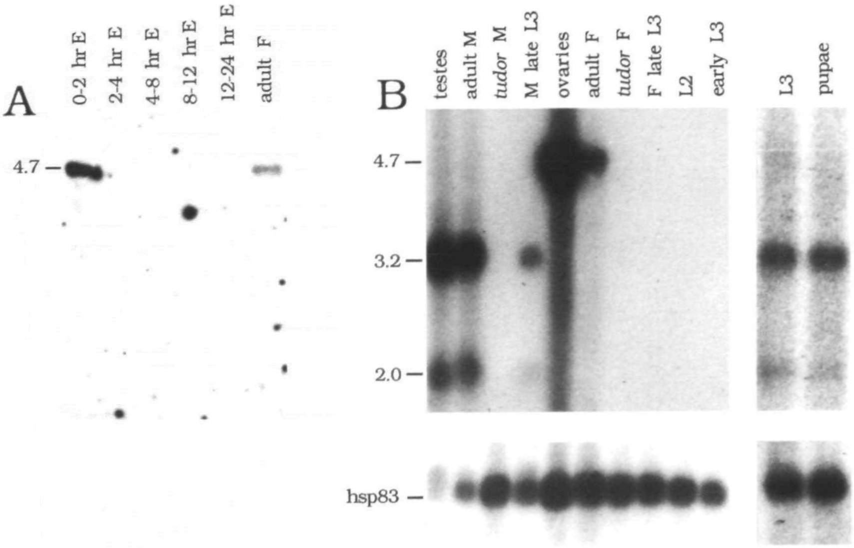


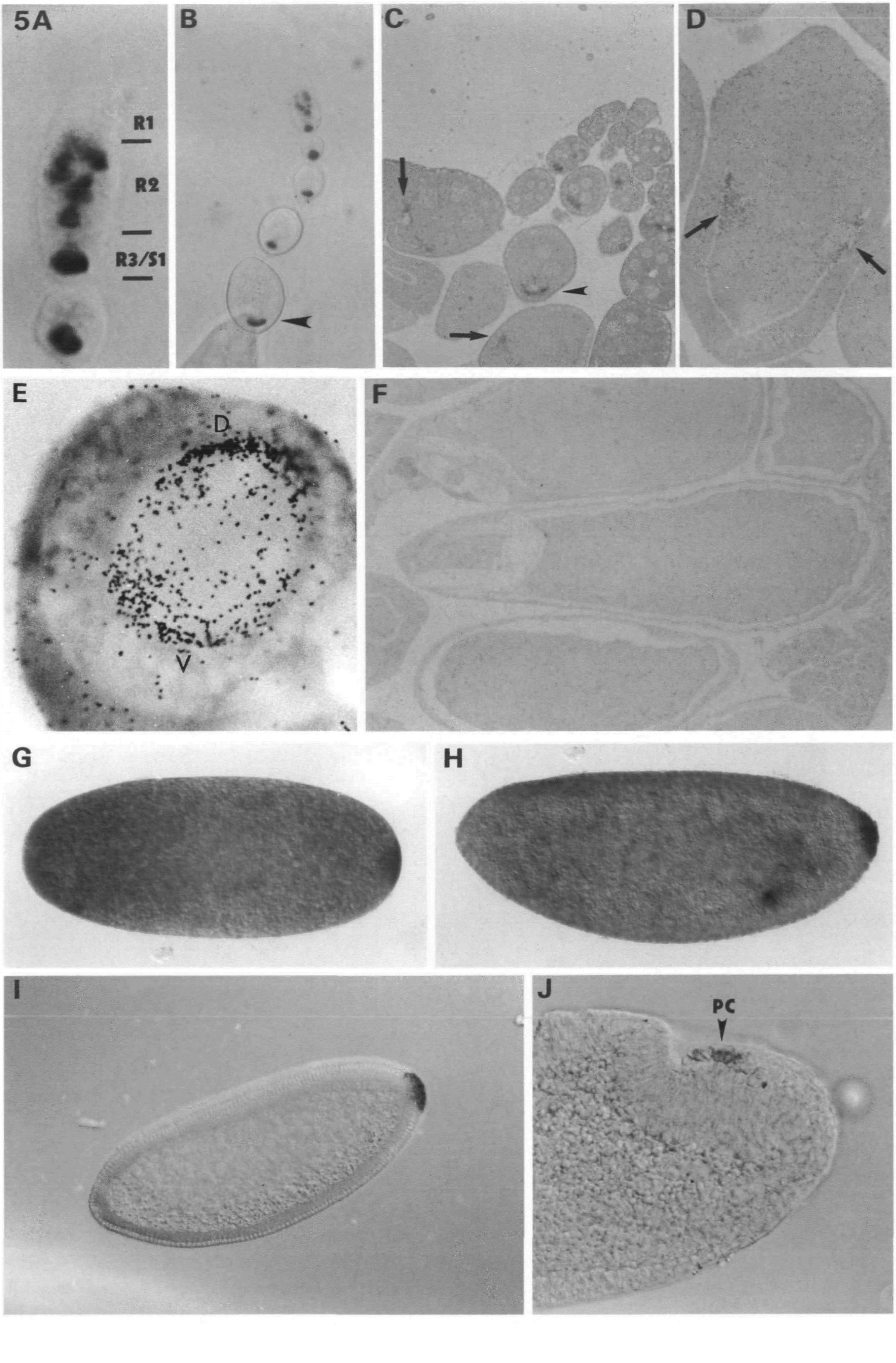
Fig. 4. Temporal and tissue specificity of the *orb* transcripts. (A) The autoradiograph in this panel shows the developmental profile of *orb* transcripts during embryogenesis: E, embryos and F, females All lanes contain 10 μg of poly(A)$^+$ RNA. As a control, the blot was reprobed with cDNAs for *Sxl* and *undifferentiated* (Bell et al., 1988; Bhat, personal communication). (B) Autoradiograph of a poly(A)$^+$ RNA blot showing the tissue specificity of the *orb* transcripts. Testes, dissected adult testes; M, male; *tudor*, F_1 adult male progeny of homozygous *tudor*^{1} females which lack a germline; M late L3, male late third instar larvae; ovaries, dissected adult ovaries; F, females; *tudor* F, F_1 adult female progeny of homozygous *tudor*^{1} females which lack a germline; F late L3, female late third instar larvae; L2, second instar larvae, early L3, early third instar larvae. All lanes contain 4 μg of poly(A)$^+$ RNA except the testes RNA lane which contains 2 μg. L3, third instar larvae and pupae lanes contain 5 μg of poly(A)$^+$ RNA. Blots were hybridized with probe I (see Fig. 1B). The 4 7 kb RNA is only detected in ovaries, adult females containing a germline, and 0-2 hour embryos The 3.2 kb and 2.0 kb RNAs are only observed in male late third instar larvae, pupae, adult testes, and adult males containing a germline As a control for RNA loading the blot presented in (B) was reprobed with hsp83 (Mason et al , 1984).

Kambysellis, 1980). An egg chamber consists of the oocyte/15 nurse cell complex surrounded by somatically derived follicle cells. The nurse cells synthesize RNA and other components which are deposited in the oocyte during the course of oogenesis, while the follicle cells secrete the vitelline membrane and chorion. As shown in Fig. 5B and 5C, *orb* RNA was found to preferentially accumulate in the most posterior cell of the egg chamber (the oocyte) during stages 1-6, while only low levels of RNA were observed in the nurse cells. At stage 7 when the oocyte can easily be distinguished from the nurse cells, the *orb* RNA in the oocyte is localized at the posterior end (Fig. 5B and 5C).

Interestingly, the pattern of *orb* RNA localization changes dramatically during subsequent stages of oogenesis. From stage 8-10 *orb* RNA no longer accumulates in the posterior of the oocyte but instead is located in the anterior of the oocyte close to the nurse cell/oocyte border (Fig. 5C and 5D). In whole-mount in situ experiments (not shown) the hybridization appears to extend as a ring around the nurse cell/oocyte border. In cross sections *orb* RNA also appears to be asymmetrically distributed along the dorsal-ventral axis at this time; higher levels of RNA are found in the dorsal and ventral cortical regions, while less RNA is observed in the lateral regions (Fig. 5E). During the last stages of oogenesis, 11-14, *orb* RNA is no longer localized in the anterior of the oocyte (Fig. 5F). Instead, weak hybridization is detected throughout the oocyte suggesting that the RNA is uniformly distributed. There may also be a decrease in the level of RNA.

Throughout oogenesis no *orb* RNA can be detected in the somatically derived follicle cells. This is consist-

Fig. 5. In situ hybridization to wild-type ovaries. (A,B, G-J) Whole-mount preparation of ovaries or embryos were probed with digoxigenin-labeled probe I or II (see Fig 1B). (C, D and F) Plastic tissue sections of ovaries were hybridized with an ^{35}S-labeled riboprobe (probe I, Fig. 1B). (A) Enlargement of the distal tip of the ovariole including the germarium and a ~stage 2 egg chamber. Regions 1-3 of the germarium are indicated. (B) Germarium and stage 2-7 egg chambers. The *orb* ovarian RNA preferentially accumulates in the presumptive oocytes in region 2 of the germarium and during stages 1-7. At stage 7 we observe posterior localization of the RNA within the oocyte. Stage 7 egg chamber is indicated by an arrowhead. (C) Stage ~3-9 egg chambers. *orb* RNA accumulates in the oocyte early and at stage 7 is posteriorly localized within the oocyte, while at stage 8-9 it is now anteriorly localized. Stage 8-9 egg chambers are indicted by arrows. (D) Early stage 10 egg chamber. *orb* RNA is no longer posteriorly localized and is now anteriorly localized. Hybridization is indicated by arrows (E) Cryostat tissue sections (cross sections) through a stage 8 egg chamber hybridized with ^3H probe E (see Fig. 1A). Notice the apparent asymmetric distribution of the *orb* RNA along the dorsal-ventral axis D, dorsal; V, ventral. (F) Stage 13 egg chamber *orb* RNA is uniformly distributed during stages 11-14. (G) Prior to pole cell formation. *orb* RNA accumulates as a cap at the posterior end while lower levels of RNA are distributed throughout the rest of the embryo. (H) Pole cell formation. *orb* RNA is included in the pole cells when formed. (I) Cellular blastoderm. *orb* RNA is detected only in the pole cells. (J) Germ band extension. *orb* RNA is still present in pole cells (PC) as they are carried onto the dorsal side of the embryo during germ band extension. Dorsal is up and anterior is to the left.


5A
R1
R2
R3/S1
B
C
D
E
D
V
F
G
H
I
J
PC

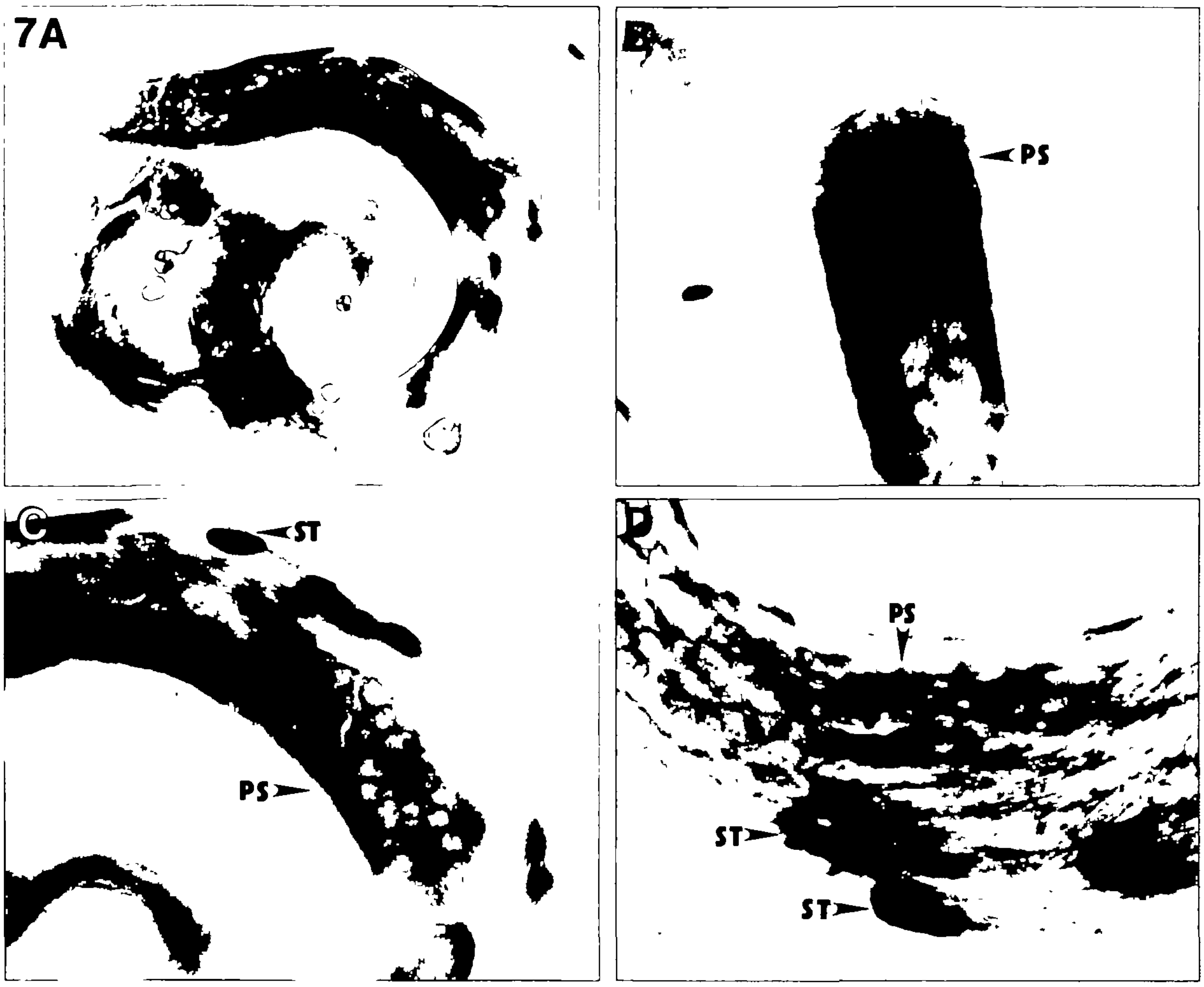


Fig. 7. In situ hybridization to wild-type testes Adult testes were hybridized with digoxigenin labeled probe III which is specific to the 3.2 kb transcript (Fig. 1B). PS, primary spermatocytes; ST, caudal tips of spermatid bundles. (A) Entire testis. (B) Distal end of the testis. *orb* RNA accumulates in the primary spermatocytes. (C) Older primary spermatocytes. (D) Caudal tip of spermatid bundle. *orb* RNA accumulates to a high level in the tip of the spermatid tails.

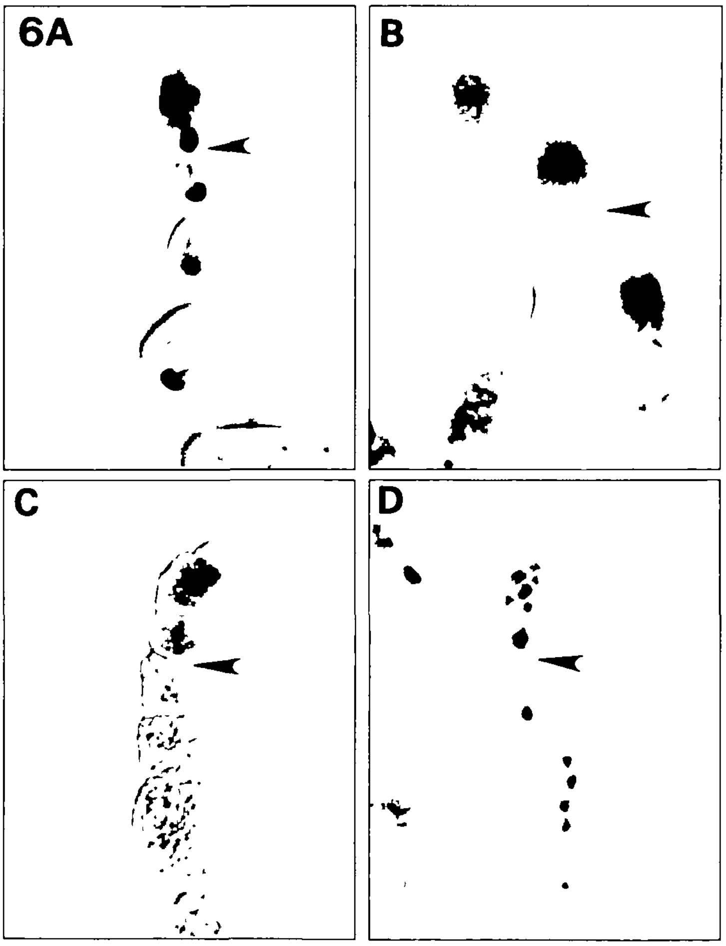


Fig. 6. In situ hybridization to *Bic-D* and *egalitarian* mutant ovaries. Wild-type and mutant ovaries were hybridized together with *orb* probe I. (A) Wild-type ovaries. (B) *egalitarian* mutant ovaries. (C) *Bic-D*^{PA66} mutant ovaries. (D) *Bic-D*^{R26} mutant ovaries. The end of the germarium (region 1-3/stage 1) is indicated by an arrowhead. *orb* RNA is more uniformly distributed in *egalitarian* and *Bic-D*^{PA66} early egg chambers than in wild type, but localized in the wild-type pattern at early stages in *Bic-D*^{R26} ovaries.

ent with the germline dependency of the ovarian transcript as determined by northern analysis of *tudor* progeny (see Fig. 4B).

Early embryos

We also examined the distribution of *orb* RNA in early embryos. Prior to pole cell formation, the highest level of *orb* RNA is found in a cap at the posterior end of the embryo, while lower levels of RNA are distributed throughout the rest of the embryo (Fig. 5G). In contrast, control embryos hybridized with a sense *orb* probe exhibit little or no staining. The posteriorly localized RNA is included in the pole cells when they first cellularize (Fig. 5H). During the syncytial blastoderm stage, the level of *orb* RNA decreases significantly except in the pole cells, and by cellular blastoderm formation it is only present in these germline cells (Fig. 5I). Consistent with the dramatic decrease in the level of *orb* RNA observed in whole-mount in situ experiments, the 4.7 kb transcript is not detected in northerns of RNA from 2-4 hour embryos (Fig. 4A). The pole cell RNA persists, at reduced levels, through stages 6-8 of embryogenesis as the pole cells are carried onto the dorsal side of the embryo during germ band extension (Fig. 5J). Interestingly, *orb* RNA appears to be uniformly distributed in the oocyte at late stages of oogenesis (11-14), and we did not observe preferential accumulation at the posterior. We suspect that the accumulation of *orb* RNA at the posterior pole in early embryos represents either the localization or stabilization of pre-existing maternal transcripts in this region at some point after egg deposition. The fact that the posterior RNA can be detected prior to the migration of pole cell nuclei would also argue that it is maternal in origin rather than the product of zygotic transcription.

Localization of orb RNA is perturbed in Bic-D and egalitarian mutant ovaries

Oocyte determination is believed to occur in the 16-cell cysts in region 2 of the germarium. *Bicaudal-D* and *egalitarian* recessive mutants appear to block this process, and the egg chambers in ovaries from homozygous mutant females fail to differentiate an oocyte (Mohler and Wieschaus, 1986; Schüpbach and Wieschaus, 1991). Since *orb* RNA preferentially accumulates in the presumptive oocytes of region 2 cysts, we were interested in examining the pattern of *orb* RNA in *Bic-D* and *egl* mutants.

We first probed northern blots of poly(A)$^+$ RNA from homozygous mutant *Bic-D* and *egl* females and determined that *orb* is expressed in both mutants (data not shown). We then used whole-mount in situ hybridization to ovaries from egl^{WU50} and *Bic-D* loss-of-function mutant females (*PA66* and *R26*) to examine the distribution of *orb* RNA. To control for variability in hybridization or staining, wild-type ovaries (Fig. 6A) were hybridized in the same tube with the mutant ovaries (Fig. 6B, C and D).

The pattern of accumulation of *orb* RNA in both *egalitarian* and *Bic-D* ovaries was very different from that of wild type. As shown in Fig. 6B and C, quite similar alterations in *orb* RNA localization were observed in egl^{WU50} and in one of the *Bic-D* alleles, $Bic\text{-}D^{PA66}$. In region 2 of the germarium in both these mutants, *orb* RNA is more uniformly distributed than in wild type, and we do not observe the striking preferential accumulation of RNA in the presumptive oocytes. In the youngest mutant egg chambers approximately the same level of *orb* RNA is found in all 16 cells, while in wild type it is clearly concentrated in one of these cells, the oocyte. In older mutant egg chambers *orb* RNA is no longer detectable. A different pattern of *orb* RNA accumulation is observed in ovaries from $Bic\text{-}D^{R26}$ mutant females (Fig. 6D). In region 2 of the germarium, the localization of *orb* RNA appears to be indistinguishable from that observed in wild type. Initially the RNA remains concentrated in the posterior-most cell (the oocyte) in the young mutant egg chambers; however, in somewhat older mutant egg chambers this localization is no longer evident, and, in fact, we do not detect any *orb* RNA.

The orb testes-specific transcripts

Since our developmental northern analysis suggested that the *orb* gene encodes sex-specific germline transcripts, we were interested in comparing the structure of the 4.7 kb ovarian/early embryonic transcript with the smaller testes-specific transcripts. To determine the approximate limits of the two testes-specific transcripts, we probed poly(A)$^+$ RNA from late third instar larvae, pupae, and adult males with fragments from the *orb* genomic phage (Fig. 1A). These experiments indicate that the 3.2 kb transcript is encoded by sequences in genomic fragments D and E, while the smaller 2.0 kb transcript appears to be contained within genomic fragment E (Fig. 1B).

To isolate *orb* testes cDNAs, an adult male cDNA library was screened with genomic fragments D and E. Although a full-length cDNA was not obtained for the 3.2 kb transcript, two partial overlapping cDNAs, 9A1 (1.4 kb) and 12A1 (2.2 kb), appear to include most of the 3.2 kb transcript. These two cDNAs were sequenced and found to overlap by 366 nucleotides to produce a composite cDNA of 3,242 bases, which is comparable in size to the larger testes transcript observed in northerns (see Figs 1B and 2). The probable structure of the 3.2 kb testes transcript is presented schematically in Fig. 1B. The testis transcript is colinear with the 4.7 kb ovarian/embryonic *orb* RNA over a region of ~2,700 bases. However, the transcripts differ at both their 5′ and 3′ ends (Fig. 1B).

At the 5′ end the ovarian/embryonic transcript has at least two exons, one encoded by fragment A′ and the other by fragment D, which are not found in the 3.2 kb testis RNA. Instead the 5′ most exon detected in this RNA (in northern blots and in male cDNAs) is derived from sequences located in the second intron of the ovarian/embryonic transcript. This would suggest that the 3.2 kb testis RNA is initiated from a different promoter than the ovarian transcript. The first ATG codon of the largest ORF in the 3.2 kb transcript is located in this male-specific 5′ exon at nucleotide 319-321. It is preceded by ACAC which differs from the translational consensus sequence by one nucleotide (Cavener, 1987).

The 3′ untranslated region of the 3.2 kb testis transcript is about 1,000 nucleotides shorter than 4.7 kb ovarian RNA. This difference in length appears to be due to sex-specific polyadenylation. Sequence analysis of several adult male cDNAs (12A1, 5B1, 2A1, and 4A1) reveals that they all have a poly(A)$^+$ terminus located about 200 nucleotides downstream of the translation termination signal. Interestingly, polyadenylation in the four cDNAs appears to occur at 3 different sites in a 20 nucleotide region. Moreover, no obvious nearby polyadenylation signals are discernable, although there is an AU-rich sequence located about 20 nucleotides upstream.

Finally, the structure of the smaller 2.0 kb testis RNA is uncertain at present. At the 5′ end our northern data suggests that it lacks the two 5′ most exons of the 3.2 kb transcript, while at the 3′ end it appears to terminate at the same position as the 3.2 kb RNA.

The 3.2 kb testis transcript is predicted to encode a protein that is nearly the same size as the ovarian protein (913 amino acids versus 915 amino acids), but has a different pI. The pI of the testis polypeptide is 7.48, whereas the ovarian is 8.0. The difference in charge between the male and female proteins is due to the N terminus which is encoded by exons unique to each sex. The first 80 amino acids of the female protein contain one of the glutamine-rich stretches, while the corresponding first 78 amino acids of the male protein contain a serine/threonine-rich region. Since the protein-coding regions of the last two exons are present in both males and females, the remainder of the *orb* protein is identical in the two sexes, and includes the two predicted RNA-binding domains, the glycine/alanine-rich region and one of the two opa sequences.

The distribution of orb RNA in testes

In *Drosophila* early stages of spermatogenesis are very similar to those of oogenesis (Cooper, 1950; reviewed by Lindsley and Tokuyasu, 1980). When a stem cell (which is located at the tip of the testis) divides, a primary spermatogonial cell is produced which undergoes four incomplete mitotic divisions to produce a 16-cell cyst. The interconnected members of this cyst, which are called primary spermatocytes, undergo a period of growth (~90 hours) where they increase in volume by ~25-fold. During this time, the cysts are pushed basally and towards the concave side of the testis. At the end of this period, meiosis occurs to produce a cyst of 64 interconnected cells which then begin to develop into spermatids. As each spermatid elongates, its caudal end extends towards the tip of the testis while its nuclear end is pushed more basally. Towards the end of differentiation, the spermatids separate and become individual spermatids leaving a waste bag or cystic bulge at their caudal ends as a result of the removal of cytoplasm and cytoplasmic organelles from the entire length of the spermatid.

In order to determine the distribution of the 3.2 kb testis-specific transcript, we performed whole-mount in situ hybridization to testes using a probe specific to this transcript (probe III, Fig 1B). The 3.2 kb transcript is first detected when the 16-cell cysts, the primary spermatocytes, are formed (Fig. 7A and B). In contrast to the localized distribution of the ovarian *orb* RNA during the early stages of oogenesis, the testis transcripts appear to be uniformly distributed in the cytoplasm of each member of the 16-cell cysts. Later in spermatogenesis, high levels of the 3.2 kb testis RNA are evident at the tip or caudal end of the spermatids in a structure which appears to correspond to the cystic bulge (Fig. 7A, C and D). In situ hybridization to testes with probe I, which hybridizes to both the 3.2 kb and 2.0 kb transcripts, shows the same pattern of distribution (data not shown).

Discussion

The possible importance of localized RNA in the specification of developmental fate in insects was initially suggested many years ago by work on early *Smittia* embryos (Kalthoff, 1973; Kandler-Singer and Kalthoff, 1976; Kalthoff, 1979). UV irradiation or RNAase treatment of the anterior region of these embryos was found to disrupt the establishment of the anterior-posterior axis generating double abdomen or *bicaudal*-like phenocopies. On the basis of these and other experiments, Kalthoff and co-workers argued that localized RNA molecules function as anterior determinants in *Smittia* embryos. The importance of localized RNAs in determining cell identity has been confirmed by more recent studies on genes involved in the establishment of anterior-posterior polarity during early *Drosophila* development (reviewed by St. Johnston and Nüsslein-Volhard, 1992).

It seems likely that RNA localization employs a strategy that is formally similar to that of protein targeting. Since most maternal transcripts are uniformly distributed within the oocyte and early embryo (*e.g.* heat shock, actin, histone and *dorsal*) (Ambrosio and Schedl, 1984 and 1985; Kobayashi et al., 1988; Steward et al., 1985), localized RNAs must have a *cis*-acting 'signal' sequence which specifies targeting. Macdonald and Struhl (1988) identified such a targeting element in the 3′ untranslated region of the *bicoid* mRNA. *orb* RNA also has a signal sequence in the 3′ untranslated region (Lantz and Schedl, unpublished data). The signal sequences are likely to be recognized by sequence-specific RNA-binding proteins which mediate RNA localization. These proteins could play a role in the targeting process not only by participating in the actual transport of the RNA but also by anchoring the RNA to the cytoskeleton once it reaches the appropriate destination within the cell.

Although a number of genes (e.g. *oskar*, *Bic-D*, *staufen*, *cappuccino*, *spire*, *swallow* and *exuperantia*) are known to be involved in the localization of RNAs during *Drosophila* oogenesis and early embryogenesis, only one of those that have been characterized at the molecular level encodes a protein product having properties suggesting that it interacts directly with RNA (reviewed by St. Johnston and Nüsslein-Volhard, 1992; Suter et al., 1989). However, vasa is a member of the RNA helicase family, and consequently may function in determining the 3-dimensional structure of the RNA or in the assembly/functioning of an RNA-protein complex rather than in the localization process per se (Hay et al., 1988; Lasko and Ashburner, 1988). From this perspective it is particularly intriguing that the germline-specific *orb* gene encodes a protein similar to a large family of sequence-specific RNA-binding proteins. The predicted protein product of the 4.7 kb *orb* ovarian RNA is a 99×10^3 M_r polypeptide which contains two closely spaced RRM type RNA-binding domains in its C-terminal half. Each domain is about 80 amino acids and shares sequence and structural features with the RRM protein family (Kenan et al., 1991). The amino acid sequence of the first RRM closely conforms to that of other members of the family. In contrast, the second RRM is atypical; at a number of conserved positions there are amino acid residues which occur only infrequently in other RRM domains. In spite of this amino acid sequence divergence, the structure of the second domain appears to have been conserved, and it has the characteristic RRM hydrophobic core.

Since other members of the RRM protein family have been shown to be involved in sequence-specific interactions with RNA (Dreyfuss et al., 1988; Mattaj, 1989), we would anticipate that the *orb* ovarian protein recognizes a specific RNA or class of RNA molecules in the developing oocyte. In this regard it may be significant that unlike some of the other RNA-binding proteins which contain two or more RRM domains, (*e.g.* yeast poly (A) binding protein or *Sex-lethal;* Adam et al., 1986; Sachs et al., 1986; Bell et al., 1988), the two *orb* RRM domains are rather different in primary amino acid sequence. Conceivably, this may indicate that the two domains in the *orb* protein have different RNA specificities.

During the very early stages of oogenesis the spatial distribution of the *orb* RNA is very similar to that reported for *Bic-D* RNA and protein (Suter et al., 1989; Wharton and Struhl, 1989; Suter and Steward, 1991). Like *Bic-D* gene products, localized *orb* RNA can first be detected in region 2 of the germarium. In this region, which contains several 16-cell cysts, *orb* RNA appears to preferentially accumulate in only one of the cells in each cyst, the presumptive oocyte. Since the other cells in each cyst (the presumptive nurse cells) have low levels of *orb* RNA, it is possible that the *orb* RNA is synthesized in these pro-nurse cells and then rapidly transported to the oocyte.

Not only does the localization of *orb* RNA resemble that of *Bic-D* RNA and protein (Suter and Steward, 1991), it is, in fact, dependent upon *Bic-D* function. In *Bic-D*^{PA66} mutant ovaries, the very early localization of *orb* RNA to the oocyte does not occur. Instead, the RNA is rather uniformly distributed in region 2 cysts and early egg chambers suggesting that *Bic-D* is required for the preferential accumulation of *orb* RNA in the oocyte. Interestingly, *Bic-D*^{PA66} causes an equivalent disruption of the localization of *Bic-D* protein. A different result is obtained in *Bic-D*^{R26} mutant ovaries. In this mutant, *orb* RNA initially accumulates in the presumptive oocytes in region 2 of the germarium and in early egg chambers (~stages 1-3), but its localization is not maintained in older chambers. A similar alteration is observed for *Bic-D* protein in the *Bic-D*^{R26} mutant. It is initially targeted to the presumptive oocytes; however, at about the same time that *orb* RNA dissipates, localized *Bic-D* protein can no longer be detected. The defects in *orb* RNA localization in *egalitarian* mutant ovaries would also suggest *Bic-D* dependence. In the *egl* mutant there is no preferential accumulation of *orb* RNA in the presumptive oocytes, and *Bic-D* protein is essentially uniformly distributed. The close correlation between the distribution of *orb* RNA and *Bic-D* protein during the early stages of oogenesis raises the possibility that *Bic-D* protein may participate directly in the localization of *orb* RNA.

Two other RNAs, *oskar* and *K10*, have also been found to preferentially accumulate in the presumptive oocyte during the early stages of oogenesis. Their pattern of accumulation is not, however, the same as *orb*. First, localized transcripts are not detected until somewhat later, late region 2 of the germarium for *oskar* and stage 2 for *K10* (Haenlin et al., 1987; Ephrussi et al., 1991; Kim-Ha et al., 1991; Cheung et al., 1992). Second, *Bic-D* mutations affect the localization of these RNAs differently than *orb* RNA (Suter and Steward, 1991). In the case of *K10*, both *Bic-D*^{PA66} and *Bic-D*^{R26} disrupt localization. In this respect, it may be significant that *K10* initially accumulates in the oocyte at a later stage than both *orb* RNA and *Bic-D* protein. In the case of *oskar* the initial accumulation in the oocyte appears to be unaffected by either *Bic-D* mutant. This would suggest that *oskar* targeting during early oogenesis may involve a pathway distinct from that for *orb* and *Bic-D*. Moreover, the fact that *oskar* localization is independent of *Bic-D* would argue that the disruption of *orb* RNA accumulation in *Bic-D* mutants cannot simply be explained by a 'change in cell identity' amongst the members of the cyst.

During previtellogenic stages *orb* RNA appears to accumulate preferentially towards the posterior end of the presumptive oocyte (data not shown). This asymmetric distribution is more clearly evident once the oocyte is distinguishable from the nurse cells at stage 7 (Fig. 5B and C). During stages 8-10, the localization of *orb* RNA in the oocyte changes; it is no longer concentrated at the posterior but instead is at the anterior and appears to be asymmetrically distributed along the dorsal-ventral axis. Interestingly, both *Bic-D* and *K10* RNA display a very similar pattern of localization during stages 1-10 (Haenlin et al., 1987; Suter et al., 1989; Cheung et al., 1992). Although *oskar* RNA also preferentially accumulates in the oocyte, the pattern is somewhat different than *orb* (Ephrussi et al., 1991; Kim-Ha et al., 1991). During stages 8-9 *oskar* RNA is found at both the posterior and the anterior of the oocyte and at later stages it is preferentially localized posteriorly. While localized *orb* RNA is not evident during the last part of oogenesis, it preferentially accumulates at the posterior pole of early embryos, like *oskar* and *nanos* RNA, and is subsequently included in pole cells, similar to *nanos* RNA (Wang and Lehmann, 1991).

The targeting of *orb* RNA during oogenesis may be important in determining the spatial distribution of *orb* protein. If this assumption is correct, one plausible function for *orb* protein would be to recognize *cis*-acting elements in specific RNAs and mediate their localization within the oocyte. While *orb* might play a role in targeting a variety of RNAs (e.g. *nanos* and *oskar*), the striking similarities between the distribution of *orb* RNA and that of *Bic-D* and *K 10* makes these two RNAs particularly good candidates. In this case, *orb* might be expected to participate during early oogenesis in some aspect of oocyte determination, while later in oogenesis it might function in establishing the correct spatial distribution of determinants required for early pattern formation in the embryo. Recently, P[ry^+]- (Dennis McKearin, personal communication) and EMS- (Lantz, unpublished observations) induced mutations in *orb* have been identified. Consistent with a function in the formation of the 16-cell cyst and/or oocyte determination, these have a tumorous ovary phenotype similar to that observed for *bag of marbles* (McKearin and Spradling, 1990). An understanding of the possible function of the gene later in oogenesis/embryogenesis will require the isolation of additional alleles. In order to mediate the localization of specific RNAs, *orb* would presumably interact with other proteins known to be involved in localizing determinants during oogenesis and/or with components of the cytoskeleton that have been implicated in the transport or anchoring of mRNAs (Yisraeli and Melton, 1988; Yisraeli et al., 1990; Pokrywka and Stephenson, 1991; Suter and Steward, 1991). Regions of the *orb* protein outside of the RNA-binding domains could potentially participate in such protein-protein interactions.

RNA localization may also be important for aspects of spermatogenesis. For example, the elongation of the spermatid may require the transport of mRNAs to the caudal end for protein synthesis and assembly. Since the predicted protein product of the *orb* 3.2 kb testis-specific transcript contains both of the RNA-binding domains, it may recognize specific RNAs and facilitate their localization. On the other hand, the *orb* testis protein may not function in precisely the same fashion as in the ovary. The testis protein differs at its amino terminus from the ovarian, and consequently may interact with different proteins,or perhaps even different RNA sequences/structures, than the ovarian protein.

We thank Suki Parks for her technical advice on the plastic ovarian tissue section experiment, Krishna Bhat for the embryonic northern blot, and Daniel Kenan for analysis of the RRM domains. We thank Trudi Schupbach, Beat Suter, Bob Clifford, Jamila Horrabin, Shubha Govind, and Daniel Bopp for their helpful comments on the manuscript and members of the Schedl and Steward laboratories for their support. This work was supported by a National Institutes of Health grant

References

Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K. and Dreyfuss, G. (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol Cell Biol* **6**, 2932-2943

Ambrosio, L. and Schedl, P. (1984) Gene expression during *Drosophila melanogaster* oogenesis. Analysis by in situ hybridization to tissue sections. *Dev Biol* **105**, 80-92

Ambrosio, L. and Schedl, P. (1985) Two discrete modes of Histone gene expression during oogenesis in *Drosophila melanogaster Dev Biol.* **111**, 220-231

Bandziulis, R., Swanson, M. and Dreyfuss, G. (1989) RNA-binding proteins as developmental regulators *Genes Dev* **3**, 431-437

Bell, L. R., Maine, E. M., Schedl, P. and Cline, T. W. (1988). *Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins *Cell* **55**, 1037-1046.

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M. and Nüsslein-Volhard, C. (1988). The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo *EMBO J* **7**, 1749-1756.

Bhat, K., McBurney, M. W. and Hameda, H. (1988). Functional cloning of mouse chromosomal loci specifically active in embryonic carcinoma stem cells *Mol Cell Biol* **8**, 3251-3259

Boswell, R. E. and Mahowald, A. P. (1985) *tudor*, A gene required for assembly of the germ plasm in *Drosophila melanogaster Cell* **43**, 97-104

Cavener, D. R. (1987) Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. *Nucl Acids res* **15**, 1353-1361

Chambers, J. C. and Keene, J. D. (1985) Isolation and analysis of cDNA clones expressing human lupus La autoantigen *Proc Natl Acad Sci USA* **82**, 2115-2119

Chambers, J. C., Kenan, D., Martin, B. J. and Keene, J. D. (1988) Genomic structure and amino acid sequence domains of the human La autoantigen *J Biol Chem* **263**, 18043-18051.

Cheung, H.-K., Serano, T. and Cohen, R. S. (1992) Evidence for a highly selective transport system during *Drosophila* oogenesis *Development* (in press)

Cobianchi, F., SenGupta, D. N., Zmudzka, B. Z. and Wilson, S. H. (1986). Structure of rodent helix-destabilizing protein revealed by cDNA *J Biol Chem* **261**, 3536-3543.

Cooper, K. W. (1950) Normal spermatogenesis in *Drosophila* In *The Biology of* Drosophila (ed , M Demerec) 1-61 New York: J Wiley and Sons

Courey, A. J., Holtzman, D. A., Jackson, S. P. and Tijan, R. (1989). Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* **59**, 827-836

Creighton, T. E. (1984). *Proteins Structures and Molecular Properties* New York W.H Freeman and Co

Dale, L., Matthews, G., Tabe, L. and Colman, A. (1989) Developmental expression of the protein product of Vg1, a localized maternal mRNA in the frog *Xenopus laevis EMBO J* **8**, 1057-1065.

Davis, L., Banker, G. A. and Steward, O. (1987). Selective dendritic transport of RNA in hippocampal neurons in culture *Nature* **330**, 477-479.

Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl Acids Res* **12**, 387-395

Dreyfuss, G., Swanson, M. S. and Pinol-Roma, S. (1988) Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation *Trends Biochem Sci* **13**, 86-91.

Ephrussi, A., Dickenson, L. K. and Lehmann, R. (1991). *oskar* organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell* **66**, 37-50

Feinberg, A. P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity *Analyt Biochem* **132**, 6-13

Garner, C. C., Tucker, R. P. and Matus, A. (1988) Selective localization of messenger RNA for cytoskeletal protein MAP2 in dendrites *Nature* **336**, 674-677.

Haenlin, M., Roos, C., Cassab, A. and Mohier, E. (1987) Oocyte-specific transcription of *fs(1)K10* a *Drosophila* gene affecting dorsal-ventral developmental polarity *EMBO J* **6**, 801-807

Hartl, F.-U. and Neupert, W. (1990). Protein sorting to the mitochondrion. Evolutionary conservations of folding and assembly *Science* **247**, 930-938

Hay, B., Jan, L. Y. and Jan, Y. N. (1988) Localization of *vasa*, a component of *Drosophila* polar granules, in maternal effect mutants that alter embryonic anteriopostenor polarity *Development* **109**, 425-433

Haynes, S. R., Rebbert, M. L., Mozer, B. A., Forquignon, F. and Dawid, I. B. (1987) pen repeat sequences are GGN clusters and encode a glycine-rich domain in a *Drosophila* cDNA homologous to the rat helix destabilizing protein *Proc Natl Acad Sci USA* **84**, 1819-1823.

Henikoff, S. (1987) Unidirectional digestion with exonuclease III in DNA sequence analysis *Methods in Enzymology* **155**, 156-165.

Hinnebush, A. (1988) Novel mechanisms of translational control in *Saccharomyces cerevisiae Trends Genet* **4**, 169-174

Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W. and Keene, J. D. (1990) RNA-binding domain of the protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins *Proc Natl Acad Sci USA* **88**, 2495-2499

Kalthoff, K. (1973) Action spectra for UV induction and photoreversal of a switch in the developmental program of an insect (Smittia) *Photochem Photobiol* **18**, 355-364

Kalthoff, K. (1979). Analysis of a morphogenetic determinant in an insect embryo (Smittia sp, Chironmidae, Diptera) In *Determinants of Spatial Organization* pp 79-126 (eds., Subtelny and I. Konigsberg) New York Academic Press.

Kandler-Singer, I. and Kalthoff, K. (1976). RNase sensitivity of an anterior morphogenetic determinant in an insect egg (Smittia sp, Chironmidae, Diptera) *Proc Natl Acad Sci. USA* **73**, 3739-3743

Kenan, D. J., Query, C. C. and Keene, J. D. (1991) RNA Recognition Towards identifying determinants of specificity *Trends in Biochem Sci* **16**, 214-220

Kim-Ha, J., Smith, J. L. and Macdonald, P. M. (1991). *oskar* mRNA is localized to the posterior pole of the *Drosophila* embryo *Cell* **66**, 23-35

King, R. C. (1970) *Ovarian Development in* Drosophila melanogaster New York. Academic Press

Kobayashi, S., Mizuno, H. and Okada, M. (1988). Accumulation and spatial distribution of poly $(A)^+$ RNA in oocytes and early embryos of *Drosophila melanogaster Develop. Growth and Differ* **30**, 251-260

Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucl Acids Res* **12**, 857-872

Kozak, M. (1989) The scanning model for translation· an update *J. Cell Biol* **108**, 229-241

Kreitman, M. and Landweber, L. F. (1989). A strategy for producing single-stranded DNA in the polymerase chain reaction a direct method for genomic sequencing *Gene Anal Techn* **6**, 84-88

Landry, S. J. and Gierasch, L. M. (1991) Recognition of nascent polypeptides for targeting and folding. *Trends in Biochem Sci* **16**, 159-163.

Lapeyre, B., Bourbon, H. and Amalric, F. (1987). Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence *Proc Natl Acad Sci. U S A.* **84**, 1472-1476

Lasko, P. F. and Ashburner, M. (1988). The product of the *Drosophila* gene *vasa* is very similar to eukaryotic initiation factor-4A *Nature* **335**, 611-617.

Lawrence, J. and Singer, R. H. (1986). Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell* **45**, 407-415

Lehner, C. and O'Farrell, P. (1989) Expression and function of *Drosophila* cyclin A during embryonic cell cycle progression. *Cell* **56**, 957-968

Lindsley, D. and Grell, R. (1968) Genetic variations of *Drosophila melanogaster Carnegie Inst. Wash Publ 627*

Lindsley, D. L. and Tokuyasu, K. T. (1980) Spermatogenesis In *The Genetics and Biology of* Drosophila, Vol 2d (eds , Ashburner, M., and Wright, T R F) 225-294 New York Academic Press.

Lipman, D. J. and Pearson, W. R. (1985) Rapid and sensitive protein similarity searches. *Science* **227**, 1435-1441

Macdonald, P. M. and Struhl, G. (1988) Cis-acting sequences responsible for anterior localization of *bicoid* mRNA in *Drosophila* embryos *Nature* **336**, 595-598

Mahowald, A. P. and Kambysellis, M. P. (1980) Oogenesis In *The Genetics and Biology of* Drosophila, Vol. 2d, (Ashburner, M., and Wright, T R F , eds) Academic Press, New York 141-224

Maniatis, T., Fritsch, E. R. and Sambrook, J. (1982) *Molecular Cloning* Cold Spring Harbor, New York Cold Spring Harbor Laboratory.

Mason, P. J., Hall, L. M. C. and Gausz, J. (1984) The expression of heat shock genes during normal development in *Drosophila melanogaster. Mol. Gen Genet* **194**, 73-78

Mattaj, I. W. (1989) A binding consensus RNA-protein interactions in splicing, snRNPs, and sex *Cell* **57**, 1-3.

Matthews, M. D. and Francoeur, A. M. (1984) La antigen recognizes and binds 3' oligouridylate tail of a small RNA *Mol Cell Biol* **4**, 1134-1140

McKearin, D. M. and Spradling, A. C. (1990) *Bag-of-marbles* a *Drosophila* gene required to initiate both male and female gametogenesis. *Genes Dev* **4**, 2242-2251.

Melton, D. A. (1987). Translocation of a localized maternal mRNA to the vegetal pole of *Xenopus* oocytes *Nature* **328**, 80-82.

Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. and Green, M. R. (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic acids Res* **12**, 7035-7056

Mohler, J. and Wieschaus, E. F. (1986). Dominant maternal-effect mutations of *Drosophila melanogaster* causing the production of double-abdomen embryos *Genetics* **112**, 803-822

Nagai, K., Oubridge, C., Jessen, T., Li, J. and Evans, P. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A *Nature* **348**, 515-520

Parks, S. and Spradling, A. (1987) Spatially regulated expression of chorion genes during *Drosophila* oogenesis *Genes Dev* **1**, 497-509

Pokrywka, N.-J. and Stephenson, E. (1991) Microtubules mediate the localization of *bicoid* RNA during *Drosophila* oogenesis. *Development* **113**, 55-66

Poole, J. S., Kauvar, B., Drees, B. and Kornberg, T. (1985) The *engrailed* locus of *Drosophila* Structural analysis of an embryonic transcript. *Cell* **40**, 37-43

Proudfoot, N. J. and Brownlee, G. G. (1982) 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**, 211-214.

Query, C. C., Bently, R. C. and Keene, J. D. (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein *Cell* **57**, 89-101

Rebagliati, M. R., Weeks, D. L., Harvey, R. P. and Melton, D. A. (1985) Localized maternal mRNAs in *Xenopus laevis* eggs *Cell* **42**, 769-777.

Reddy, R., Henning, D., Tan, E. M. and Busch, H. (1983) Identification of a La protein binding site in a RNA polymerase III transcript (4.5I RNA). *J Biol Chem* **258**, 8352-8356

Riggleman, B., Wieschaus, E. and Schedl, P. (1989) Molecular analysis of the *armadillo* locus uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene. *Genes Dev* **3**, 96-113

Sachs, A. B., Bond, M. W. and Kornberg, R. D. (1986) A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins. domain structure and expression *Cell* **45**, 827-835

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

Salz, H. K., Maine, E. M., Keyes, L. N., Samuels, M. E., Cline, T. W. and Schedl, P. (1989). The *Drosophila* female-specific sex determination gene, *Sex-lethal*, has stage-, tissue-, and sex-specific RNAs suggesting multiple modes of regulation. *Genes Dev* **3**, 708-719.

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

Schüpbach, T. and Wieschaus, E. F. (1991) Female sterile mutations on the second chromosome of *Drosophila melanogaster* II Mutations blocking oogenesis or altering egg morphology *Genetics* **129**, 1119-1136.

Shannon, K. W. and Guthrie, C. (1991) Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs *Genes Dev* **5**, 773-785.

Sillekens, P. T. G., Beijer, R. P., Habets, W. J. and van Venrooij, W. (1988) Human U1 snRNP-specific C protein: complete cDNA and protein sequence and identification of a multigene family in mammals *Nucl Acids Res* **16**, 8307-8321

Silver, P. A. (1991) How proteins enter the nucleus *Cell* **64**, 489-497

Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis *J. Mol Biol* **98**, 503-517

Spritz, R. A., Strunk, K., Surowy, C. S., Hoch, S. O., Barton, D. E. and Francke, U. (1987). The human U1-70K snRNP protein· CDNA cloning, chromosomal location, expression, alternative splicing and RNA-binding *Nucl Acids Res* **15**, 10373-10391.

Staden R. (1986) The current status and portability of our sequence handling software (Summary for May 1985). *Nucl Acids Res* **14**, 217-231.

Steward, R., Ambrosio, L. and Schedl, P. (1985) Expression of the *dorsal* gene. *Cold Spring Harbor Quant Biol* L, 223-228.

Steward, R., Ambrosio, L. and Schüpbach, T. (1987) Polarity in the oocyte and embryo of *Drosophila*. In *Molecular Approaches to Developmental Biology* (eds , Firtel, L A , and Davidson, E H) UCLA Symposia on molecular and cellular biology, new series 51, 39-50 New York Alan R, Liss, Inc

Steward, R., McNally, F. and Schedl, P. (1984) Isolation of the *dorsal* locus of *Drosophila*. *Nature* **311**, 262-265

St Johnston, D. and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo *Cell* **68**, 1-20

Suter, B., Romberg, L. and Steward, R. (1989). *Bicaudal-D*, a *Drosophila* gene involved in developmental asymmetry. localized transcript accumulation in ovaries and sequence similarity to myosin heavy chain tail domains *Genes Dev* **3**, 1957-1968.

Suter, B. and Steward, R. (1991). The role of the *Bicaudal-D* protein and its phosphorylation in *Drosophila* oocyte differentiation *Cell* **67**, 917-926.

Swanson, M. S., Nakagawa, T. Y., LeVan, K. and Dreyfuss, G. (1987) Primary structure of human nuclear ribonuclear particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins *Mol Cell Biol.* **7**, 1731-1739

Tannahill, D. and Melton, D. A. (1989). Localized synthesis of the Vg1 protein during early Xenopus development *Development* **106**, 775-785.

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

Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Lührmann, R. and Phillipson, L. (1986). Cloning of the human cDNA for the U1 RNA-associated 70K protein *EMBO J* **5**, 3209-3217.

Wang, C. and Lehmann, R. (1991) *nanos* acts as the posterior determinant in *Drosophila* *Cell* **66**, 637-647

Weeks, D. L. and Melton, D. A. (1987) A maternal mRNA localized to the vegetal hemisphere in *Xenopus* eggs codes for a growth factor related to TGF-b *Cell* **51**, 861-867.

Wharton, K., Yedvobnick, B., Finnerty, V. and Artavanis-Tsakonas, S. (1985) opa. A novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *D melanogaster*. *Cell* **40**, 55-62.

Wharton, R. P. and Struhl, G. (1989). Structure of the *Drosophila Bicaudal D* protein and its role in localizing the posterior determinant *nanos* *Cell* **59**, 881-892

Yisraeli, J. K. and Melton, D. A. (1988). The maternal mRNA Vg1 is correctly localized following injection into *Xenopus* oocytes *Nature* **336**, 592-595

Yisraeli, J. K., Sokol, S. and Melton, D. A. (1990). A two step model for the localization of maternal mRNA in *Xenopus* oocytes Involvement of microtubules and microfilaments in the translocation and anchoring of Vg1 mRNA *Development* **108**, 289-298.

(*Accepted 7 February 1992*)

Note added in proof

The sequence data in this paper has been assigned the following accession number: X64412 D. melanogaster orb mRNA by the EMBL Data Library.