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

VALERIE LANTZ¹, LINDA AMBROSIO² and PAUL SCHEDL¹

¹Department of Molecular Biology, Moffett Laboratory, Princeton University, Princeton, NJ 08544, USA
²Department 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, *Vgl* RNA is restricted to the vegetal cortex of *Xenopus* oocytes (Melton, 1987; Weeks and Melton, 1987). The asymmetric distribution of *Vgl* mRNA and protein and homology to TGF-β has led to the suggestion that the *Vgl* 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 35S-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 (1991). Cryostat sections and hybridization with a 3H-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) with the modifications of Suter and Steward (1984). Cryostat sections and hybridization with a 3H-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; Kretman 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 SI nuclease (Henkoff, 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 using 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 CTG); aa 634 Pro to Leu, (CTT to CTT).

Fly stocks

The wild-type strain Oregon R is described in Lindsey and Grell (1968). The Bic-DP406 and Bic-D265 recessive female sterile alleles were described in Steward et al. (1987), Suter et al. (1989), and Schupbach and Wieschaus (1991). The c64W150 allele is described in Schupbach and Wieschaus (1991). The tudor 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 AAC 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; Hinnebusch, 1988; Kozak, 1989).
Localized orb ovarian RNA

A

| R | R | R | S | H | H | H | R | R | R | R | S | S | S | S | H | H | H | S | S | S | H | R | R | R |
| R | R | R | R | R | R | R | R | R | R | R | R | R | R | R | R | R | R | R | R | R | R | R | R | R |

Genomic

B

A'DEF

1 KB

A

<table>
<thead>
<tr>
<th>ATG</th>
</tr>
</thead>
</table>

OPA

| C/A |

OPA

| RRM |

TAG

3'

Ovarian / Early Embryonic Transcript (4.7 kb)

| ATG/G/A |

OPA

| RRM |

TAG

5'

Male Third Instar Larval / Pupal / Adult Testes Transcript (3.2 kb)

1 KB

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) EcoRI; (S) SalI; (H) HindIII. Below is an EcoRI restriction map of the original oo18 genomic phage. Subcloned EcoRI 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 tests transcripts. Below the orb genomic region (EcoRI 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 $99 \times 10^3$ M, 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 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 (Bandzisius 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
Fig. 2. For legend see p 80
Localized orb ovarian RNA

Fig. 2. For legend see p. 80
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 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 F1 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 F1 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 Kambsellis, 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
Localized orb ovarian RNA

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 RNPI and RNPII are indicated (Bandzisius et al., 1989). The conserved basic and aromatic amino acids are indicated by asterisks. 5xI is a Drosoheada 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 U1-rich 3' ends of RNA polymerase III transcripts (Chambers and Keene, 1985; Chambers et al., 1988); RNPI is an hnRNPI 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 (Bonacci 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 germlarial 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 germlarium; 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. 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 35S-labeled riboprobe (probe I, Fig. 1B). (A) Enlargement of the distal tip of the ovariole including the gerarium and a ~stage 2 egg chamber. Regions 1-3 of the gerarium 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 gerarium 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 it 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.
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$	ext{PA66}$ mutant ovaries. (D) Bic-D$	ext{PA66}$ mutant ovaries. The end of the gerarium (region 1-3/stage 1) is indicated by an arrowhead. orb RNA is more uniformly distributed in egalitarian and Bic-D$	ext{PA66}$ early egg chambers than in wild type, but localized in the wild-type pattern at early stages in Bic-D$	ext{PA66}$ ovaries.
ent with the germine 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 germine 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 northern of RNA from 2-4 hour embryos (Fig. 6A). 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 gerarium. 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\(^{1WU59}\) 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\(^{1WU59}\) and in one of the Bic-D alleles, Bic-D\(^{PA66}\). In region 2 of the gerarium 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-D\(^{R26}\) mutant females (Fig. 6D). In region 2 of the gerarium, 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 germine 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 northern (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 RNase treatment of the anterior region of these embryos was found to disrupt the establishment of the anterior-posterior axis generating double abdomens 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 Nusslein-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 germ-line-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 $9.9 \times 10^5$ 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; Adami 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 K10 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, [1997]- (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 or localization of specific RNAs 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 Schuld and Steward laboratories for their support. This work was supported by a National Institutes of Health grant.

References


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


Lindsay, D. and Grell, R. (1986) Genetic variants of Drosophila melanogaster Carnegie Inst Wash Publ 627


(Received 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.