A Protein Component of Drosophila Polar Granules Is Encoded by vasa and Has Extensive Sequence Similarity to ATP-Dependent Helicases

Bruce Hay, Lily Yeh Jan, and Yuh Nung Jan
Howard Hughes Medical Institute
Department of Physiology
and Department of Biochemistry
University of California
San Francisco, California 94143

Summary

Determinants of pole cells, which are precursors of the germ line, are provided maternally and are localized to the posterior pole of the Drosophila egg, as are polar granules. It has been hypothesized that certain RNA molecules associated with polar granules may be necessary for pole cell determination. Using a monoclonal antibody (Mab46F11) against polar granules, we have cloned the gene for one of their components. This gene turns out to be vasa, which is required maternally for the formation of polar granules and germ cells. This polar granule component shows significant sequence similarity to elf-4A, a translation initiation factor that binds to mRNA, and to other helicases.

Introduction

The precursors of the Drosophila germ line (pole cells) form at the posterior pole of the embryo because of the action of maternally supplied and posteriorly localized cytoplasmic determinants (Jazdowska-Zagrodzinska, 1986; Okada et al., 1974; Illmensee and Mahowald, 1974; Warn, 1975; Illmensee et al., 1976; Niki, 1986; Frohnhöfer et al., 1986). Several observations associate these determinants with polar granules, which are densely staining cytoplasmic organelles without limiting membranes (Mahowald, 1962, 1968). First, polar granules are localized in the posterior pole plasm (germ plasm) in late-stage oocytes and early embryos (Counce, 1963; Mahowald, 1968). Second, both polar granules and the pole cell determinants appear to be associated with maternal RNA. UV irradiation or inhibition of protein synthesis early during embryogenesis blocks pole cell formation. Moreover, poly(A)+ RNA isolated from late-stage oocytes or early embryos can restore the ability of UV-irradiated embryos to form pole cells, although these pole cells do not develop far enough to produce germ cells (Okada and Kobeyashi, 1967). Interestingly, polar granules appear to be associated with RNA (Mahowald, 1971) only at those stages of oogenesis and embryogenesis when the polar plasm is capable of inducing germ line development, but not at late stages, e.g., after they are segregated into pole cells. Finally, the loss of maternally supplied activity of genes in the grandchildless-knirps class causes the absence of polar granules as well as the inability to form pole cells (Boswell and Mahowald, 1985; Schüpbach and Wieschaus, 1986; Lehmann and Nüsslein-Volhard, 1986; Nüsslein-Volhard et al., 1987). These observations are consistent with the hypothesis that polar granules are associated with a selective group of maternal mRNA molecules that are translated in early embryos to give rise to germ line determinants (Mahowald, 1968).

To elucidate the role of polar granules in germ cell determination, it is necessary to analyze the components of polar granules biochemically. We report here the cloning of a gene coding for one such component, a 72–74 kD protein recognized by the monoclonal antibody Mab46F11 (Hay et al., 1988). The deduced amino acid sequence suggests that this protein may be capable of binding and/or unwinding RNA.

Results

Isolation of cDNA for the Polar Granule Antigen and Localization of the Transcripts

Monoclonal antibody Mab46F11 binds to polar granules at the posterior pole of late-stage oocytes and early embryos, as well as to other germ line-specific structures throughout the fly life cycle (Hay et al., 1988). To clone the gene for this antigen, we screened an ovarian cDNA (Lambda ZAP) expression library with Mab46F11. Two classes of positive cDNA clones were isolated; clones within each class hybridize to one another and are probably derived from a single gene. Because cDNAs in the first class hybridize to transcripts expressed exclusively in germ cells (see below), whereas cDNAs in the second class hybridize to RNA throughout the embryo (not shown), we have only investigated clones in the first class and have used these cDNAs as hybridization probes to isolate full-length cDNA clones from an ovarian λgt11 library kindly provided by Dr. Laura Kalfayan. We have previously found a nuclear antigen that is present in somatic cells of late-stage embryos and is stained weakly by Mab46F11 (Hay et al., 1988), but we have not determined whether this nuclear antigen is related to the cDNAs of the second class.

cDNA from the first class (hereafter referred to simply as the cDNA) hybridizes to transcripts that show essentially the same distribution, notwithstanding one exception, as the strong Mab46F11 staining in embryos, ovaries, and testes (Figures 1 and 2). In the embryo, strong Mab46F11 staining is first localized to the posterior pole and is subsequently associated with pole cells (Hay et al., 1988). As shown in Figure 1, high levels of hybridizing transcripts are found uniformly distributed throughout the preblastoderm embryo. By the time of blastoderm formation, hybridizing transcript is essentially absent. Hybridization signal does not reappear until roughly stage 12 or early stage 13, the end of germ band shortening (staging according to Campos-Ortega and Hartenstein, 1985). When it does reappear, label is found over pole cells, which are arranged in bilateral clusters in the posterior third of the embryo (Figure 1C). At this and later stages of embryogenesis, pole cells are the only cells that contain...
Figure 1. Tissue in Situ Hybridization of Embryos with cDNA Isolated with Monoclonal Antibody Mab46F11

In each panel, anterior is to the left. (A) Parasagittal section through a preblastoderm embryo showing high levels of hybridization uniformly distributed throughout the embryo. (B) Parasagittal section through an embryo at the beginning of blastoderm formation. Hybridization intensity has decreased uniformly throughout the embryo to background levels. (C) Late stage 12, early stage 13 embryo. The only labeled cells are the pole cells (arrows). This is about the earliest stage that we can detect the hybridization signal in pole cells during embryogenesis. (D) Semi-parasagittal section through a first-instar larva. Hybridization is restricted to the pole cells (arrow).

hybridizing transcripts and show strong staining with the Mab46F11 antibody. This temporal distribution of the transcript suggests that zygotic transcription of this gene occurs primarily, if not exclusively, in pole cells (Figure 1).

In the adult female the Mab46F11 antigen is found in germ line stem cells, in nurse cells (sister cells of the oocyte), and at the oocyte posterior pole. In the adult male the antigen is present in germ cells in the testes at early stages of spermatogenesis (Hay et al., 1966). With the exception of late-stage oocytes, which show a uniform distribution of the hybridizing transcript (as in early embryos), transcript hybridizing to the cDNA is found exclusively in cells of the germ line lineage, paralleling the localization of the Mab46F11 antigen (Figure 2).

Although the Mab46F11 antigen is found sharply localized to the posterior pole of late-stage oocytes and early embryos, the transcript that encodes this antigen appears to be uniformly distributed in the cytoplasm of these cells (Figure 1). This difference may be accounted for by the following scenario. The Mab46F11 antigen is likely to be synthesized in the nurse cells and then transported into the oocyte, where it is targeted to the oocyte posterior pole (Mahowald, 1962; Hay et al., 1988). The transcript for the Mab46F11 antigen, however, probably flows into the oocyte with the bulk of nurse cell cytoplasm during the second half of oogenesis (Mahowald and Kambysselis, 1978) and does not appear to be targeted to any specific location.

Consistent with the observed distribution of transcripts for the Mab46F11 antigen, Northern analysis shows the presence of a single major transcript of about 2.2 kb, which is abundant in ovaries (not shown), female flies, and early embryos (Figure 3). Much lower levels of expression can be detected in late embryos (Figure 3) and in larvae (not shown).

**vasa Is the Gene for the Mab46F11 Antigen**

Since the Mab46F11 antigen is a component of the polar granules, a prime candidate for its structural gene might be one of the grandchildless-kniips class of maternal-effect gap genes, because mutations in these genes cause females to give rise to embryos without polar granules or pole cells (Boswell and Mahowald, 1985; Schüpbach and Wieschaus, 1986; Lehmann and Nüsslein-Volhard, 1986; Nüsslein-Volhard et al., 1987). Of the mutations in this class that we have tested (tudor, velois, oskar, staufen, vasa), only the vasa mutation (vasa\(^{PD23}\)) removed the Mab46F11 antigen. Early embryos from vasa\(^{PD23}\) mutant mothers lack the 72–74 kd protein (Hay et al., 1988) as well as Mab46F11 staining (Figure 4). In vasa\(^{PD23}\) adult females, immunoreactivity is limited to the germarium of the ovary. Substantial levels of the antigen, however, are still present in the adult male germ line of this vasa allele (data not shown). Since vasa\(^{PD23}\) is a hypomorphic allele (R. Lehmann, personal communication), these observations are compatible with the possibility that vasa codes for the Mab46F11 antigen.

The cDNAs isolated with Mab46F11 hybridize to the 35BC region on the left arm of the second chromosome (Figure 5), the same region that contains vasa. Further-
Figure 2. Tissue In Situ Hybridization of Adult Ovaries and Testis with cDNA Isolated with Monoclonal Antibody Mab46F11

(A) Low-magnification view of a section through an ovary. The germarium (germ), containing germ line stem cells and prointeriting cysts (Mahowald and Kambysselis, 1978), is located to the left. Later stages of oogenesis are located to the right. High levels of hybridization are found in the germarium and the nurse cell cytoplasm (ncc). Signal is absent from the nurse cell nucleus (ncn), however. At these stages (pre-stage 9, 10), the oocyte is relatively devoid of hybridization. The arrowhead points to an oocyte in which transport of transcript from nurse cells to the oocyte is beginning. Transcript accumulates throughout the oocyte beginning around stage 9 such that by stage 14, the mature oocyte, large amounts of transcript are found throughout the oocyte, similar to what is seen in the cleavage stage egg shown in Figure 1A.

(B) Higher-magnification view of a glancing section through a single egg chamber showing the lack of label associated with the somatic follicle cells (fc).

(C) Whole-mount squash of an adult testis showing hybridization limited to the apical tip (apex). This is the region in which germ line stem cells and early-stage spermatocytes reside.
Figure 3. Northern Blot of Maternal Poly(A)+ RNA from Female Flies and Poly(A)+ RNA from Different Embryonic Stages at 25°C

An EcoRI fragment that contains the amino half of the protein coding sequence in the full-length cDNA was used as a hybridization probe. Lane 1 shows an abundant transcript of roughly 2.2 kb in female flies. This transcript is also abundant in early (0-3 hr) embryos (lane 2) but occurs at much lower levels in 3-6 hr and 6-9 hr embryos (lanes 3 and 4). Poly(A)+ RNA from later stages (9-12 hr, lane 5; 12-16 hr, lane 6; 16-22 hr, lane 7) is overloaded (note actin controls below) to reveal the hybridization signals. Embryogenesis takes 22 hr.

Figure 4. Monoclonal Antibody Mab46F11 Staining of Whole-Mount Wild-Type Embryos and Embryos Derived from Homozygous vasapD3 Mothers

(A) Wild-type cleavage stage embryo showing Mab46F11 antigen localized to the polar plasm. (B) Wild-type embryo following pole cell formation. The Mab46F11 antigen has been incorporated into pole cells. (C) Cleavage stage vasapD3 embryo showing absence of localized Mab46F11 immunoreactivity. Also note that the background level of antigen throughout the embryo is decreased relative to the wild type. (D) vasapD3 embryo stage similar to the embryo in (B). Pole cells are absent and Mab46F11 immunoreactivity is undetectable.
transduction product has a predicted molecular mass of 71 kd. This predicted size is very similar to the size (72-74 kd) of the Mab46F11 antigen as measured on SDS–polyacrylamide geis (Hay et al., 1988), in agreement with the assignment of the vasa gene product as the Mab46F11 antigen.

The Predicted vasa Protein Sequence

We have noticed three regions of interest in the predicted vasa protein. A large internal domain of vasa has sequence similarity with eukaryotic translation initiation factor 4A (eIF-4A; Nielsen et al., 1985; Nielsen and Trachsel, 1988) and a nuclear antigen, p68, whose appearance is associated with cell proliferation (Ford et al., 1988; Figure 7). This sequence similarity has been noticed independently by Lasko and Ashburner (1988). Second, the carboxyl terminus of vasa contains multiple negatively charged residues (Figure 6), as in several single-stranded nucleic acid binding proteins. Finally, near its amino terminus the vasa protein contains a fivefold tandem repeat of the heptad F(or S)RGGE(or Q)GG (Figure 8). This sequence similarity has been noticed independently by Lasko and Ashburner (1988). The F(or S)RGGE(or Q)GG heptad repeats

Sequence Similarity with eIF-4A and Other Helicases

The vasa product shows 27% amino acid identity with eIF-4A and 31% identity with human nuclear antigen p68 (Figure 7). The function of the latter protein is unknown. The activity of eIF-4A, on the other hand, has been well characterized. eIF-4A is a component of a high molecular weight protein complex (called eIF-4F) involved in 5' cap recognition and the binding of mRNA to ribosomes. It has been suggested that eIF-4F functions to bind and unwind mRNA in an ATP-dependent manner; i.e., that eIF-4A is an ATP-dependent RNA helicase (Ray et al., 1985; Lawson et al., 1986; Abramson et al., 1987; Rhoads, 1988).

Sequence motifs thought to be required for ATP binding (Walker et al., 1982) are present in eIF-4A as well as p68 (Ford et al., 1988) and vasa (motifs I and II in Figure 7). In addition to these two motifs, five others have been found in a large number of ATP-dependent helicases (motifs Ia, III, IV, V, and VI; Hodgman, 1988a, 1988b; Lane, 1988). Six of these seven motifs can be readily identified in vasa (Figure 7). The extensive sequence similarity between vasa and eIF-4A and the presence of these motifs strongly suggest that the vasa product is also an ATP-dependent nucleic acid binding protein.

The Acidic Carboxyl Terminus of vasa

The carboxyl terminus of the vasa product is negatively charged; six of the last 12 amino acids are glutamate or aspartate residues (Figure 6). This feature is interesting in light of the sequence similarities between vasa and known helicases, since similar acidic carboxyl termini have been found in other RNA binding or single-stranded DNA binding proteins (Williams et al., 1983; Chase and Williams, 1986; Swanson et al., 1987).

The F(or S)RGGE(or Q)GG Heptad Repeats

The carboxyl terminus of the vasa product is negatively charged; six of the last 12 amino acids are glutamate or aspartate residues (Figure 6). This feature is interesting in light of the sequence similarities between vasa and known helicases, since similar acidic carboxyl termini have been found in other RNA binding or single-stranded DNA binding proteins (Williams et al., 1983; Chase and Williams, 1986; Swanson et al., 1987).

Discussion

We have cloned the gene for a polar granule component recognized by Mab46F11 (Iley et al., 1988). This gene was subsequently identified as vasa because its sequence is essentially identical to that of vasa (Lasko and Ashburner, 1988). Since the vasa223 mutation removes polar granules (Schüpbach and Wieschaus, 1988) as well as the vasa protein in oocytes and early embryos, most likely the vasa protein is either a major component of polar granules or a component important for their integrity. The failure of these mutant embryos to form pole cells, then, is consistent with the hypothesis that polar granules are involved in the determination of pole cells (Mahowald, 1988).
Sequence of the vasa Protein Suggests That It May Be an RNA Binding Helicase

The vasa sequence contains three features that are interesting in light of the association between polar granules and RNA (Mahowald, 1971) and in light of the implication of maternal RNA in pole cell determination (Okada and Kobayashi, 1987).

First, vasa has extensive sequence similarity with the translation initiation factor elf-4A and contains six of the seven motifs identified in a number of helicases. This suggests that the vasa protein may be able to associate with, and to unwind, RNA. Second, the highly acidic carboxyl terminus is reminiscent of the carboxyl termini of several single-stranded nucleic acid binding proteins. Some of these carboxyl termini appear to be involved in the interaction between the nucleic acid binding protein and other proteins (e.g., between the T4 gene 32 product and other proteins involved in DNA replication, or between the high-mobility group nonhistone proteins and histones); such interactions may alter the nucleic acidbinding affinity (Burv, et al., 1986; Hsieh, et al., 1982).

Finally, the amino-terminal region of vasa contains a fivefold tandem repeat of F(del S)ROGE(del O)QG. It has been pointed out that from purely physical considerations, the extent of a heptad repeat in any non-a-helical structure is unlikely to exceed two or three heptads (Cohen and Parry, 1986). The presence of five tandem heptads in vasa would therefore suggest an a-helical structure, although the presence of multiple glycine residues may weaken this prediction. These heptads and the flanking sequences are arranged in such a way that in an a-helix of 12' turns there would be six arginines along one face next to four phenylalanines and two serines (Figure 8B). Bearing in mind the involvement of basic residues and aromatic residues in the interaction with single-stranded nucleic acids (O'Connor and Coleman, 1983; Chaeo and Williams, 1986), we are tempted to consider the possibility that a single-stranded nucleic acid (e.g., RNA) interacts with the structure of heptad repeats so that the phosphate backbone has electrostatic interactions with arginines and the nucleotide bases interact with the phenylalanines by stacking, or with the serines by forming hydrogen bonds.

This model further raises the intriguing question of whether the postulated association with RNA has any specificity for primary sequences or secondary structures. Sequence-specific binding domains have been found at the amino-terminal region, separate from the putative helicase domain, of the SV40 large tumor antigen and of the transcription termination factor rho (Brennan et al., 1987; Gish and Botchan, 1987; Dombroski and Platt, 1988; Goetz et al., 1988). Such a model for the heptad repeats should be regarded with great caution, however, since no examples of a single a-helical structure interacting with nucleic acids have been shown. Besides, other possible functions for the heptad repeats exist, such as formation of a coiled-coil structure, although the heptad sequences in vasa do not agree very well with the preferred arrangement of heptads found in coiled coils (Cohen and Parry, 1988).
Figure 7. Homology of *vasa* Coding Sequence with Murine eIF-4AI and Human p68, and Identification of Domains Shared by a Large Number of Helicases

Alignments have been performed visually between eIF-4AI, p68, and *vasa*. Boxed regions indicate identity with the *vasa* sequence. Vertical lines and Roman numerals show our alignment of six domains within the *vasa* sequence identified as common to a large number of helicases (Hodgman, 1988a, 1988b; Lane, 1988). Letters below the protein sequence alignments indicate the Hodgman (1988a, 1988b) consensus sequences (in single-letter codes for amino acids) for these domains of helicases.

1976), either the relevant maternal RNA is restricted to the posterior pole or it has wider distribution but is only functional in the posterior pole. In the former case, the posterior localization of these maternal RNA species may be attributed to polar granules, e.g., binding of specific mRNA by the *vasa* protein. (A precedent has been found in Xenopus oocytes, where a protein binds to specific mRNA [Crawford and Richter, 1987].) Translation of such
mRNA in the embryo may follow either its release from pola-
granules or interactions between polar granule compo-
nents and other cellular constituents. If the relevant mater-
nal RNA is not restricted to the posterior pole, one could
imagine that vasa, or other polar granule components,
has to bind to these species of maternal RNA to allow for
their translation (Kozak, 1983, 1988). It remains to be de-
termined whether one or a combination of mechanisms
such as those described above, or others, are important
in pole cell determination.

Posterior Localization of the vasa Protein

Although the *vasa* protein is sharply localized to the
posterior pole as soon as it becomes detectable in the oo-
cyte (Hay et al., 1988), the *vasa* transcript is uniformly dis-
tributed in late-stage oocytes and early embryos (Figure 1).
Therefore, the targeting machinery must operate at the
level of the protein but not of the vasa mRNA. The vasa
protein may be either directly interacting with the targeting
machinery or associated with other proteins that are tar-
geted to the posterior pole—possibly other components
of polar granules. Indeed, our preliminary studies of
maternal-effect mutations of the *grandchildless-knirps*
class (tudor valois oskar staufen; unpublished results) indi-
cate that these genes may be involved in the establish-
ment of posterior localization or its maintenance.

Experimental Procedures

Stocks

*vasa*^{GOD} and deficiency stock *D(2)l75c* were provided by Drs. Trudi
Schupbach and Eric Wieschaus, Princeton University. Uiscidenc-
stocks *D(2)l472* and *D(2)l64* were obtained from Dr. Michael Ash-
burner and the Bowling Green stock center.

Immunocytochemistry

Mab46F11 is a monoclonal antibody that recognizes polar granules and
other germ line-specific structures. Immunocytochemical tech-
niques were as described in Hay et al. (1988).

cDNA Isolation and Sequencing

RNA was isolated from hand-dissected ovaries using the hot-phenol
method (Schwarz et al., 1988). Poly(A)^{+} RNA was isolated by two cy-
cles of adsorption to and elution from oligo(dT)-cellulose (Aviv and
Leder, 1972). This RNA was used to construct a Lambda ZAP expres-
sion library (Stratagene). cDNAs for the Mab46F11 antigen were iso-
lated from the expression library using standard techniques for plating,


