The *PAX* gene *egl-38* mediates developmental patterning in *Caenorhabditis elegans*

Helen M. Chamberlin[1], Robert E. Palmer[2], Anna P. Newman[2,†], Paul W. Sternberg[2], David L. Baillie[3] and James H. Thomas[1,*]

[1]Department of Genetics, Box 357360, University of Washington, Seattle, WA 98195-7360, USA
[2]HHMI/Division of Biology, 156-29, Caltech, Pasadena, CA 91125, USA
[3]IMBB/ Biology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

*Author for correspondence (e-mail: jht@genetics.washington.edu)
†Current address: Department of Biochemistry, Baylor College of Medicine, Houston, TX 77030-3498, USA

SUMMARY

Mutations in the *C. elegans* gene *egl-38* result in a discrete set of defects in developmental pattern formation. In the developing egg-laying system of *egl-38* mutant hermaphrodites, the identity of four uterine cells is disrupted and they adopt the fate of their neighbor cells. Likewise, the identity of two rectal epithelial cells in the male tail is disrupted and one of these cells adopts the fate of its neighbor cell. Genetic analysis suggests that the *egl-38* functions in the tail and the egg-laying system are partially separable, as different *egl-38* mutations can preferentially disrupt the different functions. We have cloned *egl-38* and shown that it is a member of the *PAX* family of genes, which encode transcription factors implicated in a variety of developmental patterning events. The predicted EGL-38 protein is most similar to the mammalian class of proteins that includes PAX2, PAX5 and PAX8. The sequence of *egl-38* mutant DNA indicates that the tissue-preferential defects of *egl-38* mutations result from substitutions in the DNA-binding paired domain of the EGL-38 protein. *egl-38* thus provides the first molecular genetic insight into two specific patterning events that occur during *C. elegans* development and also provides the opportunity to investigate the in vivo functions of this class of PAX proteins with single cell resolution.

Key words: *Caenorhabditis elegans*, *PAX* genes, paired domain, cell lineage, pattern formation

INTRODUCTION

An important question in animal development is how two cells become different from each other. Two general mechanisms can influence cell differences: cell interactions (non-autonomous mechanisms) and the unequal localization of factors (autonomous mechanisms) (reviewed in Jan and Jan, 1995). In either case, however, the effect is to influence the activity of the cell, often by influencing the transcriptional activity of the cell. Thus, transcription factors can play a key role in cell fate specification because they represent the link between the mechanism generating cell differences and cell differentiation. Understanding the function of a transcription factor can suggest the logic underlying a particular developmental decision, as its function can reflect the developmental choices available to a given cell at a given time. Because of this role, analysis of transcription factors can provide a unique vantage point for the particular developmental processes in which they are involved.

PAX proteins are transcription factors originally identified based on sequence similarity to the protein encoded by the Drosophila *paired* gene (Bopp et al., 1986; Walther et al., 1991). *PAX* genes from both vertebrates and invertebrates have been identified, and they play a variety of roles in animal development and disease (reviewed in Noll, 1993). PAX proteins all share the paired domain, a sequence motif that is a bipartite DNA-binding domain (Czerny et al., 1993). Different subclasses of PAX proteins also include a homeodomain, an octapeptide sequence, or both. The mammalian proteins PAX2, PAX5 (BSAP) and PAX8 are grouped together on the basis of sequence similarity within the paired domain, and they also include the octapeptide sequence and a partial homeodomain. Expression studies implicate these proteins in developmental patterning within several distinct tissues: the nervous system (PAX2, PAX5), the kidney and urogenital system (PAX2, PAX8), the thyroid gland (PAX8) and B lymphocytes (PAX5) (Dressler et al., 1990; Plachov et al., 1990; Adams et al., 1992). Mice with a genetic knockout of *Pax2* or *Pax5* have phenotypes consistent with predictions from expression: *Pax2* mutants have deformed urogenital systems and nervous system defects, and *Pax5* mutants lack mature B lymphocytes (Urbanek et al., 1994; Torres et al., 1995, 1996).

The development of the nematode *Caenorhabditis elegans* provides a system in which to identify and interpret the precise cellular consequences of mutations. *C. elegans* development is largely invariant and the cell lineage is entirely known (Sulston

and Horvitz, 1977; Kimble and Hirsh, 1979; Sulston et al., 1983). Consequently, if the fate of a particular cell is disrupted in a mutant animal, the cell lineage can indicate both what the cell should normally become and what it has become instead. The cell lineage can thus provide a precise in vivo description of the cell fate transformations that result from disruption of a particular gene.

Here we describe the genetic and molecular characterization of the *egl-38* gene in *C. elegans*. We have focused on this gene because of its roles in developmental patterning. We show that it is an essential gene that plays distinct roles in the development of the hermaphrodite egg-laying system and the development of the male tail. *egl-38* thus provides the first genetic insight into two specific patterning events that occur during *C. elegans* development. We have molecularly cloned *egl-38* and show that it is a *PAX* gene most similar to the mammalian class that includes *PAX2*, *PAX5* and *PAX8*. *egl-38* is the first genetically characterized invertebrate member of this subclass of *PAX* genes and provides the opportunity to investigate the functions of these proteins in vivo with single-cell resolution.

MATERIALS AND METHODS

Genetic strains
Nematode strains were cultured according to standard techniques (Brenner, 1974; Sulston and Hodgkin, 1988). Mutations used are described by Hodgkin et al. (1988), and as noted.

LG IV: *dpy-13(e184)*. *dpy-20(e1282)*. *egl-38(n578)* (Trent et al., 1983), *egl-38(sy294)* (this work), *egl-38(s1775)* (Clark, 1990; this work). *unc-22(s7)*. *unc-24(e138)*. *unc-31(e169)*. *eDf19*. *mDf7*. *nT1*.

LG V: *him-5(e1490)*.

Recovery of *egl-38* mutations
egl-38(sy294) was recovered in a genetic screen of over 25,000 mutagenized gametes for mutations that disrupt male tail development (H. M. C., unpublished data). *egl-38(s1775)* was identified among a set of 24 *unc-22*-linked lethal mutations recovered by Clark (1990) that were uncovered by *mDf7*, but not *sDf2*.

Analysis of *egl-38* phenotypes
Cell lineage and anatomy
Cell nuclei divisions and morphogenesis were directly observed in live animals under Nomarski optics as described by Sulston and Horvitz (1977).

Genetic dosage tests
To control for possible maternal contribution to *egl-38* phenotypes, all data in Tables 1, 2 and 3 are from mutant animals derived from heterozygous parents. In Table 3, genetic methods were utilized to count the number of predicted animals of *egl-38* genotype. In many cases, the total number of observed Egl-38 animals was less than the total predicted. The 'missing' animals were assigned to the class of Late emb./L1 lethal. This is because the earliest arrest phenotype of directly observed *egl-38* mutant embryos and early larvae was rupture during or subsequent to embryonic elongation, or immediately after hatching. These ruptured individuals generally disintegrated rapidly.

Genetic mapping of *egl-38*
Trent et al. (1983) genetically mapped *egl-38* to LG IV, between *unc-8* and *dpy-20*. We further delineated its genetic position using deficiency tests and multipoint mapping. Map data have been submitted to the *C. elegans* database, ACeDB. Our map data place *egl-38* within *eDf19* and *mDf7*, to the right of *elt-1* and to the left of *daf-14*.

Germline transformation and rescue of *egl-38* mutant phenotypes
DNAs were microinjected into the mitotic germline of hermaphrodites according to the method of Mello et al. (1991). For rescue experiments, 100 ng/µl of plasmid containing the semi-dominant *rol-6(su1006)* allele (pRF4 or pCes1943), was coinjected as a marker into *egl-38(s1775) unc-22(s7) unc-31(e169)/nT1* animals with 1-10 ng/µl of test DNA. Stable lines were tested for rescue by assaying the presence of viable Unc-22 Unc-31 progeny. Some lines were tested for rescue of *egl-38(s1775)/egl-38(sy294) trans*-heterozygous males and hermaphrodites by mating *egl-38(sy294) unc-22(s7) unc-31(e169)/+* males with transgenic hermaphrodites and assaying for the presence of viable Unc-22 Unc-31 cross progeny. No transgene that failed to rescue the lethality of *egl-38(s1775)* homozygotes rescued *egl-38(s1775)/egl-38(sy294) trans*-heterozygotes.

Isolation of *egl-38* genomic and cDNA clones
Cosmids and DNA sequence from the genomic region of *egl-38* were provided by Alan Coulson and the *C. elegans* sequencing consortium. Subclones of these cosmids were cloned into pBluescript (Stratagene) using standard methods (Sambrook et al., 1989). cDNAs for *egl-38* were not identified in three different cDNA libraries (provided by R. Barstead and P. Okkema). We determined the structure of *egl-38* RNA using first-strand cDNA synthesis and the polymerase chain reaction (RT-PCR). RNA from mixed-stage worms was provided by D. Birnby and T. Inoue. We followed the procedures provided in the Gibco/BRL 5′ and 3′ RACE kits, using random hexamers instead of gene-specific primers in the 5′ RACE first strand synthesis reaction. First-strand cDNA was subjected to PCR amplification using (1) primers to internal portions of the *egl-38* gene, derived from genomic sequence, (2) internal primers and the Universal Adapter Primer provided by Gibco/BRL, and (3) internal primers with either the Anchor Primer provided by Gibco/BRL or primers complementary to the *C. elegans* SL1 and SL2 splice leaders (Krause and Hirsh, 1987; Huang and Hirsh, 1989). Several resultant PCR products were directly sequenced by the University of Washington Biochemistry sequencing facility using an ABI PRISM sequencer (Perkin-Elmer). The PCR products identified two classes of *egl-38* transcripts (Fig. 4A), with two sites for SL1 splicing at the 5′ end of the longer transcript. These two sites for SL1 splicing are 11 nucleotides apart and do not result in changes to the coding region. The longer transcript is reported in Fig. 4B and submitted to the GenBank database (accession number AF002702).

Sequence of *egl-38* mutations
We used PCR to amplify DNA including the first exon (for *s1775*) and the paired domain (exons 2-5) of *egl-38* from mutant animals according to the single worm PCR method of Barstead et al. (1991). PCR products were directly sequenced as above. Both strands of DNA were sequenced to confirm the mutations and the DNA from at least two independent PCR reactions was sequenced.

RESULTS

Mutations in *egl-38* disrupt the fate of specific cells during *C. elegans* development
Three alleles of *egl-38* were each originally identified based on a different phenotype: E̲gg-laying defect (Egl; *n578* (Trent et al., 1983)), M̲ale ab̲normal development (Mab; *sy294* (this work)) and L̲ethality (Let; *s1775* (Clark, 1990; this work)) (Figs 1-3). We have investigated the cellular defects underlying these three phenotypes using direct observation of developing animals and cell lineage analysis.

egl-38 was identified as playing a role in hermaphrodite egg-laying by Trent et al. (1983), who found that the egg-laying
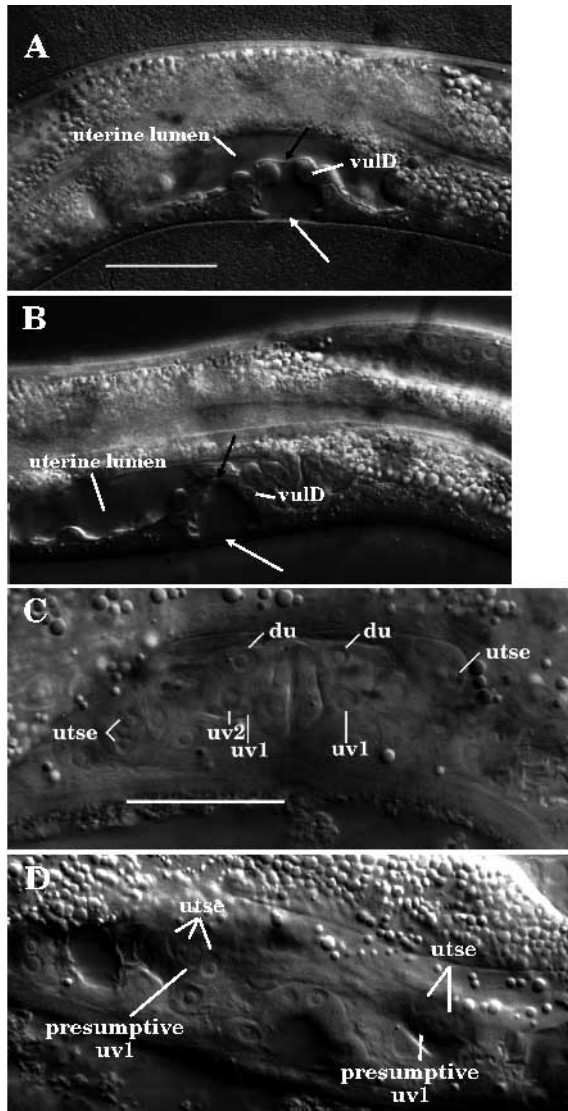
Fig. 1. Nomarski photomicrographs illustrating the cellular and morphological defects in the egg-laying system of L4 *egl-38* mutants. (A,B) Vulval morphogenesis and the vulval-uterine connection. (A) In wild-type, a thin laminar process (black arrow) separates uterine lumen from vulval lumen (white arrow). (B) In *egl-38(n578)* the tissue is thick and the vulval cells are not separated. vulD is one of two nuclei (P5.ppp, P7.paa) that exemplify the mispositioned vulval cells. (C,D) Hermaphrodite ventral uterus. (C) Wild-type. Three utse nuclei have migrated to the posterior (two are visible), one utse nucleus has migrated to the anterior and both uv1 cells have remained proximal to the vulva. The nuclei of two other uterine cell types that remain proximal to the vulva (du, uv2) are also indicated. (D) *egl-38(n578)* (medial plane relative to C). The presumptive uv1 cells have adopted the utse fate and migrated distally rather than attached to the vulva. Posterior right, dorsal up. Scale bar, 20 μm.

defect was not neuronal in origin and noted that there was a vulval morphogenesis defect. We have further analyzed the egg-laying system in *egl-38(n578)* mutant hermaphrodites and find that there are defects in the vulva and the uterus, and their subsequent connection during the L4 stage. In wild-type

animals, the cells derived from the 1° and 2° vulval lineages invaginate and form an arch during the L4 stage (Fig. 1A). The vulval nuclei anterior and posterior to the anchor cell then separate slightly. In many *egl-38(n578)* homozygotes, the vulval arch is misshapen and the nuclei are often mispositioned (Fig. 1B). The vulval cells most proximal to the anchor cell invaginate to the greatest extent and it is the position of these cells that is most abnormal in *egl-38* mutants.

Coincident with these vulval morphogenesis defects, the identity of uv1 ventral uterine cells is disrupted in *egl-38(n578)* mutants. In wild-type animals, there are four uv1 cells (left and right anterior to the prospective vulval opening, and left and right posterior). Each uv1 cell extends a process that contacts the vulval cells most proximal to the anchor cell, resulting in the attachment of the uterus to the vulva (Newman et al., 1996). During development, the uv1 nuclei remain in close proximity to the corresponding vulval nuclei and are located in characteristic positions in mid to late L4 stage. The nuclei of eight other neighboring cells, which include the sisters of the uv1 cells, normally migrate to positions distal to the prospective vulval opening (Fig. 1C). These cells subsequently fuse together with the anchor cell to form the multinucleate uterine seam cell (utse). In *egl-38(n578)* mutants, the presumptive uv1 nuclei behave like utse nuclei and migrate to positions distal to the vulval opening (Fig. 1D). We found that, in 8 of 13 animal sides followed, both the anterior and the posterior presumptive uv1 nuclei migrated like utse cells; in 4 of 13 one of the two nuclei migrated, and in 1 of 13 both did not migrate (were wild-type).

The utse is an H-shaped cell, with the nuclei on the left side and the right side of the animal connected by a bridge of cytoplasm that crosses the animal at the vulva. Morphologically, the utse cytoplasmic bridge is apparent as a thin laminar process separating the uterine and vulval lumina (Fig. 1A). In *egl-38(n578)*, the vulval cells either do not separate or what appears to be the thin laminar process dorsal to the vulva is too thick (Fig. 1B). The earlier defects in the egg-laying system suggest that this connection defect could arise from the absence of the uv1 cells, excess nuclei in the utse or the abnormal vulval cells, or a combination of all three.

The abnormal male phenotype results from specific defects in fate specification in the rectal epithelium of the tail. Many cells of the sex-specific structures in the male tail are the progeny of four cells that make up the posterior rectal epithelium: F, U, B and Y (Fig. 2C). These cells are present in both hermaphrodites and males, but divide postembryonically only in males. We followed the development of these cells throughout L2 and early L3 larval stages in five *egl-38(sy294)* mutant males. Normally, during this period, the two anterior cells F and U each divide a single time, producing left/right sister cells that are equivalent in size and fate. In contrast, the two posterior cells, B and Y, divide several times (Sulston and Horvitz, 1977). We find that, in the *egl-38* mutant animals, the identity of U, and possibly F, is disrupted. In 5 of 5 animals, the presumptive U cell divided earlier than normal, from 5-30 minutes after the division of Y.p (normally U divides about 60 minutes after Y.p; the Y cell developed normally in the mutants). In 3 of 5 animals, the U cell also divided to produce additional progeny during L2. Although this defect was variable, the timing and axes of divisions were similar to those normally produced in the Y lineage. A cell lineage from an
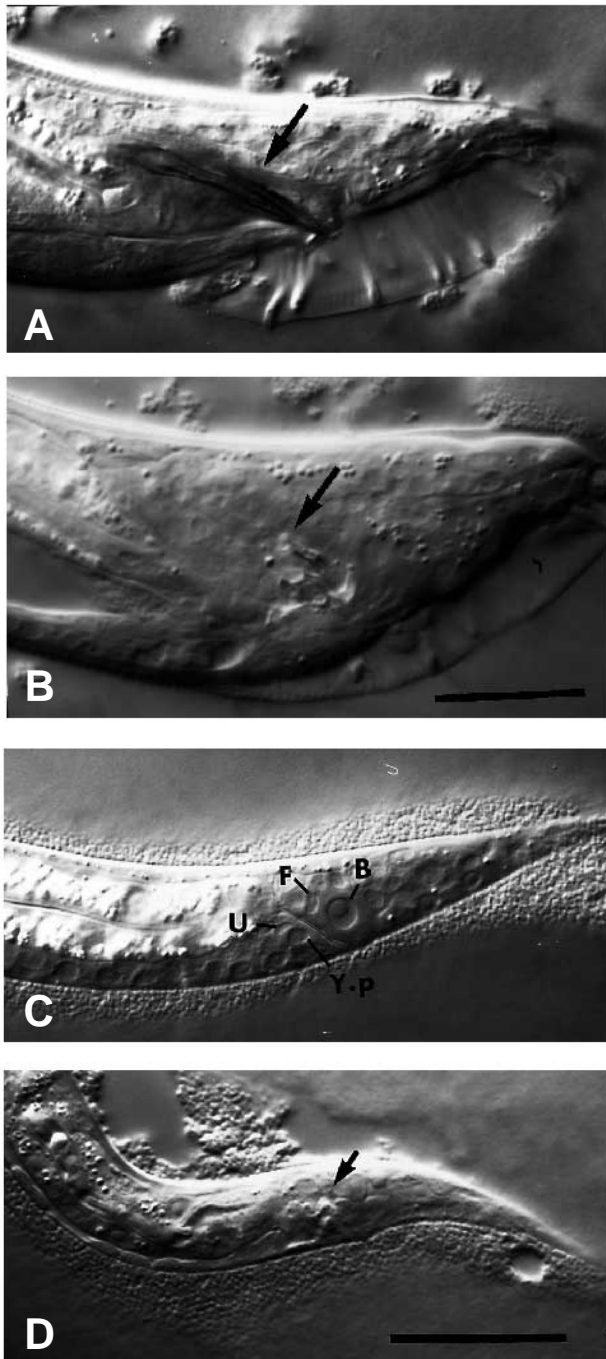
Fig. 2. Nomarski photomicrographs illustrating the cellular and morphological defects in the tail of *egl-38* mutants. (A,B) Adult male tail. (A) Wild-type males have long straight spicules. (B) In the *egl-38* mutant, there is only a small clump of spicule material. (C,D) L1 larva tail morphology. At hatching, the cells F, U, B and Y constitute the posterior rectal epithelium in both males and hermaphrodites. In males only, these four cells undergo postembryonic divisions and are hence called male-specific blast cells. (C) An L1 wild-type male, illustrating the positions of the cells F, U, B and Y.p (the posterior daughter of Y, which occupies the position of Y). The U cell is partially obscured by Y.p. (D) An L1 *egl-38(s1775)* mutant hermaphrodite. The arrow indicates approximate normal position for F and U. To best illustrate the four male-specific blast cells, the wild-type animal is older than the *egl-38* mutant animal. Anterior left, dorsal up. Scale bar, 20 μm.

animal that illustrates this transformation is in Fig. 3B. In this animal, the divisions of the presumptive U cell were similar to those of the posterior daughter of a normal Y cell. In addition, the progeny cells were morphologically similar to Y.p progeny cells (data not shown). To test whether the presumptive U cell divides like Y earlier in development, we followed the development of two additional mutants through the L1 stage. In these animals, the presumptive U cell failed to divide, providing no evidence that U behaves like Y prior to the L2 stage. The development of F is also disrupted in *egl-38(sy294)* mutant males. In 4 of 4 animals, the axis of F cell division was skewed to follow the anterior/posterior or dorsal/ventral axis. In two of these animal, cytokinesis was also unequal, producing one larger and one smaller daughter cell. An initial unequal cytokinesis is a hallmark of B, the posterior neighbor of F. However, the presumptive F cell did not divide further in L2 as B normally does. In summary, *egl-38(sy294)* mutant males have variable cell lineage defects affecting the cells U and F. We propose that *egl-38* plays a role in specifying the identity of U, and possibly F, and that in *egl-38* mutants U adopts the fate of its posterior neighbor, Y.

The development of the B cell is also disrupted in *egl-38(sy294)* mutant males. Normally the B cell develops in three steps. (1) An initial set of divisions produces ten progeny. (2) Eight of these progeny (the progeny of B.a) then participate in short-range migrations. (3) All ten progeny then undergo a second set of cell divisions (Sulston et al., 1980). In 5 of 5 *egl-38(sy294)* mutant males, the initial divisions of B were essentially normal. However, the axes of division of B.a and its progeny were frequently skewed, and the eight progeny cells failed to migrate properly. Cell ablation experiments have shown that these eight B progeny cells require F, U and Y cells for normal development, and that they also require proper positional orientation relative to each other (Chamberlin and Sternberg, 1993). Thus the abnormal development of B may result, at least in part, from the disruption of F and U. Since the B cell is the precursor to all of the cells of the spicules, this B cell lineage defect can account for the abnormal spicule morphology in the adult (Fig. 2B).

egl-38 is essential for the viability of both males and hermaphrodites. Genetic tests indicate that the arrest stage for strong *egl-38* mutants is around hatching. To better understand the causes of *egl-38* mutant lethality, we looked at dying *egl-38(sy294)/eDf19*, *egl-38(s1775)/mDf7* and *egl-38(s1775)* homozygous animals (all animals of these genotypes die; see Table 3). Mutant embryos generally elongate and are grossly normal morphologically. Subsequent to elongation, however, they tend to rupture and quickly disintegrate. The rupture can initiate at the tail, but also at the mouth, or along the ventral midline. Animals that do hatch have heavy tissue damage at the rectum (Fig. 2D) and the larvae soon disintegrate. The localized tail damage in these dying animals suggests that at least part of the lethality in *egl-38* mutants results from defects in the rectal epithelium, consistent with the cell lineage defects observed in the viable *egl-38* males. However, we can not rule out the possibility that other developmental defects also contribute to the lethality of mutant animals. In less severe *egl-38* genotypes, animals that escape lethality at hatching may still die later in larval development. The later larval lethality reflects that seen in L1; the animals develop severe tissue damage at the rectum and generally die at the larval molts. In addition,
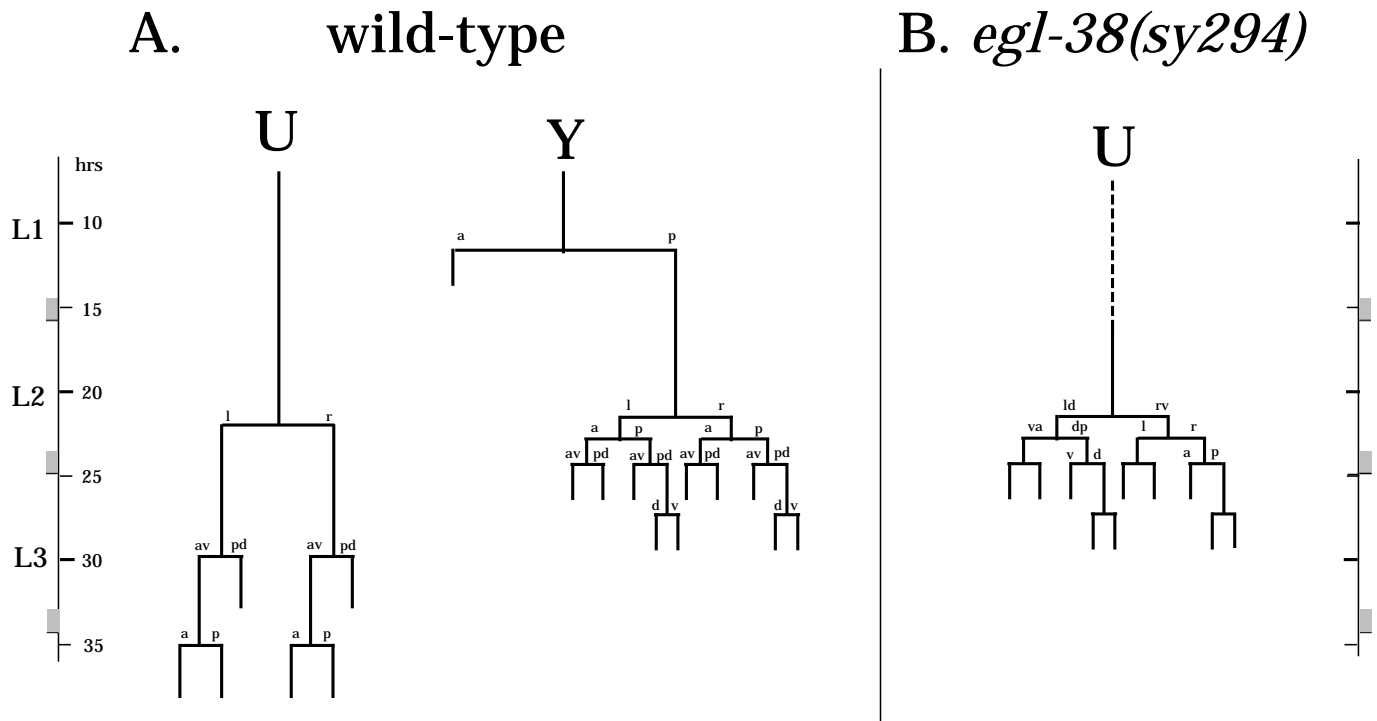
A.   wild-type   B. *egl-38(sy294)*

Fig. 3. U and Y cell lineage in a wild-type male (A) and cell lineage of the presumptive U cell in an *egl-38(sy294)* mutant male (B). Vertical lines represent cells, horizontal lines represent cell divisions. Vertical axis scale bar shows larval stage (lethargus indicated by shaded box) and developmental time after hatching. Division axes are shown: anterior (a), posterior (p), left (l), right (r), dorsal (d) and ventral (v). Dotted line indicates inferred lineage, rather than observed. In wild-type, division of U.(l/r)a is variable. Figure after Sulston et al. (1980).

hermaphrodites as well as males that do survive to adulthood can display variable morphological defects and tissue damage in the tail at the rectum.

Different mutations preferentially disrupt different functions of *egl-38*

We have used inter se complementation tests with the three alleles of *egl-38* and deficiencies that span the locus to assay function in egg-laying, male tail development and viability (Tables 1, 2, 3; summarized in Table 4). All three alleles are recessive (data not shown) and all behave as reduction- or loss-

of-function mutations. We find that the two viable alleles of *egl-38* preferentially disrupt different functions and that the three mutations do not form an allelic series. 90% of *n578* homozygous hermaphrodites are Egl and 24% of the males are Mab (Tables 1a, 2a). In contrast, 35% of *sy294* homozygotes are Egl, but 100% are Mab (Tables 1b, 2b). *n578* also disrupts viability to a lesser extent than *sy294*: 86% of *n578* homozygotes survive to adulthood, whereas only 37% of *sy294* homozygotes do (Table 3a,b). However, *n578* and *sy294* do not exhibit intragenic complementation, as 53% of *n578/sy294* hermaphrodites are Egl and 97% of males are Mab (Tables 1c, 2c). In these animals, disruption of male tail development is comparable to *sy294* homozygotes, whereas disruption of egg-laying ability is intermediate between *n578* and *sy294* homozygotes. This could reflect the possibility that *n578* retains more

Table 1. Egg-laying function of *egl-38*

egl-38 genotype	Egl (%)	n
a. *n578/n578*	90	100
b. *sy294/sy294*	35	63
c. *n578/sy294*	53	72
d. *n578/s1775*	99	131
e. *n578/mDf7*	100	72

L4 animals were placed on separate plates and scored for ability to lay eggs onto the plate over a 3- to 4-day period. Egl (%) indicates the percentage of animals that were entirely egg-laying defective. *n* indicates total number of animals tested for each genotype.
Full genotypes:
a. *egl-38(n578) dpy-20(e1282)*.
b. *egl-38(sy294) dpy-20(e1282)*.
c. *unc-24(e138) egl-38(n578)/egl-38(sy294) dpy-20(e1282)*.
d. *egl-38(n578) unc-22(s7) unc-31(e169)/egl-38(s1775) unc-22(s7) unc-31(e169)*.
e. *egl-38(n578) unc-22(s7) unc-31(e169)/dpy-13(e184) mDf7*.

Table 2. Male tail function of *egl-38*

egl-38 genotype	Mab (%)	n
a. *n578/n578*	24	71
b. *sy294/sy294*	100	71
c. *n578/sy294*	97	123
d. *n578/s1775*	100	87

Adult males were scored using Nomarski optics. Mab (%) indicates percentage of adult males with abnormal tail morphology. *n* indicates total number of animals scored for each genotype.
Full genotypes:
a. *egl-38(n578) dpy-20(e1282)*.
b. *egl-38(sy294) dpy-20(e128)*.
c. *egl-38(n578) dpy-20(e1282)/egl-38(sy294) dpy-20(e1282)*.
d. *egl-38(n578) dpy-20(e1282)/egl-38(s1775) dpy-20(e1282)*.

Table 3. Essential function of *egl-38*

egl-38 genotype	Arrest stage			
	Late emb./L1 (%)	Larva (%)	Viable (%)	*n*
a. *n578/n578*	14	0	86	182
b. *sy294/sy294*	38	25	37	266
c. *s1775/s1775*	100	0	0	218
d. *n578/s1775*	0	5	95	283
e. *n578/mDf7*	17	11	72	263
f. *sy294/s1775*	90	7	3	194
g. *sy294/eDf19*	97	3	0	289
h. *sy294/mDf7*	100	0	0	83
i. *s1775/eDf19*	100	0	0	192

Both male and hermaphrodite *C. elegans* require *egl-38* for viability and we observed no differences between the two sexes. For a.-c., progeny in all classes from the broods of heterozygous animals bearing a *cis*-marked *egl-38* allele balanced by a *trans* marker chromosome were counted and scored for phenotype. The total predicted *egl-38* progeny (*n*) was estimated by counting the total siblings homozygous for the *trans* marker chromosome. For d.-i., heterozygous animals bearing *cis*-marked *egl-38* alleles were mated with each other. Phenotypically wild-type males were counted to estimate the total predicted *egl-38* cross progeny (*n*). *egl-38* cross progeny are 2/3 of the total number of phenotypically wild-type male progeny. Late emb./L1 indicates percentage of animals that died as late embryos or early larvae. This number includes animals directly observed as well as animals calculated to have died. Larva indicates percentage of animals observed to arrest as larvae. Viable indicates percentage of animals that survived to adulthood.

Full genotypes:

a. *egl-38(n578) unc-22(s7) unc-31(e169)*.

b. *egl-38(sy294) unc-22(s7) unc-31(e169)*.

c. *egl-38(s1775) unc-22(s7) unc-31(e169)*.

d. *egl-38(n578) unc-22(s7) unc-31(e169)/egl-38(s1775) unc-22(s7) unc-31(e169)*.

e. *egl-38(n578) unc-22(s7) unc-31(e169)/dpy-13(e184) mDf7*.

f. *egl-38(sy294) unc-22(s7) unc-31(e169)/egl-38(s1775) unc-22(s7) unc-31(e169)*.

g. *egl-38(sy294) unc-22(s7) unc-31(e169)/eDf19*.

h. *egl-38(sy294) unc-22(s7) unc-31(e169)/dpy-13(e184) mDf7*.

i. *egl-38(s1775) unc-22(s7) unc-31(e169)/eDf19*.

egg-laying function than *sy294* retains male tail function (*n578* homozygotes are 90% Egl, whereas *sy294* homozygotes are 100% Mab; Tables 1a, 2b), or the possibility that the tail function is more sensitive to a general reduction of *egl-38* activity than is the egg-laying function.

The lethal mutation, *s1775*, strongly disrupts all three functions. It is homozygous lethal, and fails to complement the other alleles for egg-laying and male tail development (Tables 3c, 1d, 2d). Genetically, *s1775* is a strong loss-of-function mutation, as it behaves similar to a deficiency when tested in *trans* to the other alleles. One exception is for viability. Genetic deficiencies enhance the lethality of both *n578* and *sy294* to a greater extent than *s1775*, suggesting that *s1775* retains some residual *egl-38* activity. Alternatively, since deficiencies remove many genes, it is possible that this enhancement by the deficiencies reflects reduced viability due to haploinsufficiency of other loci included in the deficiency. Data with two different deficiencies indicate that this phenomenon does occur in the *egl-38* region: *mDf7* enhances the lethality of *sy294* more than *eDf19* does, even though both deficiencies delete *egl-38* (Table 3g,h).

Positional cloning of *egl-38*

We genetically mapped *egl-38* to linkage group IV, between *elt-1* and *daf-14*, which provided a left-most and right-most

Table 4. Summary of the functions disrupted by different *egl-38* mutations

egl-38 function	*egl-38* genotype			
	+	*n578*	*sy294*	*s1775*
egg-laying	+++	−	++	−
male tail	+++	++	−	−
viability	+++	+++	+	−

n578 disrupts *egl-38* function in the egg-laying system to a much greater extent than in male tail development or viability. *sy294* disrupts male tail development and viability to a greater extent than egg-laying. *s1775* strongly disrupts all three *egl-38* functions. +++ indicates >80% of function retained. ++ indicates 50-80% of function retained. + indicates 10-50% of function retained. − indicates <10% of function retained. Percentages estimated based on homozygous and transheterozygous phenotypes.

boundary for *egl-38* on the physical map. We tested cosmids in the region between these two genes for rescue of *egl-38* and identified two cosmids (C35E4 and ZK1258) that were capable of rescuing all three phenotypes associated with *egl-38* mutations. The overlap of the two cosmids defines an approximately 20 kb region of genomic DNA. Analysis of the sequence of DNA included in the 20 kb overlap with GENEFINDER software identified four candidate genes (Wilson et al., 1994). We were unable to further narrow down the genomic region, since subclones of C35E4 did not rescue *egl-38* mutants. However, DNA sequence of one gene in the overlap region (C04G2.7) from *egl-38* mutant animals identified mutations that indicate C04G2.7 is *egl-38* (see below).

egl-38 encodes a member of the PAX family of transcription factors

To determine the structure of *egl-38* mRNAs, we identified corresponding cDNAs using first-strand cDNA synthesis and the polymerase chain reaction (RT-PCR). We sequenced cDNA representing two classes of *egl-38* transcripts (Fig. 4A). The sequence and translation of the longer class is shown in Fig. 4B. This sequence can be translated to produce a 289 amino acid protein. The sequence and architecture of the predicted *egl-38* protein, EGL-38, is most similar to the mammalian class of PAX proteins that includes PAX2, PAX5 and PAX8. Like these mammalian proteins, EGL-38 contains a paired domain and an octapeptide sequence. Within the paired domain, EGL-38 and the human PAX proteins are 72-74% identical (Fig. 5). EGL-38 does not contain the partial homeodomain sequence found in the three mammalian proteins. However, this partial domain varies considerably among PAX proteins and differs greatly from the standard homeodomain consensus sequence, suggesting it is not under the same selective restraints as the paired domain and octapeptide sequence (Sun et al., 1997). The shorter transcript of *egl-38* results from splicing of the *C. elegans trans*-splice leader SL1 to the second exon of *egl-38*, eliminating the predicted start ATG. In this variant, the first ATG is 29 codons into the paired domain. We did not assess the relative abundance of these two mRNA species. However, we did not identify cDNA for *egl-38* in screens of three different cDNA libraries (see Materials and Methods), suggesting that *egl-38* mRNAs are rare.

We sequenced PCR-amplified genomic DNA from *egl-38* mutant animals to determine the molecular changes associated with *egl-38* mutations. The lethal mutation, *s1775*, disrupts the

splice donor site in the first intron, presumably resulting in abnormal splicing (Fig. 5B). If this unspliced RNA is translated, the resultant protein would contain the first twenty-one amino acids of EGL-38, followed by nine amino acids coded by the intron prior to the first in-frame stop codon. We predict such a protein would retain no *egl-38* activity, consistent with the gene dosage analysis. However, we have not analyzed RNA from *egl-38(s1775)* mutant animals to confirm the splicing consequences of this mutation. The two viable *egl-38* mutations are each missense mutations that change invariant glycines within the paired domain (Fig. 5A). *sy294* is a G→A transition that results in a GLY33→VAL substitution (the number indicates the position within the paired domain). *n578* is a G→A transition that results in a GLY69→GLU substitution.

DISCUSSION

The developmental logic underlying *egl-38* function

We have shown that *egl-38* plays a role in the development of *C. elegans* hermaphrodite egg-laying system and the rectal epithelium. In the hermaphrodite uterus, *egl-38* is necessary to make the presumptive uv1 cells different from their sister and neighboring cells, utse. *egl-38* is also required for normal vulval and uterine morphogenesis. Cell lineage analysis in males indicates that *egl-38* is necessary for the identity of the U cell and possibly the F cell in the rectal epithelium, and to make U different from its more posterior neighbor, Y. These fate transformations in the tail may also contribute to the lethality associated with *egl-38* mutations in both males and hermaphrodites. The precise cellular defects in *egl-38* mutants provide information about the underlying pattern formation in the two affected tissues.

Mutations in *egl-38* disrupt developmental patterning in the hermaphrodite uterus but also disrupt vulval morphogenesis. Since our current analysis does not establish a site of action for *egl-38*, the developmental defects in the uterus and vulva could each result directly from mutant *egl-38* or one defect could be a secondary conse-
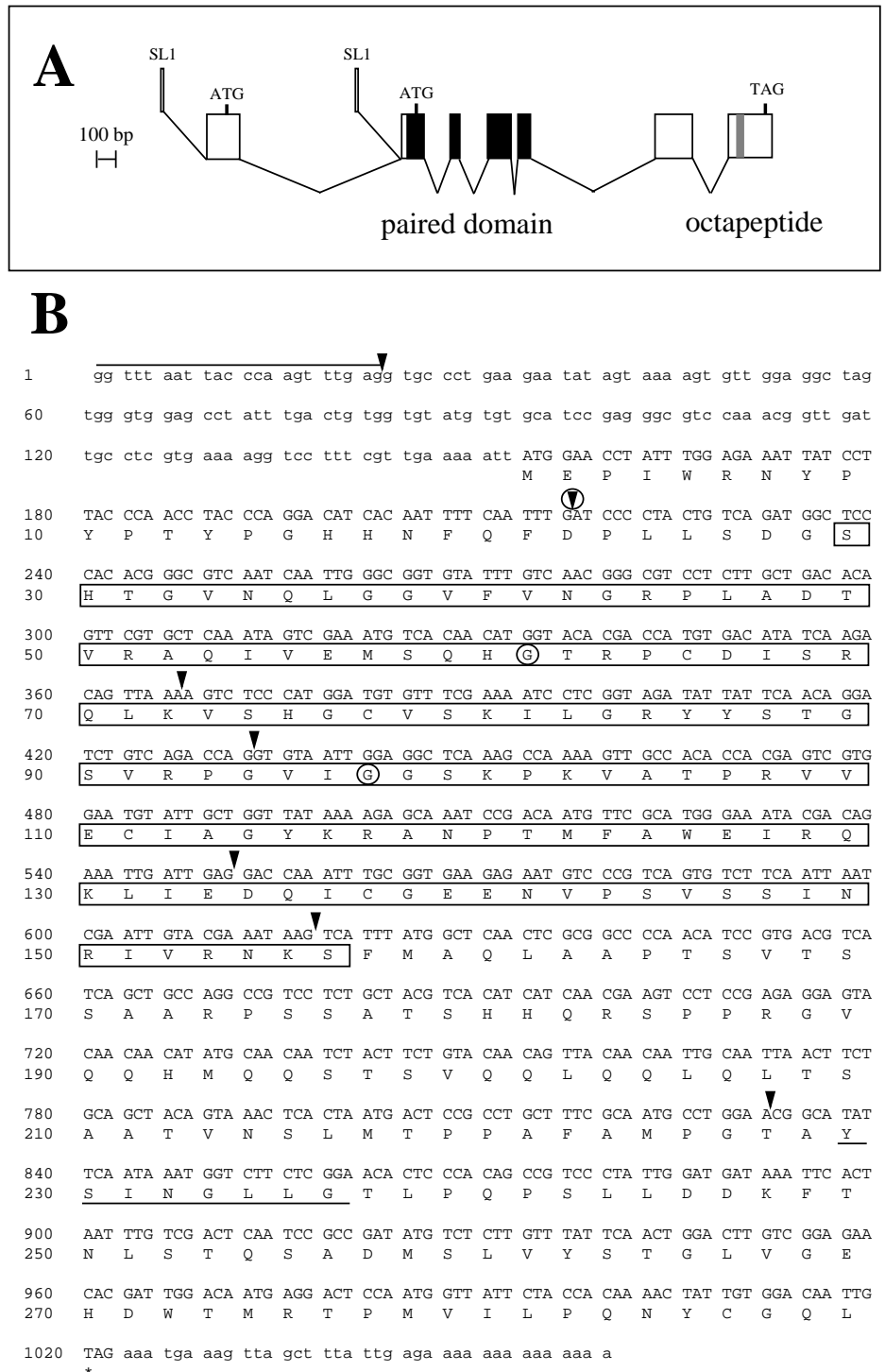

Fig. 4. (A) Genomic structure of *egl-38*. We identified two classes of *egl-38* transcripts. The larger class has seven exons that can be translated to produce a protein of 289 amino acids. The position of regions that code for sequence homology domains are indicated: the paired domain in black, the octapeptide domain in gray. The smaller transcript results from the *C. elegans* SL1 splice leader *trans*-spliced to the second exon. Conceptual translation of this transcript results in a truncated protein that lacks the amino terminus of the larger protein, including the first 28 amino acids of the paired domain. (B) Sequence of cDNA and potential amino acid translation for the longer *egl-38* transcript. The SL1 splice leader sequence is overlined. The paired domain is boxed. The octapeptide sequence is underlined. Splice boundaries are indicated by triangles. Amino acids and the splice site affected by *egl-38* mutations are circled. Sequence has been submitted to GenBank database, accession number AF002702.

quence of the other. Nevertheless, the cell fate transformation in *egl-38* mutant uteri underscores the progressive patterning of the ventral uterus. Two developmental steps can account for the specialization of ventral uterine cell types. During normal development, six ventral uterine cells, termed π cells, are first induced by the anchor cell to adopt fates different from sibling and neighbor cells in a cell interaction mediated by the gene *lin-12* (Newman et al., 1995). It is these π cells that are the precursors to both the uv1 and the utse cells. Each π cell then divides a single time and the four progeny proximal to the forming vulva differentiate as the uv1 cells, whereas the remaining eight cells differentiate as utse cells (Kimble and Hirsh, 1979; Newman et al., 1996). Since *egl-38* plays a role in making uv1 different from utse cells, it provides the first genetic insight into this final step of cell specialization in the ventral uterus.

The *egl-38* mutant phenotype also provides insight into the development of the rectal epithelium. The four male-specific blast cells that make up the posterior rectal epithelium normally each adopt a different fate (see Fig. 2C). These cells share some lineal relationships: F and U are embryonic sister cells, and B and U are left/right lineal homologues. Analysis of *egl-38* indicates that the function of a single gene distinguishes U from Y. Since Y is a neighbor, but not a lineal relative of U, a simple interpretation is that *egl-38* participates in making equipotential neighboring cells different from each other. The gene *mab-9* has also been identified to play a role in the development of these male-specific rectal epithelial cells. Chisholm and Hodgkin (1989) have shown that *mab-9* is required for the identity of the two dorsal cells, F and B, and that, in *mab-9* mutants, these cells adopt the fate normally associated with their more ventral neighbors, U and Y, respectively. It is not known whether *egl-38* and *mab-9* are required within the affected cells, or whether they influence the fates of these rectal epithelial cells in a non-autonomous manner. Nevertheless, these two genes provide the means of establishing the four distinct fates. An attractive combinatorial model is that *egl-38* functions to influence the fate of the two anterior cells (F, U) to make them different from their posterior neighbors (B, Y), and *mab-9* functions to influence the fate of the two dorsal cells (F, B) to make them different from their ventral neighbors (U, Y). Double mutant analysis is consistent with this model. *mab-9; egl-38* animals have defects in F, U and B that are consistent with an additive effect of each mutation (H. M. C., unpublished).

EGL-38 and PAX protein function

egl-38 encodes a protein that contains

a DNA-binding paired domain. The paired domain is composed of two helix-turn-helix motifs that can have distinct binding properties (Czerny et al., 1993; Xu et al., 1995; Jun and Desplan, 1996). Experimentally, PAX proteins can require both motifs, only the amino-terminal motif, or only the carboxyl-terminal motif for maximal DNA binding, depending on the particular PAX protein and target DNA (Czerny et al., 1993; Epstein et al., 1994; Bertuccioli et al, 1996). We have identified cDNA for two classes of *egl-38* transcripts. The predicted product from the longer transcript includes a complete paired domain, whereas the shorter transcript would produce a protein with a disrupted amino-terminal portion of the paired domain. How do these two transcripts relate to the functions of *egl-38*? The functional importance of the longer transcript is underscored by the fact that the *s1775* mutation that disrupts splicing of the first exon also disrupts all identified *egl-38* functions. In contrast, our genetic analysis has not suggested a function for the shorter transcript. If it has a function, either (1) the *s1775* mutation can also affect its structure or production in some way, or (2) it has a function not identified in our genetic analysis, or (3) its function is dependent on functional product from the longer transcript. For example, either the RNA or the protein product might play a negative regulatory role by interfering or competing with the longer product. However, our current results are consistent with the longer transcript mediating all the functions of *egl-38*.

Fig. 5. (A) Alignment of the paired domain from predicted EGL-38 and other PAX proteins. EGL-38 is more similar to human PAX2 (Eccles et al., 1992) than the *C. elegans* PAX6 homologue, VAB-3 (Chisholm and Horvitz, 1995; Zhang and Emmons, 1995). Sequence of Drosophila Paired (Prd) is also included (Bopp et al., 1986). Arrows indicate substitutions found in DNA from *egl-38* mutants. (B) The *egl-38(s1775)* mutation. A G→A transition mutation in DNA from *egl-38(s1775)* animals results in the disruption of the splice donor site within intron 1. (Top) The *C. elegans* splice donor consensus sequence. The GU of the intron are invariant (in bold; from Krause, 1995). (Center) The splice donor site in *egl-38(+)*. (Bottom) The defect in *egl-38(s1775)* mutants.

The role of *egl-38* in development

Analogous functions in different tissues

Mutations in *egl-38* disrupt the development of several cells and cell types. Although it is not clear whether *egl-38* function is required autonomously or non-autonomously in the affected cells, the rectal epithelium and the egg-laying system are quite developmentally distinct. The cells of the somatic gonad are mesodermal in origin, whereas the cells of the vulva and the rectal epithelium are epidermal. *egl-38* functions relatively early in development to specify blast cell identity in the male tail, whereas its role in the hermaphrodite egg-laying system is later in development, where it affects the terminal differentiation of cells. Why is a single gene (*egl-38*) involved in the development of these two distinct structures? Despite differences in tissue type, there is similarity in what the cells affected in *egl-38* mutants normally do. F and U in the rectum and the vulval and uv1 cells of the egg-laying system are part of structures that connect the inside of the worm to the outside. There may be specialized functions shared by both sets of cells that permit pore formation or maintenance of structural integrity. This function may also be related to those of other genes in the *PAX2/PAX5/PAX8* class. In *Pax2* mouse mutants, the ureters and both male and female genital tracts fail to develop properly (Torres et al., 1995). These structures normally function to connect the kidneys and gonads to the cloaca. Thus, despite the differences in species and tissue type, murine *Pax2* also plays a role in establishing the fate of cells that normally form ducts or pores.

Disparate functions in different tissues

Our genetic analysis of *egl-38* indicates that this gene has qualitatively distinct activities affecting at least two distinct sets of cells (Table 4). In particular, the three mutations in *egl-38* do not form an allelic series. Both genetic and molecular data suggest that *s1775* is a mutation that either strongly or completely eliminates *egl-38* function. The two viable alleles cause a reduction of *egl-38* function. Each viable allele was originally isolated based on a different phenotype and genetic tests indicate that they preferentially disrupt different functions of *egl-38*. Both are missense mutations affecting the paired domain. The *sy294* mutation is a substitution at an invariant glycine at the end of the first alpha helix of the paired domain (Xu et al., 1995) and thus could result in a destabilization or disruption of the folded protein. The developmental consequences of this mutation are a complete disruption of male tail development, a disruption of viability and a mild disruption of egg-laying function. In contrast, the *n578* mutation is a substitution at an invariant glycine that normally contacts the minor groove of the DNA (Xu et al., 1995). The consequences of this mutation are a strong disruption of egg-laying function, with only minor effects on viability and male tail development. These results suggest that there are differences between the targets for EGL-38 in the egg-laying system and in the tail. This may reflect differences in the relative importance of different DNA targets. Although the *n578* mutation has not been tested in vitro, tests of other point mutations generated in the PAX5 paired domain indicate that single substitutions can restrict or alter the potential DNA targets of the paired domain (Czerny et al, 1993; Czerny and Busslinger, 1995). Our genetic analysis suggests that at least the downstream genetic targets of *egl-38* differ: *egl-38* influences the terminal specialization of cells in the egg-laying system, whereas it influences subsequent cell divisions in the male tail. The *egl-38(n578)* mutation indicates that these downstream differences may be in the direct DNA targets for the EGL-38 protein.

REFERENCES

Adams, B., Dorfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurer-Fogy, I. and Busslinger, M. (1992). *Pax-5* encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev.* **6**, 1589-1607.

Barstead, R. J., Kleiman, L. and Waterston, R. H. (1991). Cloning, sequencing and mapping of an alpha-actinin gene from the nematode *Caenorhabditis elegans. Cell Motil. Cytoskel.* **20**, 69-78.

Bertuccioli, C., Fasano, L., Jun, S., Wang, S., Sheng, G. and Desplan, C. (1996). In vivo requirement for the paired domain and homeodomain of the paired segmentation gene product. *Development* **122**, 2673-2685.

Bopp, D., Burri M., Baumgartner S., Frigerior G. and Noll M. (1986). Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of Drosophila. *Cell* **47**, 1033-1040.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans. Genetics* **77**, 71-94.

Chamberlin, H. M. and Sternberg, P. W. (1993). Multiple cell interactions are required for fate specification during male spicule development in *Caenorhabditis elegans. Development* **118**, 297-324.

Chisholm, A. D. and Hodgkin, J. (1989). The *mab-9* gene controls the fate of B, the major male-specific blast cell in the tail region of *Caenorhabditis elegans. Genes Dev.* **3**, 1413-1423.

Chisholm, A. D. and Horvitz, H. R. (1995). Patterning of the *Caenorhabditis elegans* head region by the *Pax-6* family member *vab-3. Nature* **377**, 52-55.

Clark, D. V. (1990). The *unc-22* IV region of *Caenorhabditis elegans*: genetic analysis and molecular mapping. PhD Thesis, Simon Fraser University, Burnaby, BC, Canada.

Czerny, T., Schaffner, G. and Busslinger, M. (1993). DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev.* **7**, 2048-2061.

Czerny, T. and Busslinger, M. (1995). DNA-binding and transactivation properties of Pax-6: Three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). *Molec. Cell. Biol.* **15**, 2858-2871.

Dressler, G. R., Deutsch, U., Chowdhury, K., Nornes, H. O. and Gruss, P. (1990). *Pax2*, a new murine paired-box-containing gene and its expression in the developing excretory system. *Development* **109**, 787-795.

Eccles, M. R., Wallis, L. J., Fidler, A. E., Spurr, N. K., Goodfellow, P. J. and Reeve, A. E. (1992). Expression of the *PAX2* gene in human fetal kidney and Wilms' Tumor. *Cell Growth Diff.* **3**, 279-289.

Epstein, J. A., Glaser, T., Cai, J., Jepeal, L., Walton, D. S. and Maas, R. L. (1994). Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. *Genes Dev.* **8**, 2022-2034.

Hodgkin, J., Edgley, M., Riddle, D. L. and Albertson, D. G. (1988). Genetics. In *The Nematode Caenorhabditis elegans* (ed. W. Wood), pp. 491-584. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Huang, X.-Y. and Hirsh, D. (1989). A second trans-spliced RNA leader in the

nematode *Caenorhabditis elegans. Proc. Natn. Acad. Sci., USA* **86**, 8640-8644.

Jan, Y. N. and Jan, L. Y. (1995). Maggot's hair and bug's eye: Role of cell interactions and intrinsic factors in cell fate specification. *Neuron* **14**, 1-5.

Jun, S. and Desplan, C. (1996). Cooperative interactions between paired domain and homeodomain. *Development* **122**, 2639-2650.

Kimble, J. E. and Hirsh, D. (1979). Post-embryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans. Dev. Biol.* **70**, 396-417.

Krause M. and Hirsh, D. (1987). A trans-splice leader sequence on actin mRNA in C. elegans. *Cell* **49**, 753-761.

Krause, M. (1995). Transcription and translation. In *Caenorhabditis elegans: Modern biological analysis of an organism* (ed. H.F. Epstein and D.C. Shakes). *Methods in Cell Biology* **48**, 483-512.

Mello, C. C., Kramer, J. M., Stinchcomb, D. and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: Extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959-3970.

Newman, A. P., White, J. G. and Sternberg, P. W. (1995). The *C. elegans lin-12* gene mediates induction of ventral uterine specialization by the anchor cell. *Development* **121**, 263-271.

Newman, A. P., White, J. G. and Sternberg, P. W. (1996). Morphogenesis of the *C. elegans* hermaphrodite uterus. *Development* **122**, 3617-3626.

Noll, M. (1993). Evolution and role of *Pax* genes. *Cur. Op. Genet. Dev.* **3**, 595-605.

Plachov, D., Chowdhury, K., Walther, C., Simon, D., Guenet, J. L. and Gruss, P. (1990). Pax8, a murine paired box gene expressed in the developing excretory system and thyroid gland. *Development* **110**, 643-651.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* 2nd edition. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sulston, J. and Horvitz, H. R. (1977). Postembryonic cell lineages of the nematode *Caenorhabditis elegans. Dev. Biol.* **56**, 110-156.

Sulston, J. E., Albertson, D. G. and Thomson, J. N. (1980). The *Caenorhabditis elegans* male: Postembryonic development of nongonadal structures. *Dev. Biol.* **78**, 542-576.

Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans. Dev. Biol.* **100**, 64-119.

Sulston, J. E. and Hodgkin, J. (1988). Methods. In *The nematode Caenorhabditis elegans* (ed. W. Wood), pp. 587-606. Cold Spring Harbor NY: Cold Spring HarborLaboratory Press.

Sun, H., Rodin, A., Zhou, Y., Dickinson, D. P., Harper, D. E., Hewett-Emmett, D. and Li, W.-H. (1997). Evolution of paired domains: Isolation and sequencing of jellyfish and hydra Pax genes related to *Pax-5* and *Pax-6. Proc. Natn. Acad. Sci., USA* **94**, 5156-5161.

Torres, M., Gomez-Pardo, E., Dressler, G. R. and Gruss, P. (1995). *Pax-2* controls multiple steps during urogenital development. *Development* **121**, 4057-4065.

Torres, M., Gomez-Pardo, E. and Gruss, P. (1996). *Pax2* contributes to inner ear patterning and optic nerve trajectory. *Development* **122**, 3381-3391.

Trent. C., Tsung, N. and Horvitz, H. R. (1983). Egg-laying defective mutants of the nematode *C. elegans. Genetics* **104**, 619-647.

Urbanek, P., Wand, Z-Q., Fetka, I., Wagner, E. F. and Busslinger, M. (1994). Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* **79**, 901-912.

Walther, C., Guenet, J. L., Simon, D., Deutsch, U., Jostes, B., Goulding, M. D., Plachov, D., Balling, R. and Gruss, P. (1991). Pax – A murine multigene family of paired box-containing genes. *Genomics* **11**, 424-434.

Wilson, R. et al. (1994). 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans. Nature* **368**, 32-38.

Xu, W., Rould, M. A., Jun, S., Despain, C. and Pabo, C. O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations. *Cell* **80**, 639-650.

Zhang, Y. and Emmons, S. W. (1995). Specification of sense-organ identity by a *Caenorhabditis elegans Pax-6* homologue. *Nature* **377**, 55-59.