

PES-1 is expressed during early embryogenesis in *Caenorhabditis elegans* and has homology to the fork head family of transcription factors

Ian Allen Hope

Department of Pure and Applied Biology and Department of Genetics, The University of Leeds, Leeds, LS2 9JT, UK

SUMMARY

Promoter trapping has identified a gene, *pes-1*, which is expressed during *C. elegans* embryogenesis. The β -galactosidase expression pattern, directed by the *pes-1/lacZ* fusion through which this gene was cloned, has been determined precisely in terms of the embryonic cell lineage and has three components. One component is in a subset of cells of the AB founder cell lineage during early embryogenesis, suggesting *pes-1* may be regulated both by cell autonomous determinants and by intercellular signals. Analysis of

cdna suggests *pes-1* has two sites for initiation of transcription and the two transcripts would encode related but distinct proteins. The predicted PES-1 proteins have homology to the fork head family of transcription factors and therefore may have important regulatory roles in early embryogenesis.

Key words: *Caenorhabditis elegans*, embryogenesis, promoter trapping, fork head homologue, gene expression

INTRODUCTION

Six years ago a new procedure called enhancer trapping was developed for investigation of animal development (O'Kane and Gehring, 1987). The 'trapping' approach is a strategy of screening developmental patterns of gene expression and identifying regulatory elements important in the developmental control of gene expression. The details of the approach vary between species (e.g. Allen et al., 1988; Gossler et al., 1989), but all depend on fusion of the bacterial *lacZ* reporter gene, encoding the readily assayable enzyme β -galactosidase, to genetic regulatory elements of the species under study. Animals containing these constructions are then stained histochemically to reveal the β -galactosidase expression pattern. Large numbers of fusions are assayed to identify those of interest.

The variation on the trapping approach used for the nematode worm, *Caenorhabditis elegans*, was called promoter trapping (Hope, 1991; Young and Hope, 1993). Random *C. elegans* genomic DNA fragments were ligated into a plasmid upstream of a promoterless *lacZ* gene. Recombinant plasmids containing different genomic DNA inserts were assayed for promoter activity by transformation of *C. elegans* and examination of transformed lines for β -galactosidase expression. Plasmids were assayed in large pools of independent recombinant plasmid constructions because few of the random inserts would be expected to direct β -galactosidase expression. For plasmid pools which gave a β -galactosidase expression pattern of interest, the individual plasmid responsible was identified subsequently.

C. elegans development proceeds with an almost completely invariant cell lineage, which has been fully determined (Sulston and Horvitz, 1977; Kimble and Hirsh, 1979; Sulston

et al., 1983). Patterns of cell division and cell fate are largely identical between individuals of this species. The mechanisms by which *C. elegans* development proceeds is the subject of considerable research but markers of cellular identity for evaluating the consequences of experimental or genetic perturbation of *C. elegans* development consist mainly of terminal differentiation products (Cowan and McIntosh, 1985; Edgar and McGhee, 1988). Key developmental events occur during early embryogenesis prior to any terminal differentiation, yet there are few markers to distinguish cells of the early embryo. Promoter trapping could provide molecular markers that would distinguish cell identities during early embryogenesis. Furthermore, genes expressed in a restricted manner during early embryogenesis could provide subjects for investigating how distinct cell fates are specified in the initial stages of *C. elegans* development.

One plasmid identified in the original *C. elegans* promoter trap screen was pUL#24C7 (Hope, 1991). Several transgenic *C. elegans* strains generated with this plasmid, including one, UL1, for which the transforming DNA was chromosomally integrated, expressed β -galactosidase in a minority of cells during embryogenesis. The β -galactosidase expression pattern in UL1 has now been determined with cellular resolution and the *C. elegans* gene with which *lacZ* was fused in pUL#24C7 has been sequenced.

MATERIALS AND METHODS

Indirect immunofluorescence microscopy

The transformed *C. elegans* line, UL1, was maintained on NGM agar plates at 20°C as previously described (Sulston and Hodgkin, 1988). Embryos were prepared by treating mixed stage nematode cultures

with alkaline hypochlorite (Sulston and Hodgkin, 1988) with an additional wash in M9 buffer. Embryos in M9 buffer were applied to 8 well multitest microscope slides (3 µl per well; Flow labs), subbed by dipping in 0.05 mg/ml bovine serum albumin. A coverslip was applied and after freezing on dry ice, flipped off with a razor blade. The embryos on the slide were fixed in methanol and then acetone, 5 minutes each, both at -20°C.

After washing in PBS (0.15 M NaCl, 10 mM phosphate pH 7.2), 10 µl monoclonal antibody diluted in PBS was added per well. The monoclonal antibody recognizing β-galactosidase (Promega) was used at 1/250 dilution. The monoclonal antibody recognizing *C. elegans* P-granules, which are restricted to the cells of the germ line (Strome and Wood, 1983), was generously provided by Susan Strome as a hybridoma culture supernatant and was used at 1/2 dilution. The slide was incubated overnight at 4°C, washed in 3 changes of PBS over a 30 minute period and 10 µl FITC-goat anti-mouse IgG (Sigma), diluted 1/60 in PBS containing 12.5 µg/ml RNase A, was added per well. After 2 hours at room temperature, the slide was washed in PBS as before and the embryos were mounted in 90% glycerol, 0.1× PBS, 25 µg/ml propidium iodide. Propidium iodide is a DNA stain and was used to reveal positions of nuclei. Stained embryos were examined using a BioRad confocal microscope with a 63× oil immersion objective to generate a series of optical sections through individual embryos.

β-galactosidase expression patterns observed on fixed embryos had to be related to the *C. elegans* cell lineage, which was originally determined by following the development of live embryos (Sulston et al., 1983). The confocal microscope images were compared with an image of an embryo at the same stage from a recording of *C. elegans* embryogenesis made on optical discs using a 4D microscope (Hird and White, 1993). The two images were aligned using the germ line cells to direct orientation. The cells of the 4D microscope image corresponding to those containing β-galactosidase in the confocal microscope image could then be identified and, by playing the 4D microscope recording backwards in time, the ancestry of the stained cells was determined.

Molecular biology

RNA was extracted from nematodes, which had been frozen in liquid nitrogen, using a modification of a procedure developed by David Pilgrim (personal communication). Mixed stage UL1 nematodes had been grown on agar plates whereas mixed stage, wild-type N2 nematodes (Brenner, 1974) had been grown in liquid culture (Sulston and Hodgkin, 1988). Extraction was achieved by vortexing the frozen nematode pellet with glass beads, guanidine isothiocyanate and phenol/chloroform/isoamyl alcohol. Extracted RNA was precipitated with ethanol, then with 2.2 M LiCl, and again with ethanol, with resuspension in water each time.

In preparation for reverse transcription, 2 µg total RNA and either 2.5 pmole of specific primer (primer 1) or 25 pmole of non-specific primer (oligo-dT/adaptor primer) in 10 µl of water was heated to 65°C for 3 minutes and then placed on ice. Reverse transcription was performed on this template using the BRL reverse transcriptase system in a 30 µl reaction volume. PCR conditions were modified from those of Frohman (1990). 20 µl reactions in 67 mM Tris-HCl pH 8.8, 6.7 mM MgCl₂, 16.6 mM (NH₄)₂SO₄, 1.5 mM each dNTP, contained 5 pmole of each primer and 1 µl of either a reverse transcriptase reaction or a 1/100 dilution of a previous PCR. Incubation was at 95°C for 5 minutes, 55°C for 5 minutes, 72°C for 20 minutes followed by 40 cycles of 95°C for 40 seconds, 55°C for 1 minute, 72°C for 3 minutes, with a final incubation at 72°C for 15 minutes.

Primers used in these reactions are as follows:

- Primer 1. M13 -40, GTTTTCCCAGTCACGAC (NEB).
- Primer 2. CGCAGTCGACGGTTTAATTACCCAAGTTTG.
- Primer 3. CTTTGGGTCCTTTGGCCA.
- Primer 4. GTCAGTCGACATGGACTGCTCTGAATAGCC.
- Primer 5. GAATCTGCAGAGGGCTTCAACATCTCGAC.

Primer 6. GTCA GTCGACGTACAATGCATTGATCGCC.

Primer 7. TGTACTGCAGATGGCAAACTCGGTTTCGCC.

Primer 8. GTCGGTTCGACAAAACTCAGAAGGCTATTC.

Oligo-dT/adaptor primer. GTGCTCCACCGCGGTGGCGGCCG-CTTTTTTTTTTTTTTTTT.

Adaptor primer. GTGCTCCACCGCGGTGGCG.

Primer 3 was synthesized by the MRC Laboratory of Molecular Biology, Cambridge. Primers 2,4,5,6,7,8, the oligo-dT/adaptor primer and the adaptor primer were synthesized by the Department of Biochemistry and Molecular Biology, Leeds. The oligo-dT/adaptor primer and the adaptor primer were designed by Sarah Gurr and Mike McPherson.

PCR amplification of the 5' end of the *pes-1/lacZ* fusion transcript

To avoid assuming that the splicing pattern predicted from the genomic sequence was correct, the 5' end of the *pes-1/lacZ* fusion transcript was amplified from UL1 strain cDNA using primers matching the start of the *lacZ* coding region and part of the *trans*-spliced leader, SL1 (Nonet and Meyer, 1991). First strand cDNA synthesis was primed on UL1 total RNA using primer 1. A PCR using this cDNA was primed with primer 1 and another oligonucleotide (primer 2) matching part of the *trans*-spliced leader, SL1 (Krause and Hirsh, 1987; Nonet and Meyer, 1991). At least nine DNA products of distinct size were observed. Therefore, a second PCR was performed using products of the initial PCR, the SL1 based primer again and a third oligonucleotide (primer 3) matching a part of the vector plasmid, which was expected to be included in the *pes-1/lacZ* transcript, 5' of the *lacZ* coding region. The second PCR gave three DNA products of approximately 410 bp, 310 bp and 170 bp. These three products were purified by agarose gel electrophoresis and sequenced directly. The sequence obtained showed that all three products were derived from *pes-1*. The smallest fragment gave very clear sequence and appeared relatively pure. The two larger fragments, however, were contaminated with other reaction products. Therefore, these gel purified DNA products were reamplified using the SL1 based primer and an oligonucleotide (primer 4) matching the *pes-1* genomic sequence just upstream of the point of fusion with *lacZ*. In both cases DNA products of a single size were generated, approximately 50 bp smaller than the template fragment, as expected. (Fig. 1 shows the three products generated using the SL1 based primer and primer 4 in a PCR reaction using unfractionated PCR products.) Direct sequencing of the products of this third round of PCR showed that they were no longer contaminated with other DNA products.

PCR amplification of the 3' end of the *pes-1* mRNA

Two oligonucleotides (primers 5 and 6) were designed to match sequences known, from analysis of the 5' end, to be present in the transcript (Fig. 1). Another oligonucleotide (primer 7) was made to match the genomic sequence just downstream of the point of fusion with *lacZ*, a region predicted to be included in the *pes-1* transcript. Total RNA was prepared from a wild-type *C. elegans* strain, N2 (Brenner, 1974), as well as the transgenic strain, UL1. An oligo-dT/adaptor oligonucleotide was used to prime first strand cDNA synthesis, priming on the poly(A) tail, for both RNA preparations. Nested series of PCRs on cDNA prepared from a wild-type *C. elegans* strain, N2 (Brenner, 1974), as well as the transgenic strain, UL1, using primers 5, 6 and 7 with the adaptor primer, which matches part of the oligo-dT/adaptor primer, failed to give *pes-1*-specific bands. Therefore, another oligonucleotide (primer 8) was prepared to match the genomic sequence just after the predicted *pes-1* stop codon. Two rounds of PCR, starting from the oligo-dT/adaptor primed cDNA, using primer 6 with the adaptor primer and then primer 7 with primer 8, generated a single product of 410 bp. This means, however, that the length of the 3' untranslated region for *pes-1* transcripts is not known.

PCR products were purified by electrophoresis in low-melting point

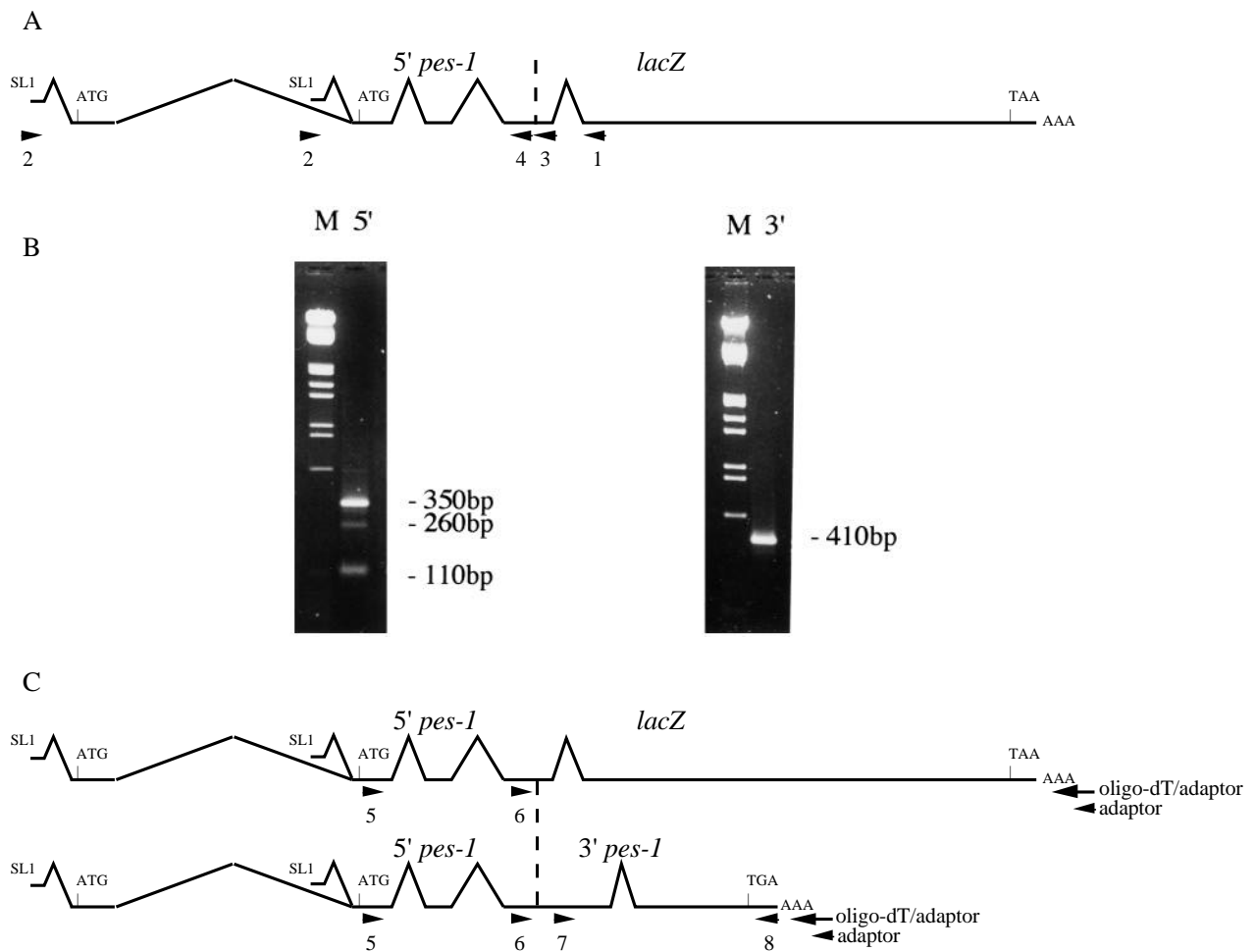


Fig. 1. PCR amplification of *pes-1* and *pes-1/lacZ* transcripts. (A) The positions of primers used to amplify the 5' end of the *pes-1/lacZ* transcript are indicated by the arrowheads with the primers numbered as described in the text. The position of the fusion of *pes-1* and *lacZ* is indicated by the dashed line. SL1 is the *trans*-spliced leader, ATG and TAA indicate the potential sites for initiation and termination of translation and AAA represents the poly(A) tail. (B) PCR products analysed by agarose gel electrophoresis. 5' indicates the track which contains the products, with sizes indicated, derived from amplification of the 5' end of the *pes-1/lacZ* transcript. UL1 RNA was reverse transcribed using primer 1, amplified using primers 1 and 2 and reamplified using primers 2 and 4. 3' indicates the track containing the product derived from amplification of the 3' end of the *pes-1* transcript. UL1 RNA was reverse transcribed using the oligo-dT/adaptor primer, amplified using the adaptor primer and primer 6, and reamplified using primers 7 and 8. M identifies the molecular mass marker tracks containing λ DNA digested with *Hind*III and *Eco*RI. (C) The positions of primers used to amplify the 3' end of the *pes-1* transcript are indicated as in A. The UL1 RNA preparation contained both the *pes-1* transcript and the *pes-1/lacZ* transcript whereas the N2 RNA preparation contained only the *pes-1* transcript. TGA indicates the site of translation termination for the *pes-1* transcript.

agarose gels (FMC-Sea Plaque). Direct sequencing of purified PCR products was carried out as described (McPherson et al., 1992). PCR products were cloned using a TA cloning kit (Invitrogen). Genomic and cDNA sequencing using the USB sequenase 2.0 kit, was performed on single stranded DNA, prepared by subcloning into M13mp18/19 vectors (Yanisch-Perron et al., 1985). Amino acid sequence homology was detected using Sweep (Akrigg et al., 1992) to search the Owl composite database and using Blastx (Altschul et al., 1990) to search the swissprot database.

RESULTS

β -galactosidase expression pattern in UL1

The UL1 strain of *C. elegans* is a transgenic strain carrying pUL#24C7, a plasmid selected in a promoter trap screen of

developmental expression patterns because of the embryonic, β -galactosidase expression that the plasmid directed. UL1 embryos were examined at various stages of development to determine the precise developmental distribution of β -galactosidase expression.

The earliest β -galactosidase expression detected during embryogenesis was in the AB lineage as the AB lineage increases from 16 to 32 cells. This set of cell divisions, producing an embryo of 44 cells, occurs at approximately 105 minutes after cleavage of the zygote. Staining is strongest and appears first in the four granddaughters of ABalp, found on the anterior ventral surface of the embryo (Fig. 2A). Staining around these cells (the daughters of ABaraa, ABplaa and ABplpa) and in four cells on the dorsal/posterior surface (the daughters of ABpraa and ABprpa) appears slightly later and is

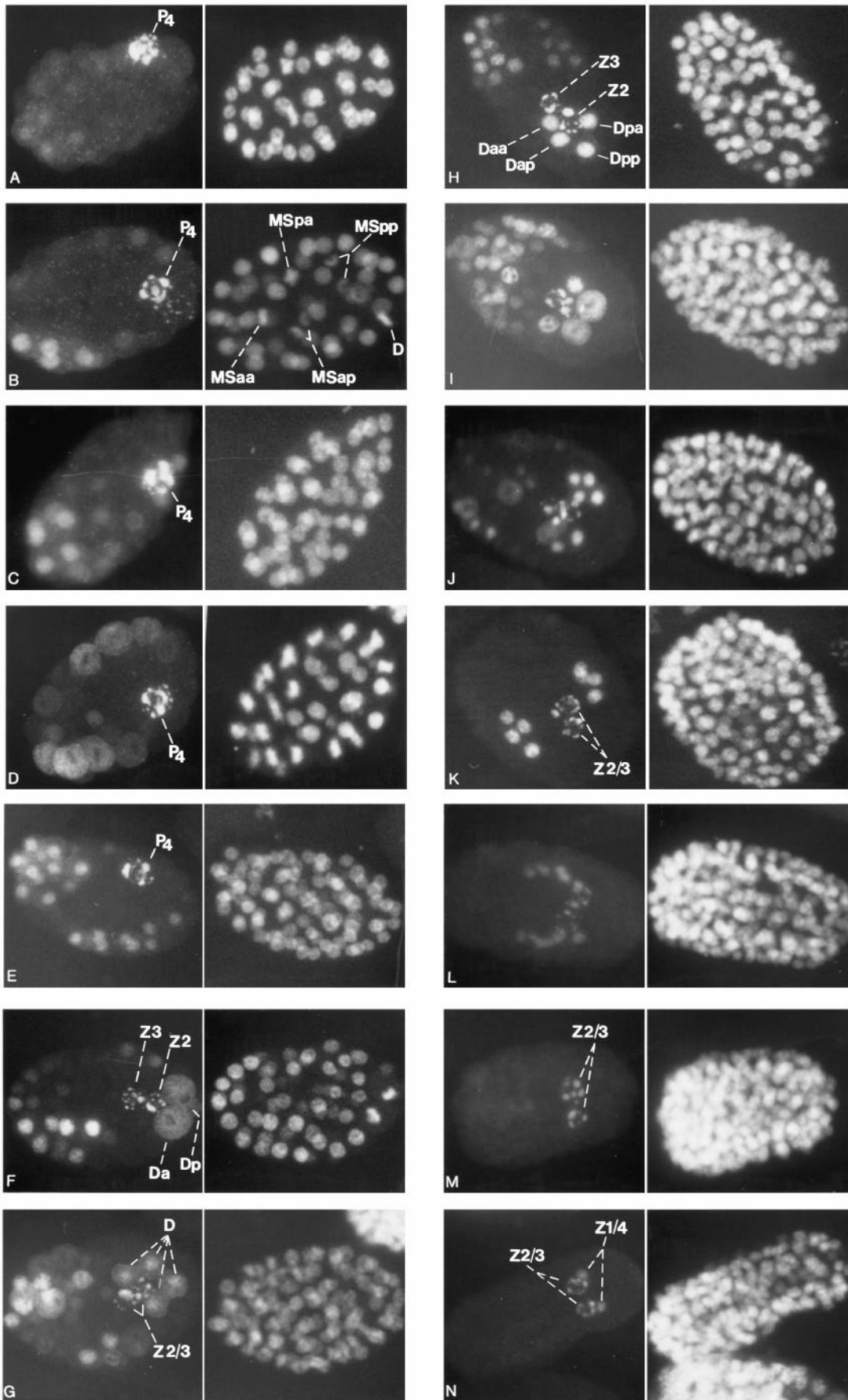


Fig. 2. β -galactosidase expression pattern in UL1 during early embryogenesis. Successive panels are images of progressively older embryos obtained using a confocal microscope. Each panel is a summation of optical sections through a part (B,D,F,H,J,K,L,N) or through the whole (A,C,E,G,I,M) of an embryo. The left half of each panel shows the location of β -galactosidase and the position of the P-granules of the germ line cells (P₄ in A-E, Z₂ and Z₃ in F-N). The right half of each panel shows the distribution of DNA in the same embryo. The stage of development of a stained embryo was determined by counting the number of nuclei present through the optical sections. Cells descended from a particular founder cell divide at the same time during early embryogenesis and this was used in cell identification. For example, in the 46-cell embryo of B the 4 descendants of the founder cell MS are dividing and the D founder cell is about to divide. The images are presented with anterior to the left but the direction of the dorsal-ventral axis varies between panels because the staining procedure meant that embryos could not be viewed in a particular orientation. (A) 43-cell embryo. One of the AB descendants has not yet divided in this round of AB lineage cell divisions. β -galactosidase can be detected in anterior AB descendants. (B) 46-cell embryo. Cells of the MS lineage are dividing and the D founder cell is about to divide. β -galactosidase is now at higher levels in 4 anterior AB descendants and detectable in additional AB descendants on the anterior-ventral and posterior-dorsal surfaces. (C) 51-cell embryo. The MS lineage and D founder cell have completed cell division. (D) 58-cell embryo. Cells of the AB lineage are in the process of dividing, some have just completed cell division, others are about to start. (E) 87-cell embryo. AB lineage cell divisions have finished and two patches of β -galactosidase containing cells are present, 20 cells anteriorly and 8 cells posteriorly. (F) 100-cell embryo. P₄ has now divided to give Z₂ and Z₃. Da and Dp are about to divide and β -galactosidase is detectable in these cells. (G) 103-cell embryo. Division in the D cell lineage (cells labelled D) is now complete and another round of cell division in the AB lineage has begun. (H) 185-cell embryo. β -galactosidase levels are now very high in the D lineage and beginning to drop in the AB lineage. The posterior group of AB cells expressing β -galactosidase have begun to move towards the anterior. (I) 189-cell embryo. Cell division is about to begin in both the D lineage and the AB lineage. (J) 220-cell embryo. β -galactosidase in the AB lineage has almost disappeared. The cell divisions in the D lineage are close to completion, generating 8 cells containing β -galactosidase. (K) 338-cell embryo. The round of cell division in the AB lineage is now complete. There are now 256 descendants of AB, none of which contain detectable levels of β -galactosidase. β -galactosidase is restricted to the 8 descendants of D. (L) 385-cell embryo. Another round of cell division in the D lineage has created 16 cells containing β -galactosidase, but at a lower level now. These cells have begun to move towards the anterior away from Z₂ and Z₃. (M) This embryo has at least 400 cells but elongation has not yet begun. β -galactosidase was not detected in any cells of embryos at this stage of development. (N) Elongation has begun and two cells, Z₁ and Z₄, begin to express β -galactosidase. The subcellular distribution of β -galactosidase was affected by cell division. In non-dividing cells the β -galactosidase was nuclear localized, possibly a consequence of the nuclear localization signal encoded in the vector part of pUL#24C7. In a cell beginning to divide the nuclear envelope breaks down and the β -galactosidase spreads throughout the cell, with the exception of the condensed chromosomes from which it appears to be excluded. (e.g. cells of the AB lineage in D and cells of the D lineage in F.) Immediately after cell division and reformation of the nuclear envelope the β -galactosidase remains throughout the cell with the concentration in the nucleus increasing as the level in the surrounding cytoplasm decreases. (e.g. cells of the AB lineage in E and cells of the D lineage in G.) Disappearance of β -galactosidase from the cytoplasm after cell division may depend on degradation of the enzyme or uptake into the nucleus.

weaker (Fig. 2B,C). This gives two physically separated groups of staining cells within the AB lineage. β -galactosidase levels increase in these 14 AB descendants and their daughters (Fig. 2E,F) but by the next set of cell divisions in the AB lineage, which give 128 AB cell descendants in the 168 cell embryo, β -galactosidase in the AB lineage is beginning to disappear. The numbers and positions of staining cells are consistent with β -galactosidase being restricted to the 56 granddaughters of the original 14 staining cells but by this stage there were too many cells to identify all the cells of the fixed embryos without additional cell markers. As β -galactosidase in the AB lineage decreases, the expression pattern becomes symmetrical about the embryo's plane of bilateral symmetry. The anterior group appears to divide into two clusters, one remaining in the midline and the other appearing as a mirror image of the previously dorsal/posterior group of staining cells, which has now moved towards the anterior (Fig. 2I,J).

The emergence of symmetry within the staining cells of the AB cell lineage is consistent with the fates of these cells as described in the embryonic cell lineage (Sulston et al., 1983). Each of the cells making up the dorsal/posterior cluster of staining cells (i.e. ABpraaa, ABpraap, ABprpaa and ABprpap) is one half of a cell pair related by fate. The cells of each pair arise asymmetrically but subsequently develop with mirror symmetry on opposite sides of the embryo. The symmetric partners of the dorsal/posterior group of staining cells (i.e. ABalppp, ABplaap, ABplpaa and ABplpap) are members of the ventral/anterior group of staining cells.

The second component of the expression pattern is in the lineage of the D founder cell and begins as β -galactosidase levels peak in the AB lineage. β -galactosidase was detected in the daughters of D as they are about to divide in the 100 cell embryo, approximately 160 minutes after the initial cleavage of the zygote (Fig. 2F). β -galactosidase persists in all the cells of the D lineage as this lineage increases to 4 cells (Fig. 2H), 8 cells (Fig. 2K) and 16 cells (Fig. 2L). At this point the D cells move anteriorly to become body wall muscle cells and the β -galactosidase disappears.

Embryos at this stage, just prior to elongation, do not stain for β -galactosidase (Fig. 2M) but as embryonic elongation commences the third component of the expression pattern appears in two cells, Z₁ and Z₄ (Fig. 2N). As previously described (Hope, 1991), β -galactosidase activity increases in these two cells during the rest of embryogenesis and disappears shortly after hatching when these two cells divide and give rise to the somatic part of the gonad during postembryonic development.

The expression pattern is summarized in Fig. 3. After determination of the expression pattern, the gene with which *lacZ* was fused in pUL#24C7 was called *pes-1* (pattern expression site).

Genomic DNA sequence of *pes-1*

As a first step towards characterization of *pes-1*, the genomic sequence was determined (Fig. 4). The start of the *pes-1* gene was sequenced from the genomic DNA insert of pUL#24C7. The rest of the gene was sequenced from the cosmid T28H11, a clone from the *C. elegans* physical genome map (Coulson et al., 1991) known to contain *C. elegans* genomic DNA including and flanking the insert of pUL#24C7 (Hope, 1991).

The genomic sequence around the site of fusion with *lacZ*

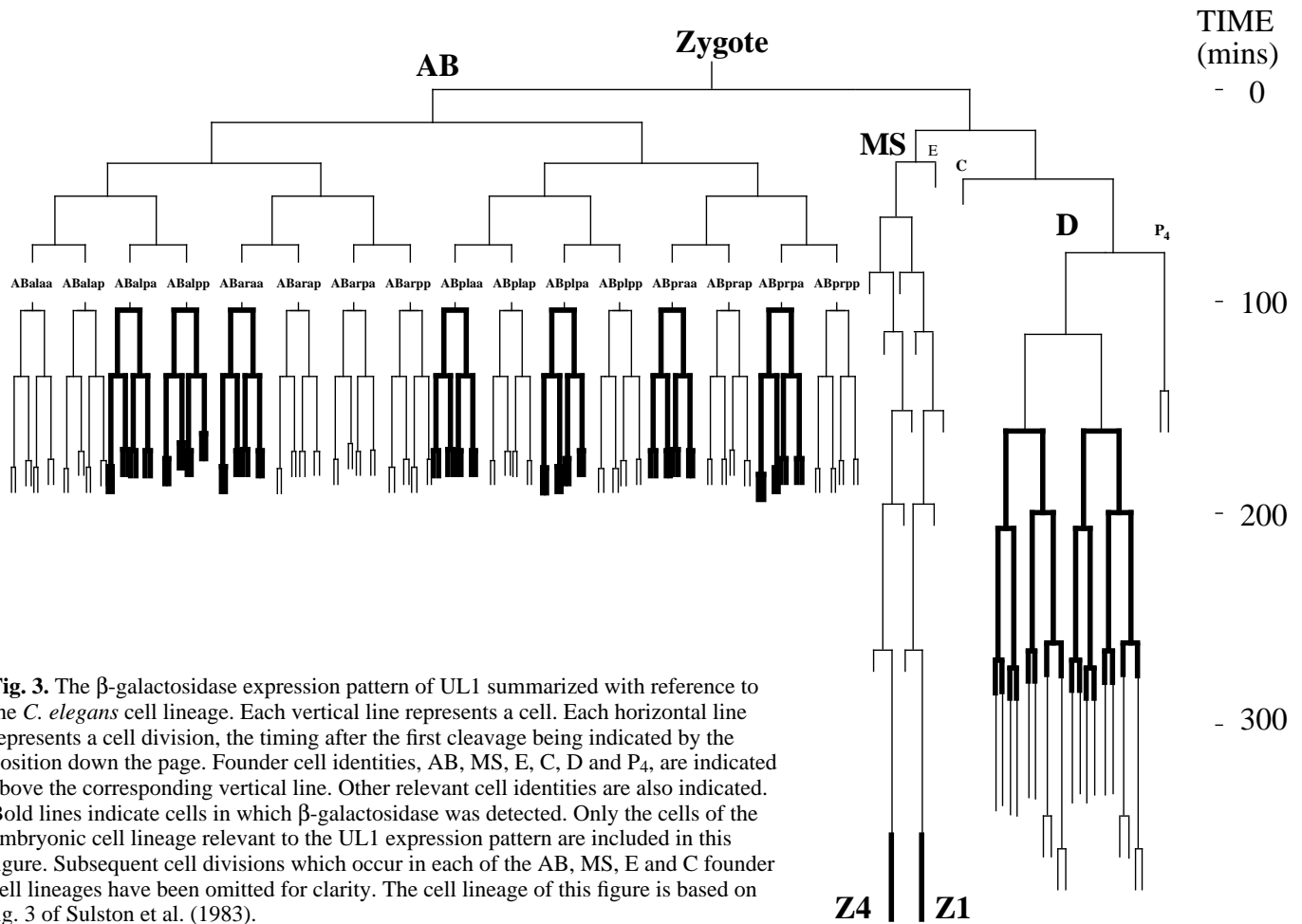


Fig. 3. The β -galactosidase expression pattern of UL1 summarized with reference to the *C. elegans* cell lineage. Each vertical line represents a cell. Each horizontal line represents a cell division, the timing after the first cleavage being indicated by the position down the page. Founder cell identities, AB, MS, E, C, D and P₄, are indicated above the corresponding vertical line. Other relevant cell identities are also indicated. Bold lines indicate cells in which β -galactosidase was detected. Only the cells of the embryonic cell lineage relevant to the UL1 expression pattern are included in this figure. Subsequent cell divisions which occur in each of the AB, MS, E and C founder cell lineages have been omitted for clarity. The cell lineage of this figure is based on fig. 3 of Sulston et al. (1983).

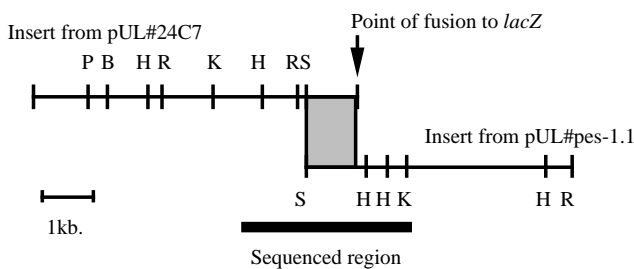


Fig. 4. Restriction maps, approximately to scale, of the inserts of pUL#24C7 and pUL#pes-1.1 used in subcloning DNA fragments for sequencing. The 6.4 kb. genomic DNA insert of pUL#24C7 contains the start of the *pes-1* gene. The 2.2 kb of the pUL#24C7 insert directly adjacent to *lacZ* was sequenced. A DNA fragment containing the rest of the gene was subcloned as a 5.3 kb. *SphI/EcoRI* fragment from the cosmid T28H11 (provided by Alan Coulson and Ratna Showneken) into pUC19 (Yanisch-Perron et al., 1985) producing the plasmid pUL#pes-1.1. From this plasmid, 1 kb of *pes-1*, downstream from the point of fusion with *lacZ* was sequenced. Restriction enzyme sites marked are: B, *Bam*HI; H, *Hind*III; K, *Kpn*I; P, *Pst*I; R, *Eco*RI; S, *Sph*I. The end of the insert fused to *lacZ* in pUL#24C7 is indicated with the arrow, the region of overlap between the two clones is indicated by the shaded box and the region sequenced is indicated by the heavy line.

in pUL#24C7 is shown in Fig. 5. The reading frame of the exon with which *lacZ* was fused could be inferred because expression of β -galactosidase was expected to depend on transcriptional and translational fusion to a *C. elegans* gene. From this starting point and using the consensus sequences for splicing donor and acceptor sites in *C. elegans* (Fields, 1990), a splicing pattern was predicted for the *pes-1* gene (Fig. 5). The ease with which this exon/intron structure could be predicted suggested that *pes-1* was a real *C. elegans* gene and that fusing random fragments of *C. elegans* genomic DNA to the *lacZ* reporter gene, in the promoter trap approach, had not simply revealed a cryptic promoter. The splicing pattern still needed to be confirmed, however, with cDNA analysis.

pes-1 transcripts

Messenger RNA produced by the *pes-1* gene was investigated by sequencing cDNA generated using the polymerase chain reaction (PCR; see Fig. 1) (Frohman, 1990). Three *pes-1*-specific products, of approximately 350, 260 and 110 bp, were obtained for the 5' end of the mRNA and sequenced. The 260 bp product corresponded to the transcript predicted from the genomic sequence, whereas the 350 bp and 110 bp products were not expected. The 110 bp product appears to be anomalous for various reasons. First, this product contains an additional 12 nucleotides at the 5' end and the origin of this segment was not found in the 3.1 kb of genomic sequence

```

1      * * * * *
agaacacacaaactcctagttttggcatcgtctgctgttaagtgattcatcaaacatctttctgtttcacaatctttttaattccgcaagtcg
101      * * * * *
gotttcaaaacttattagataaattgttttttactactacataataacatttagATGACGTCATCAATCAAAATCTGATGCTCCACAATTTCTGCT
                                M T S S I K S D A P Q F L L
                                ↑
201      * * * * *
CGATTGGATAACTGTCTTCACTTCTCCAACACCGCGAAAACCGCATCTCCAGGtaagttgcaatgctaatacatagtttaaggaagcatcgggggt
D L D N C S S L P P T P P K T A S P
301      * * * * *
ttgaaaatcogtgaagtaacaacattcatgaaacagtaatttttttcgaaagcacaaaaatcttaattttcccaagcttgaataaaaaactt
* * * * *
401      * * * * *
tgtgaaatcccatatccattacataaattggcagatgtaattatgctcggtttatgtgttcacagagcaattcacactcattgtttctctttttgcca
* * * * *
501      * * * * *
agtttgccgaaaaagcgtcgggattgggtttcccatcattgtgttgagtaggtgctatgatagggtgtaattggattgggttacggtcggaatttgcg
* * * * *
601      * * * * *
gaaaagtattttgggctctcttcgaaaccattatgtgtcaaaaacagtaatttattcaactgtttcaacgtagttaaagacaataggttctattcga
* * * * *
701      * * * * *
ggattgtgtagatatacaggagtaggagagaaaagtaacttttaaccttttaattatctgtctaaataatctccacacagtcctagcaaaaaaaaaa
* * * * *
801      * * * * *
cattttcgaaaagttcaatgttttcatgtgcataaaaatgtttatgttatttagcaaaaatactagcatttagctctatataataaccatctaccgta
* * * * *
901      * * * * *
tattctctattagagcggcgtatcattagacggcgtatctattagagcggcaccctaactggctcttgaaaattagagcgcagctctctattagtcg
* * * * *
1001      * * * * *
ggcaccctctttttttgggtgagagagccacctcaatattgtattcttcgaagacattttgttcttctgaatgaattcatttaaatattgaattt
* * * * *
1101      * * * * *
tagtgtctctgtgaggcaaaattacacaattttcataaattttccataaaaatagaggttttagaaggaaaatttgaggttactactcccaaaagttag
* * * * *
1201      * * * * *
taacttcaatagaaaattagttatcgaaaatagtcgcccagatccagtagaagcagctctcctaataagagcgcagctcgcgaaggcccggaataa
* * * * *
1301      * * * * *
gtgcccgtgcccctctaataatagaatccggtacatttcatggttgatgaacgggcaactaaagccaggtattgttcatgcagaagatgaggttcgag
* * * * *
1401      * * * * *
atgtattgtcagcaccactctgaaattgaacacatcaccagatctcgtgggtgatgaactggaaaagtgaggagatgggctagattatacatttaata
* * * * *
1501      * * * * *
tgggactgacttaaatatttactgtttcaatgttttaaatagttttgttccatttcaactgttttctgttacgcaattttcaacgcttaatttag
* * * * *
1601      * * * * *
ctgtgcataaataaattctgttgagctgtgttctctcatgaattgctgactcaaaactcgagatgctctgtagagctaatcgaagtttacagaa
* * * * *
1701      * * * * *
tagttaaactctgttctccaaaaaatctgatttctcttccctgacaccgactttcagatttatttatttctgtaactacaaatttcagGTAATTCA
                                G N S
1801      * * * * *
AAAATGAAGGGCTTCAACATCTCGACTTGTGCCTCGACTTGGACAGCTCAACTTCCAGCTTCTGCTCCGTTTACCACGgttaatttttactgtttca
K M K G F N I S D L C L D L D S S T S S S C S V S P
                                ↑
1901      * * * * *
aacagtttcaaacacgcttaaaatgtgoccatatgtaatttaattgttccagCTTCTCTCCACACCCGCTCTGAATCAGTCGGTCAACAACAG
                                A S S F H T R S E S V G Q Q Q
2001      * * * * *
TCTGGTAGAAATTCACCAAGtaaggcaaaatattttgttttaaaatgtcaagtttcaactactatgatagtttcataccattcagcagatttttaatt
S G R N S P
2101      * * * * *
actatgtttacaaaaatttaccatactttataatttttagTCTCAAGCTCCACCGAATCACCAACCAAAGACCAAAATACTCGTACATGCATTGAT
                                V S S S T E S P T K R P K Y S Y N A L I
2201      * * * * *
CGCCATGGCTATTTCAGAGCAGTCCATCAAAAAGTTTGAGAGTCTCCGAGATCTACAATAATATTTCTTCCAACTTCTCCTACTACAAAAACAGAGCCA
A M A I O S S P F K S L R V S E I Y K Y I S S N F S Y Y K N O K P
2301      * * * * *
CTTCAATCGGCAAAACTCGGTCGCGCACAACTGTCTTTGCTATAAGAGTTCCGAAAGTTTCGAACTTTGGATGAAAAGgtttgttttaacaggactt
L Q W Q N S V R H N L S L H K E F R K V R T L D G K
2401      * * * * *
aatagaataatagttttcttttcagGCAGCTACTGGGCAATGACTCGCGATTAGGAAGTATGTTTATATCAGCAACAACCTCGGAAAGCTTCGTCGC
G S Y W A M T A D L G T D V Y I S N N C G K L R R
2501      * * * * *
CAAAAAGCAAAAGTGGCAAGTTCCACCAATGCAACAACATTTCCAATCCCTCAGCTTCCAAACCAAAACATTCAACAATTTGTATGCAAAATCCTC
Q K S K V A K F P P M Q Q H F P I P Q L P T Q N I H Q L C M Q N P
2601      * * * * *
AAATCTTGCTACATTGCTTCAAAATATGTATCTTCAAAATATGCAAAAACCTCCAGAATATCCCCATGTTCTGGCTTTCCAATAATCCAGTTCCCAT
Q I L A T L L Q N M Y L Q N M Q N L Q N I P M V P G F P I I P V P I
2701      * * * * *
TAATCTACTCTCTTCAATTTTCTTAAAGTCTTGATAGAAAGCTTCTGAGTTTTTgtttaaattgattgaattaactgtgttccaaaatcagtttt
N P T S F H F P K S S
2801      * * * * *
ttttgttttcccagattgttttaaatattgtattgttgatttgataaatttagaaaactagattctctgtataaagcttgaactcaccg
* * * * *
2901      * * * * *
caacaattagtgaaataaagcttccaaaacaagaagacaagattagtgattgcattaattgaaggggttctctctctttgaaactccggtgggttgc
* * * * *
3001      * * * * *
aaaactacttattttctccgctgttagttagaatgacccgggtcaaaaataaaagctctgagatatactccgatttaatttttttcaaatgttct
* * * * *
3101      * * * * *
atttaccagtgattttccaataaacaataatctgtaactgtgctttaaatttcaactccggtatagatatttctcattcaaaaagaattaccaaga
* * * * *
3201      * * * * *
agcagaaaacttctactttgaagtattataaattcaaggtacc 3245

```

Fig. 5. The genomic sequence of *pes-1*. Exons are in upper case letters. Introns and untranscribed DNA are in lower case letters. The amino acid sequence of the predicted translation products are indicated with the single letter code, and the two putative sites for initiation of translation are indicated by the small, upward-pointing arrows. The region of homology to the fork head family of proteins is indicated by underlining of the amino acid sequence. Immediately upstream of the fork head homologous domain there is a serine rich region (16 of 36 residues). Downstream of the fork head homologous domain there is a large region lacking charged residues but which contains six repeats of xQN, where x = L (three times), M (twice) or T (once). The *Sau3AI* site at which *pes-1* was fused to *lacZ* is indicated by underlining of the nucleotide sequence. Priming sites used in PCR amplification of cDNA are indicated by horizontal arrows with numbers as described in the text and Fig. 1. The two *pes-1* transcripts differ in the presence/absence of the first exon, nucleotides 159-255. The first intron contains 11 degenerate copies of a 17 bp element, T(C/A)TCTATTAG(T/A)GCGGCA, significance unknown. The sequence of the small PCR product observed when amplifying the 5' end of the *pes-1* transcript includes sequence matching the *trans*-spliced leader SL1 joined to 12 nucleotides of unknown origin (CAATATAAATCG) joined to *pes-1* nucleotides 2146 upto the point of fusion with the plasmid vector.

presented here. Second, the position in *pes-1* with which the 12 nucleotide section is fused does not correspond to the consensus for a potential splice site (Fields, 1990). Third, the first ATG of a transcript corresponding to this PCR product would not be in the correct reading frame, as predicted from the genomic sequence. The 110 bp product may be a PCR artefact resulting from inappropriate hybridization of products derived from transcripts of *pes-1/lacZ* and of a distinct *C. elegans* gene unconnected with *pes-1*.

The 350 bp product corresponded to a *pes-1* transcript with an additional exon of 97 nucleotides at the 5' end. Inspection of the genomic sequence revealed the origin of the 97 nucleotides and that splicing of the extra exon on to the rest of the transcript would require the loss of a 1.5 kb intron. Presumably there are two sites at which *pes-1* transcription may initiate.

A specific DNA fragment of 410 bp was obtained for the 3' end of the *pes-1* mRNA. Sequencing revealed that the product was derived from the *pes-1* transcript and confirmed the splicing pattern predicted for this downstream portion of *pes-1*. N2 RNA as well as UL1 RNA served as template for this PCR product showing that *pes-1* transcript is produced in wild-type *C. elegans*.

Combining the PCR analysis of the 5' and 3' ends of *pes-1* mRNA suggests that the gene produces two transcripts. If the first AUG of each transcript initiated translation, both transcripts would have very short 5' untranslated regions, the 22 nucleotides of the *trans*-spliced leader, SL1 (Krause and Hirsh, 1987), only for the longer transcript and 33 nucleotides, including SL1, for the shorter transcript. SL1 is a 22 nucleotide long, RNA segment of specific sequence which is transferred through a *trans*-splicing reaction to the 5' ends of many *C. elegans* mRNAs (Krause and Hirsh, 1987). Most *C. elegans* genes appear to produce transcripts with SL1 present, although the transcripts of any particular gene may not all be *trans*-spliced. Translation from the first AUG of the two transcripts would produce proteins of 264 and 228 amino acids. The larger protein would have the same sequence as the smaller protein but with an additional 36 N-terminal amino acids encoded by the extra exon. These alternative transcripts are both translated because an anti- β -galactosidase monoclonal antibody, used to probe a western blot of total protein from the UL1 strain of *C. elegans*, detected two major β -galactosidase fusion proteins of appropriate sizes (Hope, 1991).

DISCUSSION

A *C. elegans* gene, *pes-1*, has been cloned on the basis of the developmental expression pattern produced upon fusion with a *lacZ* reporter gene. This β -galactosidase expression pattern has now been determined at the level of individual cells. Confirmation that a real gene has been identified comes from genomic DNA and cDNA analysis.

Searches of protein sequence databases with the PES-1 amino acid sequence, as deduced from the cDNA sequences, revealed homology to the fork head family of transcription factors (Weigel and Jackle, 1990; Fig. 6). Fork head homologues have been found in vertebrates (Lai et al., 1990, 1991; Tao and Lai, 1992; Li et al., 1991; Knochel et al., 1992; Dirksen and Jamrich, 1992), *Drosophila* (Weigel et al., 1989;

```
pes-1. SPTRKRFKYSYNALIAAIAIOSSEPKSLRVSEIVKYLSSNFSSYKNOKPLQWQ
fkh...YTHAKPPYSYISLIIMAIQNNPTRLTLSEIVQFIMDLFFEFRR...QNQRWQ
slp-1. SDTKKPPYSYNALIMMAIQDSPEORLTLNGIYOYLINRPPYFRANK...RGWQ
ycv-5. ELAKKPPYSYATLICTAILQSQECKLTLSCIVYEWIVHVPFYR...OKDASWQ

pes-1. NSVRHNLSTLHKERKVRRTLD...GKGSYVAMTADLCTDVIISNNCKLRKQ
fkh...NSIRHSLSEINDCFVKLPRTPDKPKGSEFWLHPDSENMEENG...CYLRRQK
slp-1. NSIRHNLSTLHKCFIKIPRSWDDPKGNYWILDESAE...EVFVIGETLTKLRRKN
ycv-5. NSIRHNLSTLNDAFIRIKTEKSCD...GKGFVWVVRPCAEIKFKKGENRGYEFVKD
```

Fig. 6. Homology of PES-1 protein with other members of the fork head family. The amino acids shown, using the single letter code, are: 89-198 for the larger predicted PES-1 protein, 206-304 for fork head protein (fkh; Weigel et al., 1989), 116-216 for sloppy paired 1 protein (slp-1; Grossniklaus et al., 1992) and 105-204 for the yeast protein (ycv-5; YCR902 in Benit et al., 1992). Amino acid identities between these proteins are indicated with white letters in black boxes. No significant homology was detected for regions of PES-1 outside of this fork head homologous domain.

Hacker et al., 1992; Grossniklaus et al., 1992) and yeast (Benit et al., 1992), and two were recently identified from *C. elegans* (Miller et al., 1993; M.Azzaria and J.McGhee, personal communication). The region showing the greatest similarity between members of this family of proteins has been shown to have specific DNA binding activity (Lai et al., 1990, 1991). It is likely that all members of the fork head group of proteins, including PES-1, are transcription factors involved in the developmental control of other genes. The diversity in the developmental expression of the fork head homologues suggests that they are not involved in just one aspect of development. The fork head homologues identified so far may be the first of a large family of transcription factors with many roles in animal development.

PES-1 appears to be a diverged member of the fork head family with the sequence similarity to other family members restricted to 100 amino acids of the potential DNA binding domain. *Drosophila* fork head (Weigel et al., 1989), *Xenopus* XFKH-1 (Dirksen and Jamrich, 1992) and the rat hepatocyte nuclear factors 3 α , β and γ (Lai et al., 1991) appear to form a closely related sub-group with more than 80% amino acid identity in comparisons of the DNA binding domains of any two proteins. Sloppy paired 1 and 2 of *Drosophila* (Grossniklaus et al., 1992) and rat BF-1 (Tao and Lai, 1992) make up a separate subgroup; amino acid identity being at least 75% for this domain. In contrast, in comparisons between these subgroups, YCV5 of yeast (Benit et al., 1992) and PES-1, amino acid identity over this region is approximately 50%. PES-1 could be the first member of a new, fork head family subgroup which may include vertebrate and *Drosophila* proteins as yet uncharacterized.

A role for PES-1 as a transcription factor could explain why the fates of the cells expressing the PES-1/ β -galactosidase fusion protein appear to be unrelated. Z1 and Z4 produce the somatic part of the gonad which consists of cells of mainly epithelial character (Kimble and Hirsh, 1979), whereas the D lineage produces only body wall muscle cells, and the stained cells of the AB lineage produce a variety of cell types including muscle cells, nerve cells and hypodermis (Sulston et al., 1983). In general, genes expressed to confer appropriate characteristics for a particular cell type are not coordinately activated by a single transcription factor. Rather, different cell types are specified by different subsets of transcription factors acting in

a combinatorial manner such that any particular factor may be involved in specifying various cell types. Hence, PES-1 could be involved as a transcription factor in the specification of several distinct cell fates during *C. elegans* embryogenesis.

The existence of two different PES-1 proteins may also be significant for the potential role of *pes-1* in specification of distinct cell fates. The two *pes-1* transcripts appear to encode different proteins which could have distinct regulatory activities. The expression patterns of the two *pes-1* transcripts, and therefore PES-1 proteins, may be unconnected and the UL1 β -galactosidase expression pattern may be simply the sum of the separate components. The 1.5 kb intron spliced out in formation of the larger of the two transcripts is unusually large for *C. elegans* (Blumenthal and Thomas, 1988) and is in sharp contrast to the other *pes-1* introns (77, 123 and 47 bp). This large intron may contain the regulatory elements necessary specifically for control of expression of the smaller transcript.

Although the expression pattern for PES-1 may be different from that observed for PES-1/ β -galactosidase, the link of *pes-1/lacZ* expression with founder cell lineages has implications for the mechanisms by which *C. elegans* development is controlled. Nematodes, with their invariant cell lineage, were considered classic examples of mosaic development (e.g. Laufer et al., 1980), each cell appearing to develop autonomously, cell fate being specified by intrinsic determinants segregating to particular daughter cells at each division. Recent experiments, however, have shown that regulative cell interactions have a major role in early *C. elegans* development (Priess and Thomson, 1987; Wood, 1991; Schnabel, 1991). The invariant cell lineage is presumably a consequence of all regulative interactions being precise and reliable from one embryo to another, perhaps with back up mechanisms supporting primary signals.

One of the components of the PES-1/ β -galactosidase expression pattern appears to be concerted expression within the AB lineage as if something intrinsic to this founder cell lineage has influence over *pes-1/lacZ* expression. However, expression is only observed in some members of the AB cell lineage and this distinction within the AB lineage could arise from signals emanating from other parts of the embryo. Two primary, intercellular, inductive signals have been proposed for early *C. elegans* development (Schnabel, 1991). The MS and C founder cells are thought to send independent signals to adjacent AB cell descendants along the ventral and dorsal surfaces, respectively. These primary interactions could regulate *pes-1* directly because the earliest expression of the PES-1/ β -galactosidase fusion protein occurs in two groups of AB descendants, a large, ventral group of cells and a smaller, dorsal group. The primary signals, however, are thought to act at the 8 AB cell stage (Schnabel, 1991), whereas the expression pattern suggests the non-autonomous component of the specification to express PES-1/ β -galactosidase within the AB lineage occurs at the 16 AB cell stage. Nevertheless, *pes-1* could provide an example of cell intrinsic determination and intercellular regulative interactions converging at the level of the regulation of a single gene.

There appears to be a delay between the birth of cells that give rise to each component of the UL1 expression pattern and β -galactosidase reaching detectable levels. When detected in the D cell lineage, β -galactosidase activity was found throughout the D cell lineage. But expression is first detected in this lineage two cell generations, or approximately 90 minutes,

after the D founder cell is produced. Expression within the AB cell lineage was not detectable until approximately 105 minutes after the AB founder cell is born. Z1 and Z4 are generated at 250 minutes after the first cell division of embryogenesis but β -galactosidase was not detected in these cells until elongation had begun at approximately 350 minutes of development. The similar length of delay for the three components of the expression pattern, approximately 100 minutes, may be a coincidence or may be relevant to the mechanisms controlling *pes-1* expression. The delay could be simply the time required to attain detectable levels of PES-1/ β -galactosidase after the gene begins to be expressed. Alternatively the delay could reflect a specific temporal distinction between specification to express *pes-1* and actual initiation of *pes-1* expression.

The UL1 expression pattern during early embryogenesis and the sequence homology with a family of transcription factors suggest *pes-1* occupies an intermediate position along a developmental regulatory pathway. Therefore, it will be important to examine the mechanisms by which expression of the *pes-1* gene is controlled and the role of PES-1 protein in *C. elegans* embryonic development.

I thank Siegfried Hekimi for guidance in use of the confocal microscope, Joel Rothman for providing a 4D recording of wild-type embryonic development, Susan Strome for the anti-P-granule monoclonal antibody, Noel Carter for his restriction map data for T28H11, Denise Ashworth for synthesis of oligonucleotides, David Coates and Andrew Lynch for the database searches, David Coates for the shared molecular biology facilities, and Andrew Lynch, Alison Messom and Jane Young for comments on the manuscript. This work would not have been possible without grants received from The Royal Society, the Medical Research Council and the Human Frontiers Science Program Organization.

REFERENCES

- Akrigg, D., Attwood, T. K., Bleasby, A. J., Findlay, J. B. C., North, A. C. T., Maughan, N. A., Parry-Smith, D. J., Perkins, D. N. and Wootton, J. C. (1992). SERPENT - An information storage and analysis resource for protein sequences. *CABIOS* **8**, 295-296.
- Allen, N. D., Cran, D. G., Barton, S. C., Hettle, S., Reik, W. and Surani, M. A. (1988). Transgenes as probes for active chromosomal domains in mouse development. *Nature* **333**, 852-855.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
- Benit, P., Chanet, R., Fabre, F., Faye, G., Fukuhara, H. and Sor, F. (1992). Sequence of the *sup61-RAD18* region on chromosome III of *Saccharomyces cerevisiae*. *Yeast* **8**, 147-153.
- Blumenthal, T. and Thomas, J. (1988). *Cis* and *trans* mRNA splicing in *C. elegans*. *Trends Genet.* **4**, 305-8.
- Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71-94.
- Coulson, A., Kozono, Y., Lutterbach, B., Shownkeen, R., Sulston, J. and Waterston, R. (1991). YACS and the *C. elegans* genome. *Bioessays* **13**, 413-417.
- Cowan, A. E. and McIntosh, J. R. (1985). Mapping the distribution of differentiation potential for intestine, muscle and hypodermis during early development in *Caenorhabditis elegans*. *Cell* **41**, 923-32.
- Dirksen, M. L. and Jamrich, M. (1992). A novel, activin-inducible, blastopore lip-specific gene of *Xenopus laevis* contains a fork head DNA binding domain. *Genes Dev.* **6**, 599-608.
- Edgar, L. G. and McGhee, J. D. (1988). DNA synthesis and the control of embryonic gene expression in *C. elegans*. *Cell* **53**, 589-99.
- Fields, C. (1990). Information content of *Caenorhabditis elegans* splice site sequences varies with intron length. *Nucl. Acids Res.* **18**, 1509-1512.
- Frohman, M. A. (1990). RACE: Rapid amplification of cDNA ends. In *PCR*

- Protocols: A Guide to Methods and Applications* (ed. M. A. Innis, Gelfand, D. H., Sninsky, J. J. and White, T. J.), pp. 28-38. San Diego: Academic Press Inc.
- Gossler, A., Joyner, A. L., Rossant, J. and Skarnes, W. C.** (1989). Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463-465.
- Grossniklaus, U., Pearson, R. K. and Gehring, W. J.** (1992). The *Drosophila sloppy paired* locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Genes Dev.* **6**, 1030-1051.
- Hacker, U., Grossniklaus, U., Gehring, W. J. and Jackle, H.** (1992). Developmentally regulated *Drosophila* gene family encoding the fork head domain. *Proc. natl. Acad. Sci. USA.* **89**, 8754-8758.
- Hird, S. N. and White, J. G.** (1993). Cortical and cytoplasmic flow polarity in early embryonic cells of *Caenorhabditis elegans*. *J. Cell Biol.* **121**, 1343-55.
- Hope, I. A.** (1991). Promoter trapping in *Caenorhabditis elegans*. *Development* **113**, 399-408.
- Kimble, J. and Hirsh, D.** (1979). The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. *Dev. Biol.* **70**, 396-417.
- Knochel, S., Lef, J., Clement, J., Klocke, B., Hille, S., Koster, M. and Knochel, W.** (1992). Activin A induced expression of a *fork head* related gene in posterior chordamesoderm (notochord) of *Xenopus laevis* embryos. *Mech. Dev.* **38**, 157-165.
- Krause, M. and Hirsh, D.** (1987). A *trans*-spliced leader sequence on actin mRNA in *C.elegans*. *Cell* **49**, 753-761.
- Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, R. H. and Darnell, J. E.** (1990). HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes Dev.* **4**, 1427-1436.
- Lai, E., Prezioso, V. R., Tao, W., Chen, W. S. and Darnell, J. E.** (1991). Hepatocyte nuclear factor 3 α belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene *fork head*. *Genes Dev.* **5**, 416-427.
- Laufer, J. S., Bazzicalupo, P. and Wood, W. B.** (1980). Segregation of developmental potential in early embryos of *Caenorhabditis elegans*. *Cell* **19**, 569-77.
- Li, C., Lai, C., Sigman, D. S. and Gaynor, R. B.** (1991). Cloning of a cellular factor, interleukin binding factor, that binds to NFAT-like motifs in the human immunodeficiency virus long terminal repeat. *Proc. natl. Acad. Sci. USA* **88**, 7739-7743.
- McPherson, M. J., Oliver, R. P. and Gurr, S. J.** (1992). The polymerase chain reaction. In *Molecular Plant Pathology* (ed. S. J. Gurr, McPherson, M.J., and Bowles, D.J.), pp. 123-146. Oxford: IRL Press.
- Miller, L. M., Gallegos, M. E., Morisseau, B. A. and Kim, S. K.** (1993). *lin-31*, a *C. elegans* HNF-3/*fork head* transcription factor homolog, specifies three alternative cell fates in vulval development. *Genes Dev.* **7**, 933-947.
- Nonet, M. L. and Meyer, B. J.** (1991). Early aspects of *Caenorhabditis elegans* sex determination and dosage compensation are regulated by a zinc-finger protein. *Nature* **351**, 65-68.
- O'Kane, C. J. and Gehring, W. J.** (1987). Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. natl. Acad. Sci. USA* **84**, 9123-9127.
- Priess, J. R. and Thomson, J. N.** (1987). Cellular interactions in early *C. elegans* embryos. *Cell* **48**, 241-250.
- Schnabel, R.** (1991). Cellular interactions involved in the determination of the early *C. elegans* embryo. *Mech. Dev.* **34**, 85-100.
- Strome, S. and Wood, W. B.** (1983). Generation of asymmetry and segregation of germ-line granules in early *C. elegans* embryos. *Cell* **35**, 15-25.
- 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. and Hodgkin, J.** (1988). Methods. In *The nematode Caenorhabditis elegans*. (ed. Wood, W.B.), pp. 587-606. New York: Cold Spring Harbor Laboratory.
- Sulston, J. E. and Horvitz, H. R.** (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110-156.
- Tao, W. and Lai, E.** (1992). Telencephalon-restricted expression of BF-1, a new member of the HNF-3/*fork head* gene family, in the developing rat brain. *Neuron*. **8**, 957-966.
- Weigel, D., Jurgens, G., Kuttner, F., Seifert, E. and Jackle, H.** (1989). The homeotic gene *fork head* encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. *Cell* **57**, 645-658.
- Weigel, D. and Jackle, H.** (1990). The *fork head* domain: A novel DNA binding motif of eukaryotic transcription factors? *Cell* **63**, 455-456.
- Wood, W. B.** (1991). Evidence from reversal of handedness in *C. elegans* embryos for early cell interactions determining cell fates. *Nature* **349**, 536-538.
- Yanisch-Perron, C., Vieira, J. and Messing, J.** (1985). Improved cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.
- Young, J. M. and Hope, I. A.** (1993). Molecular markers of differentiation in *Caenorhabditis elegans* obtained by promoter trapping. *Dev. Dynamics* **196**, 124-132.

(Accepted 2 December 1993)

Note added in proof

The nucleic acid sequences described in this paper have been entered into the EMBL data library. The accession numbers are: Z28375, Z28376 and Z28377.