Development 121, 1253-1262 (1995)
Printed in Great Britain © The Company of Biologists Limited 1995

The *C. elegans* neuronally expressed homeobox gene *ceh-10* is closely related to genes expressed in the vertebrate eye

Pia C. Svendsen and James D. McGhee*

Department of Medical Biochemistry, University of Calgary, Health Sciences Centre, 3330 Hospital Drive, NW, Calgary, Alberta, CANADA T2N 4N1

*Author for correspondence (e-mail: jmcghee@acs.ucalgary.ca)

SUMMARY

We describe the homeobox gene *ceh-10* from the nematode *Caenorhabditis elegans*. The homeodomain of *ceh-10* is closely related to the homeodomains of two genes recently cloned from the vertebrate retina, *Chx10* from mice and *Vsx-1* from goldfish. We show that the sequence conservation extends well beyond the homeodomain and includes a region (named the CVC domain) of roughly 60 amino acids immediately C-terminal to the homeodomain. As assayed in transgenic worms, the promoter region of *ceh-10* directs expression of a *lacZ* reporter gene to a small number of neurons. We draw a parallel between the bipolar cells of the inner nuclear layer of the vertebrate retina, which express *Chx10* and *Vsx-1*, and an interneuron in *C. elegans* called AIY, which expresses *ceh-10*. AIY receives synaptic input from a sensory cell, just as do bipolar cells of the vertebrate retina. In *C. elegans*, the sensory cell AFD is not known to be photosensitive but is known to be thermosensitive; moreover, a cell with similar position in the amphids of other nematodes has been suggested indeed to be photosensitive. Our results emphasize the highly conserved nature of sensory regulatory mechanisms and suggest one way in which photosensitive organelles might have originated in evolution.

Key words: *C. elegans*, *ceh-10*, homeobox, neural expression, sensory interneurons, photosensitivity

INTRODUCTION

The genes coding for the homeodomain-containing proteins, i.e. homeobox genes, have been classified by chromosomal arrangement either as 'complex' or 'dispersed' (Bürglin, 1994; Gehring et al., 1994). Homeobox genes of the complex class are organized as chromosomal clusters and appear to be involved in specifying positional identity along the body axis; they display the intriguing relationship between conserved homeodomain sequence, chromosomal position and expression pattern in the developing embryo. The homeotic gene clusters have been best described in *Drosophila* and in mice (McGinnis and Krumlauf, 1992; Krumlauf, 1994; Lawrence and Morata, 1994) but a small set of apparently homologous genes are arranged in a similar cluster in the nematode *C. elegans* (Wang et al., 1993; Kenyon, 1994).

For the large class of homeobox genes not organized in homeotic clusters but rather dispersed throughout the genome, there is increasing evidence that conserved homeodomain sequence also indicates a conserved function. A striking example of such conservation has recently been reported by Quiring et al. (1994). They showed that the *Drosophila* gene *eyeless* is highly related in protein sequence to the mammalian *Pax-6* genes (Hill et al., 1991; Ton et al., 1991; Walther and Gruss, 1991). Remarkably, both *eyeless* and *Pax-6* are expressed in the developing eye of the respective embryos and mutation severely affects eye formation (Hill et al., 1991, Ton et al., 1991; Walther and Gruss, 1991; Glaser et al., 1994; Quiring et al., 1994).

In the present paper, we describe a second set of highly conserved homeodomain genes that may also be involved in the formation or function of photoreceptors. The *C. elegans ceh-10* gene was previously cloned by hybridization to a generic homeobox oligonucleotide probe (Hawkins and McGhee, 1990) but nothing was known about *ceh-10* function. Recently, two genes with homeodomains highly related to *ceh-10* have been cloned from vertebrates: *Chx10* from mice (Liu et al., 1994) and *Vsx-1* from goldfish (Levine et al., 1994). The intriguing observation is that both *Chx10* and *Vsx-1* are expressed preponderantly in the vertebrate retina (Levine et al., 1994; Liu et al., 1994). *C. elegans* has no obvious photoreceptors and is not phototactic but does have a photoresponse (Burr, 1985). It thus became of great interest to explore further whether *ceh-10* can indeed be considered as the homolog of the *Chx10*/*Vsx-1* genes and to determine what kinds of cells express *ceh-10* in *C. elegans*.

MATERIALS AND METHODS

Worm strains and culture methods

C. elegans were cultured as described by Brenner (1974). The strain used for transformation with β-galactosidase fusion vectors was the standard wild-type strain N2.

Genomic and cDNA clones

DNA manipulations were done as described by Sambrook et al. (1989). Cosmids in the region of *ceh-10* were identified using a lambda clone (JM#L1003) containing the *ceh-10* homeobox sequence (Hawkins and McGhee, 1990). Southern blots of cosmid restriction fragments probed with a 0.9 kb *Bam*HI *ceh-10* homeobox clone (Hawkins and McGhee, 1990), and sequencing of positive fragments, identified a 3.4 kb *Eco*RI fragment containing the entire coding region of the *ceh-10* gene. This fragment includes 1.1 kb of sequence 5′ to the presumed translational start codon. A 6.4 kb *Cla*I fragment contains an additional 2.1 kb of 5′ flanking sequence.

5′- and 3′-RACE PCR cDNA clones of the *ceh-10* message were amplified as described by Frohman et al. (1988) from poly(A)+ RNA isolated from a mixed stage *C. elegans* population. The nucleotide sequence of the 0.9 kb *Bam*HI genomic clone containing the *ceh-10* homeobox sequence (Hawkins and McGhee, 1990) was used to design *ceh-10*-specific primers that would produce 5′-RACE and 3′-RACE products overlapping a *Xho*I restriction site in the homeobox. Primers for 5′-RACE were as follows: 5′-dGCCG-TACCATTGCTCCATATAGCC-3′ (nucleotides 2262-2285 of genomic sequence shown in Fig. 1C) and 5′-dCTGCCATTATG-GTACTTTTTCCCC-3′ (nucleotides 2232-2255); the primer for 3′-RACE was 5′-dCGGAGCCCCAGACTCATGTCAGCT-3′ (nucleotides 1568-1591).

β-galactosidase fusion vectors

The *lacZ* translational fusion vector pPD16.51 (Fire et al., 1990) was used to make constructs pS310 and pS311. pS310 has a 3.8 kb *Cla*I/*Xho*I fragment containing 3.2 kb of *ceh-10* 5′ flanking sequence and the coding region for amino acids 1-151 of CEH-10 fused upstream of a nuclear localization signal (NLS) and the *E. coli lacZ* gene. pS311 is a deletion product of pS310 from which the coding region for amino acids 62-151 of CEH-10 (which includes the putative NLS, the first 16 amino acids of the homeodomain as well as intron 2) along with the NLS of vector pPD16.51 have been removed.

Germline transformation

Constructs (pS310 at 140 μg/ml or pS311 at 80 μg/ml) were coinjected with the pRF4 plasmid (at 60 μg/ml). pRF4 contains a *rol-6* allele conferring a semi-dominant 'roller' phenotype to transformed worms (Mello et al., 1991). Such injections yielded transiently transformed strains carrying extrachromosomal arrays of the transforming DNA. These strains show 50-90% transmission of the roller phenotype. To generate heritably transformed strains with integrated arrays of the transforming DNA, L4 stage worms from two independently transformed strains carrying extrachromosomal arrays of pS310/pRF4 (*pEx1, pEx3*) were subjected to 2400 rads of γ-rays from a ^{137}Cs source. F_2 clones (*pInt1* and *pInt2* from *pEx1*, and *pInt3* from *pEx3*) generating 100% roller progeny were selected as described by Way et al. (1991).

Histochemical staining for β-galactosidase activity

Embryos were prepared for histochemical staining by permeabilization and fixation as described in Edgar and McGhee (1986). Postembryonic stages were prepared as follows. Larvae and adults were washed off plates using M9 buffer; after one wash in the same buffer, worm pellets were resuspended in 2.25% paraformaldehyde (freshly prepared), frozen on dry ice, then thawed quickly at 37°C; worms were fixed for 5 minutes at 20°C followed by three washes with phosphate buffer (pH 7.2; Edgar and McGhee, 1986). Staining solution (Fire, 1992) was added after fixation; incubation at room temperature (20°C) produced the best staining results. Staining was usually complete after 2 hours and was stopped by three washes with PBS buffer. The first of these washes included 4,6-diamino-2-phenylindole (DAPI) at 1 μg/ml to allow visualization of DNA.

Immunostaining for β-galactosidase protein

Adult transgenic worms (integrated strain *pInt1*) were permeabilized and reacted with a mouse monoclonal antibody against β-galactosidase (Promega), followed by treatment with a rhodamine-labelled goat anti-mouse IgG (Jackson Immunoresearch), essentially as described by Siddiqui et al. (1989). The first wash after incubation with the secondary antibody included 1 μg/ml DAPI in PBS.

RESULTS

Characterization of the *C. elegans ceh-10* gene

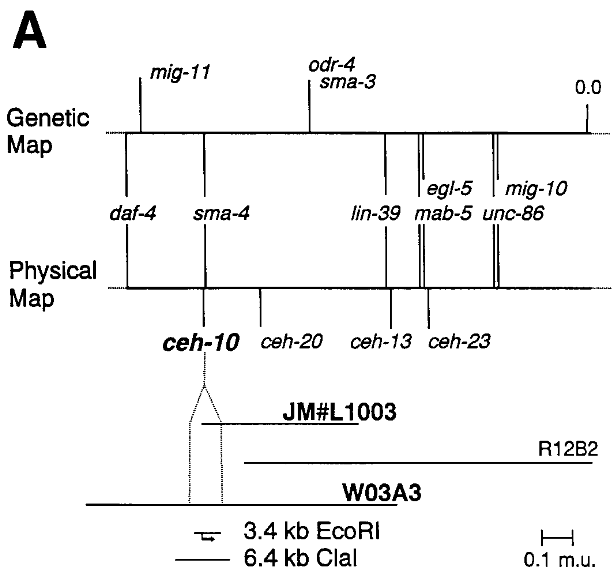
Fig. 1A shows the position of *ceh-10* on the current physical map of the *C. elegans* genome (Coulson et al., 1988), aligned with the *C. elegans* genetic map. *ceh-10* is located close to the middle of chromosome III, approximately 2 mb to the left of the *C. elegans* homeotic cluster. The sequence of a 3.4 kb *Eco*RI fragment obtained from cosmid W03A3 and containing the complete *ceh-10* gene was determined. Two independent *ceh-10* cDNA clones were produced by both 5′- and 3′-RACE (Frohman et al., 1988), using primers within the previously described homeobox sequence (Hawkins and McGhee, 1990). The cDNAs are full length, since they terminate in poly(A) at the 3′ ends and in the SL1 trans-spliced leader sequence (Krause and Hirsh, 1987) at the 5′ ends. The cDNA size (1.2 kb) agrees well with the size of the most abundant mRNA detected on northern blots (Hawkins and McGhee, 1990). Comparison between the *ceh-10* genomic and cDNA sequences indicates four introns, as shown schematically in Fig. 1B. The genomic sequence, the cDNA sequence and the predicted *ceh-10* protein sequence are aligned on Fig. 1C.

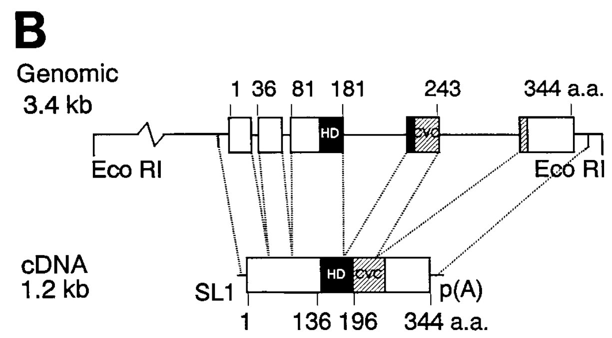
The *ceh-10* protein is predicted to be 344 amino acids long, with the homeodomain towards the middle. *ceh-10* belongs to the Paired-like class of homeobox genes (Bürglin, 1994; Gehring et al., 1994) because: (i) there is no 'paired box'; (ii) a conserved region to the N-terminal side of the homeodomain is absent, and; (iii) the residue at position 50 of the homeodomain (residue 185 of the protein) is glutamine.

ceh-10, Vsx-1 and Chx10 share a highly conserved protein domain immediately downstream of the homeodomain

The most striking feature of the *ceh-10* protein sequence is its high level of sequence identity to the *Chx10/Vsx-1* genes. The regions of highest amino acid conservation shared among the three proteins are aligned on Fig. 2A. The homeodomains of *Chx10* and *Vsx-1* are 95% identical to each other and show 82% and 78% amino acid identity respectively (92% similarity, calculated as described by Altschul et al., 1990) to the homeodomain of *ceh-10*. The region of high sequence conservation extends a further 57-61 amino acids toward the C terminus. In this extended domain, the *ceh-10* protein shows 74% amino acid identity to *Chx10* and 68% amino acid identity to *Vsx-1* (82% and 81% similarity, respectively). We suggest that this domain be referred to as the CVC domain, standing for *Chx10*/*Vsx-1* and *ceh-10*.

The *ceh-10*, *Chx10* and *Vsx-1* genes share additional regions of similarity outside of the homeodomain and the CVC domain. As shown in Fig. 2B, the sequence (6 matches in 8 residues) and position of the 'octapeptide' motif (Bürglin, 1994; Gehring et al., 1994), as well as a serine/threonine-rich





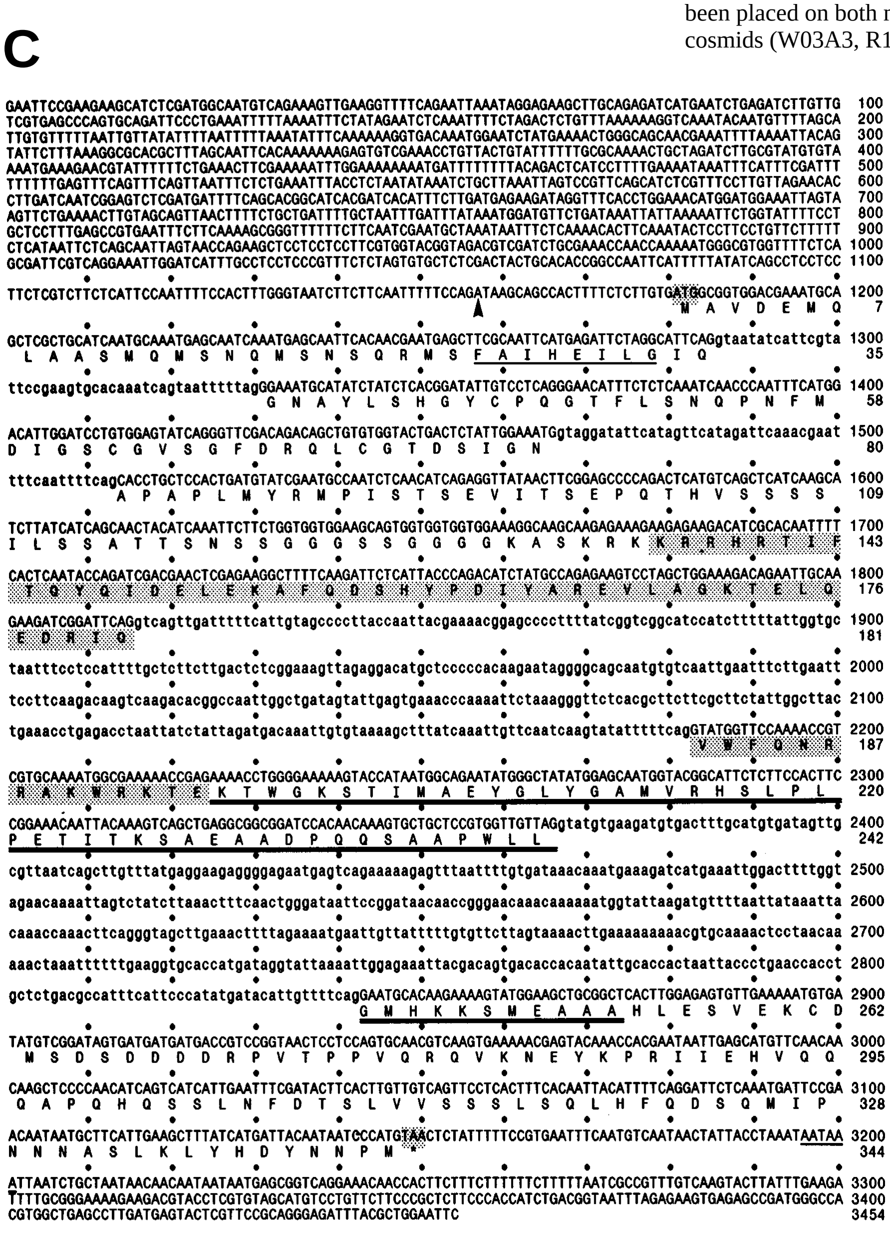


Fig. 1. Characterization of the *C. elegans ceh-10* gene. (A) Alignment of the *C. elegans* physical and genetic maps in the region of the *ceh-10* gene. The region shown is left of centre on linkage group (LG) III (middle of LGIII is indicated by 0.0 above the genetic map; the scale bar indicates distance in map units (m. u.)). Genes shown above the genetic map have only been mapped genetically; genes shown below the physical map have only been mapped physically; genes between have been placed on both maps. Below the physical map are two overlapping cosmids (W03A3, R12B2) that were identified from the cosmid library using the original lambda clone JM#L1003 as probe (Hawkins and McGhee, 1990); only W03A3 contains the 5′ end of the *ceh-10* gene and this orients the transcriptional direction (rightward, towards centre of LGIII). Subclones containing the entire *ceh-10* coding region are shown below the cosmids. The bent arrow extending from the 3.4 kb *Eco*RI fragment indicates *ceh-10* transcription. (B) A schematic drawing compares the structures of the *ceh-10* genomic and cDNA clones. Numbers indicate residue positions in the deduced amino acid sequence. The 3.4 kb *Eco*RI genomic clone includes the entire cDNA sequence split by four introns as well as 1.1 kb and 0.3 kb of upstream and downstream non-coding region respectively. Intron 3 is found at a conserved position in the homeodomain of Paired class homeobox genes while intron 4 interrupts the coding sequence of the newly identified CVC domain (see text). (C) The *ceh-10* genomic and cDNA sequences are aligned. The cDNA sequence starts at the arrowhead, just at the acceptor of the trans-spliced SL1 leader. The actual transcription start site is not known. The deduced amino acid sequence is assumed to start at the first in-frame methionine codon downstream of the trans-splice site. The homeodomain is highlighted and the CVC domain is indicated by thick underlining. The octapeptide (amino acids 26-32) is underlined. The poly(A) tail is added after nucleotides 3214-16 (cDNA clone pS414) or after 3211-3213 (cDNA clone pS416). The nucleotide sequence of clone pS414 agreed with the genomic sequence; the nucleotide sequence of clone pS416 showed an A to G change at position 1269 (glutamic acid to glycine), presumably due to an error in PCR amplification. The sequence of *ceh-10* has been submitted to GENBANK/EMBL data libraries (accession number U19995).

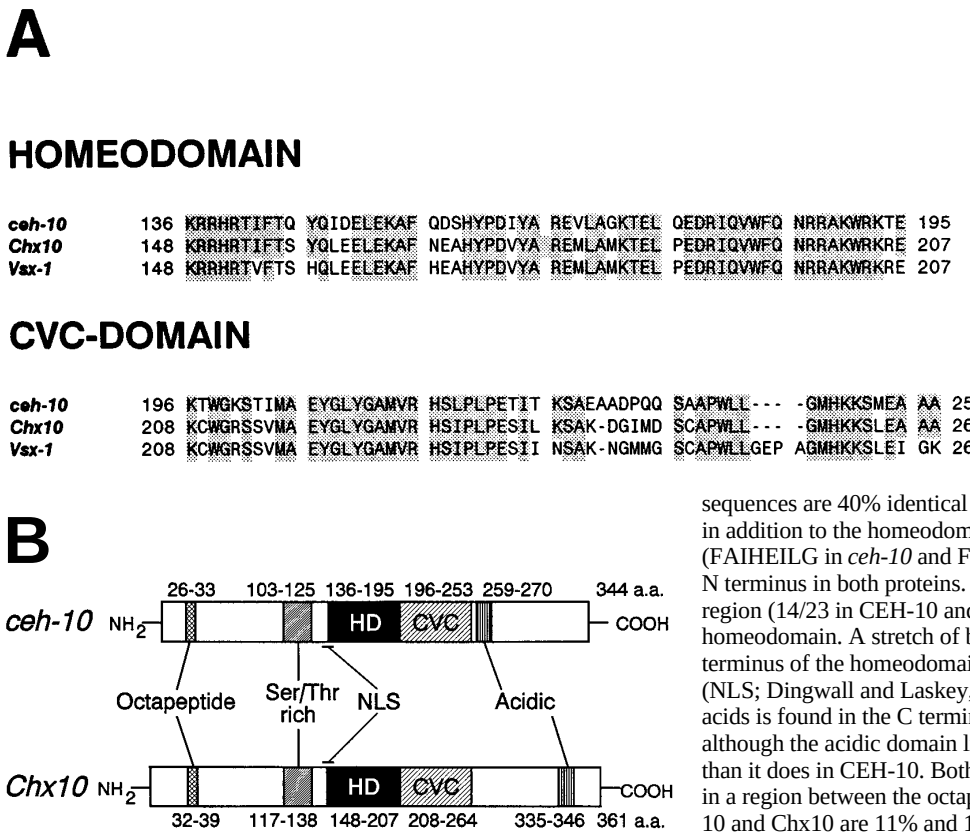


Fig. 2. Conserved features of the *ceh-10* gene. (A) Amino acid sequence alignment of the homeodomain and the CVC domain of *ceh-10*, *Chx10* (Liu et al., 1994) and *Vsx-1* (Levine et al., 1994) are shown. The numbers refer to the positions of these contiguous domains within their respective protein sequences. The highlighted regions indicate amino acids identical to the corresponding residues in *ceh-10*. The alignment suggests a single amino acid insertion in the *ceh-10* CVC domain (position 230), indicated by dashes at the same positions in *Chx10* and *Vsx-1*. *(B) Chx10* and *ceh-10* protein sequences are 40% identical overall and share several conserved features in addition to the homeodomain and CVC domain. The octapeptide (FAIHEILG in *ceh-10* and FGIQEILG in *Chx-10)* is found close to the N terminus in both proteins. Both proteins have a serine/threonine-rich region (14/23 in CEH-10 and 11/22 in Chx10) N-terminal to the homeodomain. A stretch of basic amino acids overlapping the N terminus of the homeodomain is a potential nuclear localization signal (NLS; Dingwall and Laskey, 1991). A small region high in acidic amino acids is found in the C terminus (7/12 in CEH-10, 8/12 in Chx10) although the acidic domain lies more towards the C terminus in Chx10 than it does in CEH-10. Both proteins are also proline-rich; for example, in a region between the octapeptide and the Ser/Thr-rich regions, CEH-10 and Chx10 are 11% and 15% prolines respectively.

region, a putative nuclear localization signal and an acidic stretch of amino acids are all well conserved between *ceh-10* and *Chx10*. The *Vsx-1* gene shows comparable regions of similarity (not shown). Interestingly, a four amino acid insertion in the CVC domain (residue 254) of *Vsx-1* aligns with the position of an intron in *ceh-10*.

A partial sequence of a homeobox gene (*Vsx-2*), closely related to, but distinct from, *Vsx-1*, has also been cloned from the goldfish retina (Levine and Schechter, 1993). Within both the homeodomain and the CVC domain, *ceh-10* shows an even higher sequence identity to *Vsx-2* than it does to *Vsx-1* (E. Levine and N. Schechter, personal communication). Highly related genes (both in the homeodomain and the CVC domain) have also been cloned from the retina of zebrafish (E. Boncinelli, Istituto San Raffaele, Milan; personal communication) and chicken (Y. Nakano, Mitsubishi Kasei Institute of Life Sciences, Tokyo; personal communication).

ceh-10 is expressed in a small number of *C. elegans* neurons

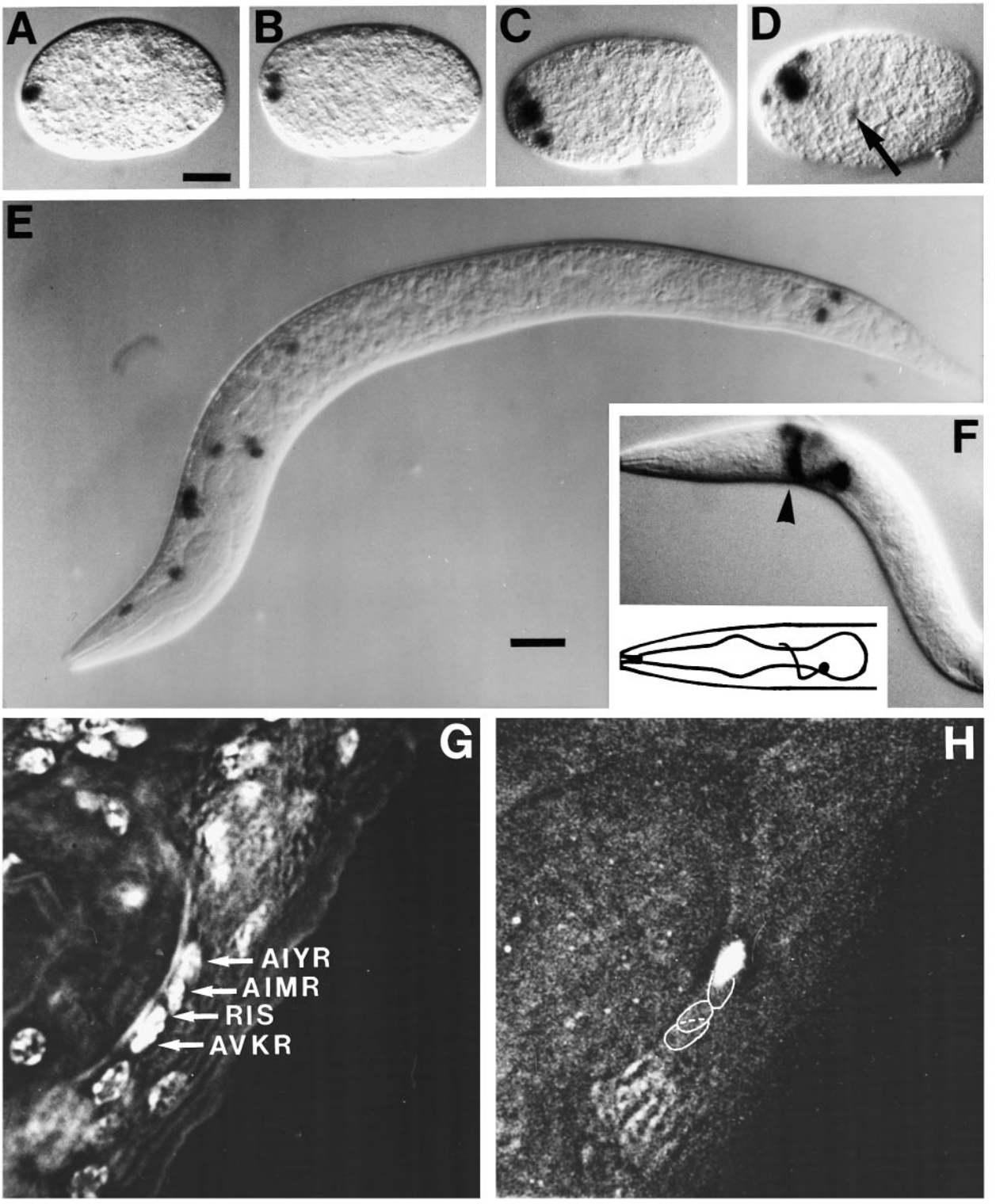
To determine where *ceh-10* is expressed in *C. elegans*, 3.2 kb of *ceh-10* 5′-flanking region and part of the *ceh-10* coding region were fused to a *lacZ* reporter gene (Fire et al., 1990). A number of independent transgenic lines were generated (Mello et al., 1991) and all produced basically similar expression patterns. Transforming arrays in two independent lines were integrated into the genome (Way et al., 1991), again with no obvious change in expression pattern.

The large majority of *ceh-10* mRNA is found in the embryo (Hawkins and McGhee, 1990) and likewise the majority of *ceh-10* directed reporter gene expression is also found within the embryo. As shown in Fig. 3A, the earliest point at which *ceh-10/lacZ* expression can be detected is in a single nucleus

Fig. 3. Expression of *ceh-10*. (A-E) The expression pattern of pS310 *ceh-10/lacZ* trans-genes (Fire et al., 1990) in embryos and larvae (scale bar = 10 μm). Anterior is left. (A) The premorphogenesis embryo shows the first detectable β-galactosidase staining, a single nucleus at the extreme anterior end of the embryo. This embryo has approximately 350 cells (determined by comparing nuclear patterns revealed by DAPI co-staining to line drawings of live embryos; Sulston et al., 1983). (B) Two cells stain in late proliferation stage embryos. (C) The comma stage embryo has six nuclei staining at the anterior end of the embryo. As the embryo lengthens two more nuclei appear in the staining pattern; (D) the 1.5-fold embryo has six anterior staining nuclei and two staining nuclei (arrow) near the excretory pore (only one of these two nuclei is seen in this plane of focus). The position of these two nuclei is consistent with their being AIY cells (see text). (E) An L1 larva is shown. The identity of the stained nuclei in L1 stage worms (see text) was determined by Dr C. Bargmann (UCSF, personal communication; cells identified as occasionally staining for β-galactosidase activity include sensory neuron support cells CEPshDL and OLQso). Two to four cells stain in the tail anterior to the pre-anal ganglion, but their identity has not yet been determined. (F) An example of process staining of the AIY cell is shown; the arrowhead marks the position of the ventral side of the worm at the nerve ring. (G,H) Transgenic worms (*pInt 1*) costained with the DNA stain DAPI and immunostained for β-galactosidase protein. (G) A confocal image (0.5 μm slice thickness) of DAPI stained nuclei in the right ventral ganglion (see Fig. 3 in White et al., 1986). (H) Rhodamine fluorescence detected in the same image plane. Outlines of the four nuclei labelled in G are superimposed on the rhodamine image in H, identifying AIYR as the *ceh-10*-expressing cell.

at the extreme anterior pole of the embryo, at mid-proliferation stage when the embryo has approximately 350 cells (as determined by DAPI fluorescence; not shown). In late proliferation stage embryos, two nuclei (Fig. 3B) and then four nuclei (not shown), all at the embryo anterior, express the transgene. As the embryo begins morphogenesis, six nuclei, all at the very anterior pole, stain for β-galactosidase activity (Fig. 3C). By the 1.5-fold stage (Fig. 3D), the embryo has eight staining nuclei, six at the anterior and two additional cells (see arrow on Fig. 3D) just ventral to the posterior bulb of the pharynx. By the time the embryo has completed morphogenesis, up to twelve staining nuclei can be detected.

The positions of the staining cells in embryos suggest that *ceh-10* is expressed in neurons (Sulston et al., 1983). The anatomy of the *C. elegans* nervous system has been comprehensively described, both in the first (L1) larval stage and in



adults (Ward et al., 1975; Ware et al., 1975; Sulston and Horvitz, 1977; White et al., 1986; Chalfie and White, 1988; Bargmann and Horvitz, 1991). It is in these post-embryonic stages that identification of the *ceh-10*-expressing cells should be the most feasible and most certain. Fig. 3E shows a representative transgenic L1 larva stained for β-galactosidase activity. *ceh-10/lacZ* expression in L1 larvae is somewhat variable, in that staining in a particular cell is not always seen in every larva; the likely reason is that the level of *ceh-10* mRNA declines after hatching (Hawkins and McGhee, 1990) and β-galactosidase staining in L1 larvae presumably results from enzyme perdurance. Most *ceh-10/lacZ* staining occurs in the region around the nerve ring, which is dense with the nuclei and processes of neurons. Hence, the incomplete penetrance of staining turns out to be a practical advantage, in that staining nuclei overlap less frequently and are easier to identify. The expressing anterior nuclei were kindly identified by Dr C. Bargmann (San Francisco) as: AIYL/R (interneurons), CEPDL/R (mechanosensory neurons), RID (motor neuron), ALA (lateral neuron), RMED (nerve ring motor neuron), AINL/R (interneurons) and AVJL/R (ventral neurons). We show below that *ceh-10* is also expressed in the CAN cells (the excretory Canal Associated Neuron). Thus *ceh-10* is expressed in the same type of cells that express vertebrate *Chx10/Vsx-1*, namely neurons.

An antibody to CEH-10 is not yet available to corroborate the β-galactosidase staining pattern. However, a number of studies have compared the cells that express various *lacZ* reporter genes with cells in which the endogenous gene product can be detected immunologically. Agreement between the two expression patterns is generally excellent, both for genes encoding structural proteins and genes encoding regulatory

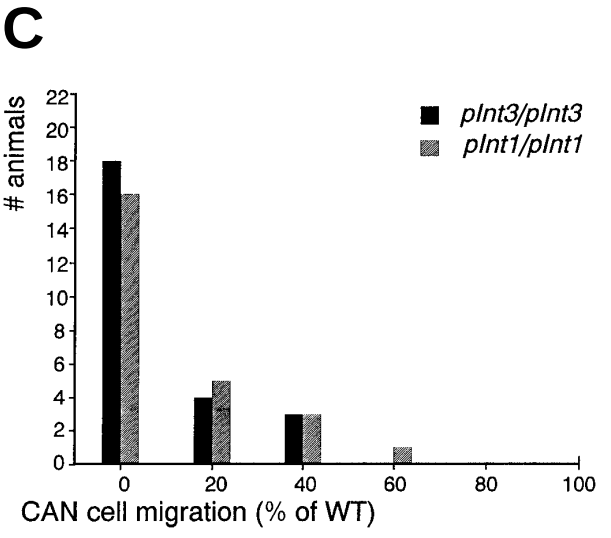


Fig. 4. Wit phenotype and CAN nuclear migration in *ceh-10/lacZ* transgenic worms. (A) The Wit phenotype seen in a *ceh-10/lacZ* transgenic adult worm (integrated strain *pInt*1) is shown. (B) An example of a transgenic L1 larva stained for β-galactosidase is shown with a CAN nucleus indicated by the arrowhead; scale bar = 10 μm. Homozygous worms (*pInt1/pInt1*, *pInt3/pInt3*) were scored for the extent of migration of CAN nuclei, relative to their wild-type migration distance, using β-galactosidase staining. In many cases, only one of the two CAN nuclei had migrated far enough for its position to be scored. (C) The histogram shows the position of the CAN nucleus that had migrated furthest in individual worms, expressed as a percentage of the pharynx to gonad primordium distance (*n*=25 in each class; *pInt1* and *pInt3* are independent strains carrying integrated arrays of pS310 and pRF4).

factors (Krause et al., 1990; Hamelin et al., 1992; Okkema and Fire, 1994). We now go on to discuss one particular pair of cells that express *ceh-10*, the AIY(L/R) interneurons.

ceh-10 is expressed in the AIY interneurons

Are there any *ceh-10*-expressing cells in *C. elegans* that could possibly be homologous to the cells of the vertebrate retina that express *Chx10* and *Vsx-1*? We suggest that the AIY(L/R) pair of interneurons may be such cells.

We first provide unambiguous identification of AIY as a *ceh-10*-expressing cell. Our first approach was to modify the *ceh-10* reporter gene construct by deleting the nuclear localization signal, such that the cell body and neural processes rather than nuclei were labelled. An example of an L1 larva transgenic for such a modified construct is shown in Fig. 3F; the position of the indicated cell body and the extending process is completely consistent with its identity as AIY; (compare to the line drawing taken from White et al., 1986). However, a neighbouring neuron AIM has a process with a roughly similar trajectory to that of AIY (White et al., 1986) and a suitably cautious interpretation of Fig. 3F is that the *ceh-10*-expressing neuron in this region is either AIY or AIM.

To provide unambiguous evidence that it is indeed AIY that expresses *ceh-10*, we took advantage of the increased spatial resolution provided by confocal microscopy, coupled to the fact that sufficient β-galactosidase protein perdures into the adult worm to allow expressing nuclei to be detected by indirect immunofluorescence. Fig. 3G shows a confocal image of a DAPI-stained transgenic adult worm, focusing on four nuclei AIYR, AIMR, RIS and AVKR in the ventral ganglion, slightly ventral and towards the right of the pharynx (White et al., 1986). Fig. 3H shows the same focal plane but now imaged by a rhodamine-labelled antibody to β-galactosidase. The expressing nucleus can be unambiguously aligned with AIY, not AIM. Furthermore, neighbouring focal planes (not shown) demonstrate that the *ceh-10*-expressing nucleus lies more lateral than does the neighbouring non-expressing nucleus, providing further proof that the expressing nucleus is indeed AIY (White et al., 1986). The implications of *ceh-10* expression in the AIY cells will be considered in the Discussion.

ceh-10 is also expressed in the CAN cells

We first suspected CAN (Canal Associated Neuron) cell expression of *ceh-10* because every strain of worms transformed with the *ceh-10/lacZ* fusion construct produced individuals showing a 'withered tail' (Wit) phenotype, as shown in Fig. 4A. This phenotype was observed with multiple independently produced strains, with two different fusion constructs and with strains that had been extensively outcrossed after integration of the transforming array. The Wit phenotype has previously been shown to result from ablation, improper function, or incomplete migration of the CAN cell (Manser and Wood, 1990); the distinctive Wit morphology has been proposed to result from a defective association between the CAN cell and the excretory canal, with consequent effects on osmoregulation (Hedgecock et al., 1987; Manser and Wood, 1990).

The CAN nucleus normally undergoes posterior migration late in embryogenesis from its birthplace in the head to a defined position in midbody (Hedgecock et al., 1987). In untransformed control larvae, CAN nuclei could be detected using Nomarski optics but in ten transgenic larvae examined, only one showed a CAN nucleus at the expected position. Moreover, ß-galactosidase staining nuclei can be found in transgenic larvae at variable positions between the pharynx (roughly the origin of CAN migration) and the gonadal precursor cells (roughly the termination of CAN migration); an example of such a staining nucleus is indicated in Fig. 4B and a histogram summarizing the incomplete CAN migrations found in two transformed strains is shown in Fig. 4C.

Worms heterozygous for the (integrated) transforming array provide further evidence for a link between *ceh-10* expression and CAN cell migration: (i) heterozygous worms show a greatly decreased incidence of the Wit phenotype (<5%, compared to nearly 100% in adult homozygotes); (ii) in roughly half of the heterozygous worms examined, at least one CAN nucleus could be identified at the wild-type position, and; (iii) β-galactosidase staining nuclei could be detected at a significantly more posterior position than in the homozygote.

In summary, the *ceh-10/lacZ* reporter gene is expressed in the CAN cells but apparently perturbs normal CAN cell function. This fact will be used below to suggest where *ceh-10* lies in the regulatory hierarchy.

DISCUSSION

The conserved homeodomains and CVC domains suggest that *ceh-10* is the *C. elegans* homolog of the vertebrate *Chx10/Vsx-1* genes

In general, homeodomains found in *C. elegans* genes within the homeotic cluster show 50-75% amino acid identity with their presumed homologs in vertebrates (Bürglin, 1994; Gehring et al., 1994). The highest degree of conservation so far reported between *C. elegans* and vertebrate homeodomains is 90% (Bürglin and Ruvkun, 1992). Thus, by our reckoning, the sequence conservation between *ceh-10* and *Chx10/Vsx-1* is the second highest yet found between nematode and vertebrate homeodomains. The implication is that the function of *ceh-10* and *Chx10/Vsx-1* is also highly conserved.

The argument that *ceh-10* can indeed be regarded as the nematode homolog of *Chx10/Vsx-1* is further strengthened by the presence of the conserved CVC domain, extending some 60 amino acids from the C-terminal boundary of the homeodomain. The function of this domain is not known but two obvious possibilities come to mind. The CVC domain could be involved in binding DNA, as are the PRD- and POU-domains associated with certain classes of homeodomains (Treisman et al., 1991; Dekker et al., 1993). The CVC domain could possibly be involved in protein-protein interactions with another regulatory factor; it is through a region C-terminal to the homeodomain that the regulatory protein MATα2 interacts with its sometimes partner MATa1 (Mak and Johnson, 1993; Vershon and Johnson, 1993).

It has been suggested (Liu et al., 1994) that *aristaless* may be a *Drosophila* homolog of *Chx10*. However, *aristaless* does not contain a CVC domain, suggesting that the true *Drosophila* homolog has yet to be identified.

The *ceh-10*-expressing AIY(R/L) interneurons are possible homologs of the *Chx10/Vsx-1*-expressing bipolar cells of the vertebrate retina

If *ceh-10* can be regarded as the homolog of the *Chx10*/*Vsx-1*

genes, as suggested by the sequence conservation, then can the *ceh-10* expression pattern in the nematode be related to the expression patterns of *Chx10/Vsx-1* in vertebrates? In the developing mouse, *Chx10* is predominantly expressed in the developing eye (Liu et al., 1994). In the adult retina, *Chx10* expression appears confined to cells within the inner nuclear layer, most probably bipolar cells (Liu et al., 1994). *Vsx-1* is also expressed in the inner nuclear layer of the mature goldfish retina, in a subset of cells that presumably includes bipolar cells (and possibly Müller glial cells; Levine et al., 1994). Bipolar cells are interneurons that receive synaptic input from the photoreceptor cells and, in the most direct pathway, synapse onto the ganglion cells that transmit signals to the brain (Tessier-Lavigne, 1991).

The AIY interneurons receive synaptic input from the AFD cells of the *C. elegans* amphids (Ward et al., 1975; Ware et al., 1975; White et al., 1986). There is good evidence that AFD is thermosensitive (Perkins et al., 1986). A more direct proof that AFD is thermosensitive has recently been obtained by laser ablation; ablation of AIY also alters thermal responsiveness (Dr Ikue Mori, Kyushu University; personal communication). Whether AFD is also photosensitive has not yet been addressed but it has been noted that the AFD cell, with its numerous microvillar projections, is similar to the rhabdomeric type of photoreceptor organelles found in arthropods (Burr, 1985). Moreover, in the photophobic marine nematode *Oncholaimus vesicarius*, the photoreceptor has been suggested to be a single neuron that occupies roughly the same position in the amphid as does the AFD cell in the *C. elegans* amphid (Burr and Burr, 1975). Thus, we suggest that the following two phenomena may be homologous: (1) *Chx10* and *Vsx-1* are expressed in vertebrate retina interneurons that receive synaptic input from photoreceptors, and; (2) *ceh-10* is expressed in a nematode interneuron that is a synaptic partner of a sensory cell that is at least potentially photosensitive.

We note several further intriguing comparisons that can be made between the vertebrate cells that express homologs of *ceh-10* and the cells in *C. elegans* that express *ceh-10/lacZ*. Liu et al. (1994) report expression of *Chx10* in a portion of the upper lip in embryonic mouse; *ceh-10/lacZ* expression is high in CEPD neurons, which are sensory neurons whose processes extend to the tip of the *C. elegans* mouth. Both *Chx10* and *ceh-10* are expressed in a small subset of motor neurons. Finally, Y. Nakano (Mitsubishi Kasei Institute of Life Sciences Tokyo, Japan; personal communication) has found that the apparent homolog of *ceh-10* cloned from the chicken retina is also expressed embryonically in a small number of kidney cells; *ceh-10* is expressed in CAN cells, which are intimately associated with the worm osmoregulatory structure, the excretory canal. A more complete and accurate comparison of the expressing cells in the different organisms seems warranted.

ceh-10 may not be a 'master regulatory gene'

It has been suggested that the *Drosophila eyeless* and mammalian *Pax-6* genes operate at a high level in the genetic hierarchy regulating eye development, since mutants do not form eyes (Hill et al., 1991; Ton et al., 1991; Walther and Gruss, 1991; Glaser et al., 1994; Quiring et al., 1994). Where would *Chx10/Vsx-1/ceh-10* fit into the same hierarchy? It can be argued that these genes might operate at a lower level in the regulatory pathway, based on the observations associated with expression of *ceh-10* in the CAN cells.

We suggest that the *ceh-10* gene is normally expressed in CAN cells but that the *ceh-10/lacZ* fusion gene (which in such transgenic arrays can exist in tens to hundreds of copies; see, for example, Mello et al.,1991 and Stringham et al., 1992) causes some antimorphic effect that inhibits correct CAN migration. Possible mechanisms of such an antimorphic effect include the titration of regulatory factors by high copy numbers of promoter elements (see, for example, Johnston and Kucey, 1988; Pai et al., 1992), disrupted binding to DNA, disrupted interaction with other regulatory proteins (especially interesting because Paired class homeodomain proteins have been shown to form homodimers on DNA; Wilson et al., 1993) and some less defined 'trans-sensing' mechanism (Tartof and Henikoff, 1991) possibly arising from pairing between the transgenic array and the normal chromosomal locus. Nonetheless, whatever the inhibitory mechanism, the CAN cell still exists and can be recognized in these worms. In other words, cell identity has not been completely altered but cell function has been disrupted. Hence, *ceh-10* (and by implication, *Chx10* and *vsx-1* as well) might not control other regulatory genes but instead might control 'working' genes, associated with cellular differentiation and allowing the cell to interpret its surroundings. The products of such genes might be receptors or cell adhesion molecules that permit AIY to correctly interact with its AFD partner or that permit the CAN cell to undergo its characteristic migration. The recent identification of mutations in the *Chx10* gene with the ocular retardation phenotype in mice (McInnes et al., 1994) should greatly clarify the regulatory role of *Chx10*.

An alternative interpretation of the CAN cell behaviour is that *ceh-10* is indeed a 'master regulatory gene' but that the antimorphic effect exerted by the fusion construct corresponds to a weak allele. We argue against this interpretation since the Wit phenotype conferred by the *ceh-10* transgene is strong, can be detected even in L1 larvae and is close to 100% penetrant in adults. This phenotype is just as severe as that caused by null alleles in other genes that affect CAN migration (Manser and Wood, 1990). Overexpression of the β-galactosidase reporter protein itself is unlikely to be the cause of the inhibited CAN cell migration: a number of different homeobox gene/*lacZ* fusion constructs have been shown to be expressed in migratory cells (HSNs, QL, QR and CAN cells) with no reported interference with cell migration (Wang et al., 1993).

Mechanisms of evolution

If the same classes of transcription factors are indeed found in diverse types of photosensitive organs (extrapolating from the results of Hill et al., 1991; Ton et al., 1991; Walther and Gruss, 1991; Quiring et al., 1994; and the current paper), then questions are raised about the distinction between convergent and divergent evolution. One could imagine that evolution could appear divergent at a molecular level but convergent at a morphological level. A family of homologous regulatory molecules, all diverging from common ancestors, might control molecular events needed for the formation of photoreceptors in all organisms. At the same time, different organisms might initiate the regulatory pathways in which these factors function by quite different (and non-homologous) tissue-based mechanisms, such as the interaction between different germ

layers or the formation of imaginal discs. It will be especially important to isolate a *Drosophila* homolog of *ceh-10/Vsx-1/Chx10* and to see if it too is expressed in an interneuron connected to a photoreceptor.

We should like to thank Dr C. Bargmann for kindly identifying *ceh-10/lacZ*-expressing cells in the L1 larva and Dr J. Manser for his help in identifying CAN nuclei. We should also like to thank Drs E. Boncinelli, E. Levine, I. Liu, R. McInnes, I. Mori, Y. Nakano and N. Schechter for sharing their results prior to publication, M. Chung for producing several transgenic strains and D. Bazett-Jones and M. Schoel for help with the confocal microscopy. This work was supported by the Medical Research Council of Canada, the Alberta Heritage Foundation for Medical Research, and the Howard Hughes Medical Institute.

REFERENCES

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.

Bargmann, C. I. and Horvitz, H. R. (1991). Control of larval development by chemosensory neurons in *Caenorhabditis elegans*. *Science* **251**, 1243-1246.

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

Burr, A. H. (1985). The photomovement of *Caenorhabditis elegans*, a nematode which lacks ocelli. Proof that the response is to light not radiant heating. *Photochem. Photobiol.* **41**, 577-582.

Burr, A. H. and Burr, C. (1975). The amphid of the nematode *Oncholaimus vesicarius*: Ultrastructural evidence for a dual function as chemoreceptor and photoreceptor. *J. Ultrastuct. Res.* **51**, 1-15.

Bürglin, T. R. (1994). A comprehensive classification of homeobox genes. In *A Guidebook of Homeobox Genes* (ed. D. Duboule), pp. 25-72. Oxford, UK: Oxford University Press.

Bürglin, T. R. and Ruvkun, G. (1992). New motif in PBX genes. *Nature Genet.* **1**, 319-320.

Chalfie, M. and White J. (1988). The nervous system. In *The Nematode Caenorhabditis elegans* (ed. W. B. Wood), pp. 337-391. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Coulson, A. R., Waterston, R., Kiff, J., Sulston, J. and Kohara, Y. (1988). Genome linking with yeast artificial chromosomes. *Nature* **335**, 184-186.

Dekker, N., Cox, M., Boelens, R., Verrijzer, C. P., van der Vliet, P. C. and Kaptein, R. (1993). Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature* **362**, 852-855.

Dingwall, C. and Laskey, R. A. (1991). Nuclear targeting sequences – a consensus? *Trends Biochem, Sci.* **16**, 478-481.

Edgar, L. G. and McGhee, J. D. (1986). Embryonic expression of a gut-specific esterase in *Caenorhabditis elegans*. *Dev. Biol.* **114**, 109-118.

Fire, A. (1992). Histochemical techniques for locating *E. coli* β-galactosidase activity in transgenic organisms. *Gene Analyt. Techn. Applic.* **9**, 151-158.

Fire, A., Harrison, S. W. and Dixon, D. (1990). A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93**, 189-198.

Frohman, M. A., Dush, M. K. and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002.

Gehring, W. J., Affolter, M. and Bürglin, T. (1994). Homeodomain proteins. *Annu. Rev. Biochem.* **63**, 487-526.

Glaser, T., Jepeal, L., Edwards, J. G., Young, S. R., Favor, J. and Maas, R. L. (1994). *PAX6* gene dosage effect in a family with congenital cataracts, aniridia, anopthalmia and central nervous system defects. *Nature Genet.* **7**, 463-471.

Hamelin, M., Scott, I. M., Way, J. C. and Culotti, J. G. (1992). The *mec-7* β-tubulin gene of *Caenorhabditis elegans* is expressed primarily in the touch receptor neurons. *EMBO J.* **11**, 2885-2893.

Hawkins, N. C. and McGhee, J. D. (1990). Homeobox containing genes in the nematode *Caenorhabditis elegans*. *Nucl. Acid Res.* **18**, 6101-6106.

Hedgecock, E. M., Culotti, J. G., Hall, D. H. and Stern, B. D. (1987). Genetics of cell and axon migration in *Caenorhabditis elegans*. *Development* **100**, 365-382.

Hill, R. E., Favor, J., Hogan, B. L. M., Ton, C. C. T., Saunders, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. D. and van Heyningen, V. (1991). Mouse *small eye* results from mutations in a paired-like gene. *Nature* **354**, 522-525.

Johnson, R. N. and Kucey, B. L. (1988). Competitive inhibition of *hsp70* gene expression causes thermosensitivity. *Science* **242**, 1551-1554.

Kenyon, C. (1994). If birds can fly, why can't we? Homeotic genes and evolution. *Cell* **78**, 175-180.

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

Krause, M., Fire, A., Harrison, S. W., Priess, J. and Weintraub, H. (1990). CeMyoD accumulation defines the body wall muscle cell fate during *C. elegans* embryogenesis. *Cell* **63**, 907-919.

Krumlauf, R. (1994). *Hox* genes in vertebrate development. *Cell* **78**, 191-201.

Lawrence, P. A. and Morata, G. (1994). Homeobox genes: their function in *Drosophila* segmentation and pattern formation. *Cell* **78**, 181-189.

Levine, E. M. and Schechter, N. (1993). Homeobox genes are expressed in the retina and brain of adult goldfish. *Proc. Natl. Acad. Sci. USA* **90**, 2729-2733.

Levine, E. M., Hitchcock, P. F., Glasgow, E. and Schechter, N. (1994). Restricted expression of a new paired-class homeobox gene in normal and regenerating adult goldfish retina. *J. Comp. Neurol.* **348**, 596-606.

Liu, I. S. C., Chen, J.-de, Ploder, L., Vidgen, D., van der Kooy, D., Kalnins, V. I. and McInnes, R. R. (1994). Developmental expression of a novel murine homeobox gene (*Chx10*): Evidence for roles in determination of the neuroretina and inner nuclear layer. *Neuron* **13**, 377-393.

Mak, A. and Johnson, A. D. (1993). The carboxy-terminal tail of the homeo domain protein α2 is required for function with a second homeo domain protein. *Genes Dev.* **7**, 1862-1870.

Manser, J. and Wood, W. B. (1990). Mutations affecting embryonic cell migrations in *Caenorhabditis elegans*. *Dev. Genet.* **11**, 49-64.

McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* **68**, 283-302.

McInnes, R. R., Basu, S., Novak, J., Ploder, L., Liang, M.-Y., Hawes, N., Taylor, B., Roderick, T., Goldman, D., Hankin, M. and Burmeister, M. (1994). The ocular retardation (oc^J) mouse has an ochre mutation in the homeobox gene *Chx10* : direct evidence for *Chx10* as a major determinant of retinal development. *Amer. J. Human Genet.* **55S**, A3.

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.

Okkema, P. G. and Fire, A. (1994). The *Caenorhabditis elegans* NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in pharyngeal muscle. *Development* **120**, 2175-2186.

Pai, B. S., Pai, R. B. and Johnson, R. N. (1992). Overexpression of *c-myc* by amplification of negative promoter domain. *J. Biol. Chem.* **267**, 12428-12431.

Perkins, L. A., Hedgecock, E. M., Thomson, J. N. and Culotti, J. G. (1986). Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol.* **117**, 456-487.

Quiring, R., Walldorf, U., Kloter, U. and Gehring, W. J. (1994). Homology of the *eyeless* gene in mice and *Aniridia* in humans. *Science*, **265**, 785-789.

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

Siddiqui, S. S., Aamodt, E., Rastinejad, F. and Culotti, J. (1989). Anti-tubulin monoclonal antibodies that bind to specific neurons in *Caenorhabditis elegans*. *J. Neurosci.* **9**, 2963-2972.

Stringham, E. G., Dixon, D. K., Jones, D. and Candido, E. P. M. (1992). Temporal and spatial expression patterns of small Heat Shock (hsp16) genes in transgenic *Caenorhabditis elegans*. *Mol. Biol. Cell* **3**, 221-233.

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

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.

Tartof, K. D. and Henikoff, S. (1991). Trans-sensing effects from *Drosophila* to humans. *Cell* **65**, 201-203.

Tessier-Lavigne, M. (1991). Phototransduction and information processing in the retina. In *Principles of Neural Science* (eds. E. R. Kandel, J. H. Schwartz and T. M. Jessell), pp. 400-418. Norwalk, Connecticut: Appleton and Lange.

Ton, C. C. T., Hirvonen, H., Miwa, H., Weil, M. M., Monaghan, P., Jordan, T., van Heyningen, V., Hastie, N. D., Meijers-Heijboer, H., Drechsler, M., Royer-Pokora, B., Collins, F., Swaroop, A., Strong, L. C., and Saunders, G. F. (1991). Positional cloning and characterization of a paired

box- and homeobox-containing gene from the Aniridia region. *Cell* **67**, 1059-1074.

Treisman, J., Harris, E. and Desplan, C. (1991). The paired box encodes a second DNA-binding domain in the Paired homeodomain protein. *Cell* **59**, 553-562.

Vershon, A. K. and Johnson, A. D. (1993). A short, disordered protein region mediates interactions between the homeodomain of the yeast α2 protein and the MCM1 protein. *Cell* **72**, 105-112.

Walther, C. and Gruss, P. (1991). *Pax-6*, a murine paired box gene, is expressed in the developing CNS. *Development* **113**, 1435-1449.

Wang, B., Müller-Immergluck, Austin, J., Robinson, N. T., Chisholm, A. and Kenyon, C. (1993). A homeotic gene cluster patterns the anteroposterior body axis of *C. elegans*. *Cell* **74**, 29-42.

Ward, S., Thomson, N., White, J. C. and Brenner, S. (1975). Electron microscopical reconstruction of the anterior sensory anatomy of the nematode *Caenorhabditis elegans*. *J. Comp. Neurol.* **160**, 313-338.

Ware, R. W., Clark, D., Crossland, K. and Russell, R. L. (1975). The nerve ring of the nematode *Caenorhabditis elegans*: Sensory input and motor output. *J. Comp. Neurol.* **162**, 71-110.

Way, J. C., Wang, L., Run, J.-Q. and Wang, A. (1991). The *mec-3* gene contains cis-acting elements mediating positive and negative regulation in cells produced by asymmetric cell division. *Genes Dev.* **5**, 2199-2211.

White, J. G., Southgate, E., Thomson, J. N. and Brenner, S. (1986). The structure of the nervous system of *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B. Bio. Sci.* **314**, 1-340.

Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. and Desplan, C. (1993). Cooperative dimerization of Paired class homeo domains on DNA. *Genes Dev.* **7**, 2120-2134.

(Accepted 24 January 1995)