Muscle and nerve-specific regulation of a novel NK-2 class homeodomain factor in Caenorhabditis elegans

Brian D. Harfe1,2 and Andrew Fire1,*

1Carnegie Institution of Washington, Department of Embryology, 115 W. University Parkway, Baltimore, MD 21210, USA
2The Johns Hopkins University, Graduate Program in Biology, Baltimore, MD 21218, USA
*Author for correspondence (e-mail: fire@mail1.ciwemb.edu)

Accepted 17 November 1997; published on WWW 13 January 1998

SUMMARY

We have identified a new Caenorhabditis elegans NK-2 class homeobox gene, designated ceh-24. Distinct cis-acting elements generate a complex neuronal and mesodermal expression pattern. A promoter-proximal enhancer mediates expression in a single pharyngeal muscle, the donut-shaped m8 cell at the posterior end of the pharynx. A second mesodermal enhancer is active in a set of eight nonstriated vulval muscles used in egg laying. Activation in the egg laying muscles requires an ‘NdE-box’ consensus motif (CATATG) which is related to, but distinct from, the standard E-box motif bound by the MyoD family of transcriptional activators. Ectodermal expression of ceh-24 is limited to a subset of sublateral motor neurons in the head of the animal; this activity requires a cis-acting activator element that is distinct from the control elements for pharyngeal and vulval muscle expression. Activation of ceh-24 in each of the three cell types coincides with the onset of differentiation. Using a set of transposon-induced null mutations, we show that ceh-24 is not essential for the formation of any of these cells. Although ceh-24 mutants have no evident defects under laboratory conditions, the pattern of ceh-24 activity is apparently important for Rhabditid nematodes: the related species C. briggsae contains a close homologue of C. elegans ceh-24 including a highly conserved and functionally equivalent set of cis-acting control signals.

Key words: ceh-24, Vulval muscle, Pharyngeal muscle, Muscle, Nerve, Motor neuron, NK-2 homeodomain, NdE-box, Caenorhabditis elegans, Caenorhabditis briggsae

INTRODUCTION

In the past few years, many of the tools of molecular genetics have been used to extend our knowledge of the determination and differentiation of non-skeletal muscles. These muscles differ from skeletal muscles in their structure, in their contractile properties and in lacking expression of the myogenic helix-loop-helix (MyoD) family of transcription factors (Rudnicki and Jaenisch, 1995). A variety of regulatory factors have been proposed to coordinate non-skeletal muscle differentiation. These include members of the NK-2 family of homeodomain proteins (Harvey, 1996). In Drosophila, the NK homeodomain factors tinman and bagpipe are involved in the subdivision of the mesoderm and in the determination of cell fates in the dorsal mesoderm (Azpiazu and Frasch, 1993; Bate, 1993; Bodmer, 1993). The vertebrate gene Nkx2-5 has been shown to express in myocardiogenic progenitor cells and in mature cardiomyocytes (Lints et al., 1993; Komuro and Izumo, 1993). Mice containing a targeted disruption in Nkx2-5 do not initiate heart looping morphogenesis and die as embryos (Lyons et al., 1995). Several other transcription factors including GATA-4/5/6 (Laverriere et al., 1994), MEF-2 (Yu et al., 1992), MHOX (Cserjesi et al., 1992) and HF-1b (Zhu et al., 1993) have been implicated in vertebrate cardiac muscle development. Many of these regulatory components are members of structural families with roles in diverse muscle and non-muscle differentiation. In particular, NK homeodomains have been implicated in a wide variety of developmental processes.

The muscles of C. elegans can be grouped into three types: striated body wall muscles used in locomotion, single-sarcomere pharyngeal muscles used for feeding and defecation (Waterston, 1988). These groups of muscle cells are distinguished by lineage as well as their location and cell morphology. An analysis of regulatory components responsible for myosin heavy chain expression in C. elegans pharyngeal muscles led to identification of an NK-2 class homeodomain factor, CEH-22 (Okkema and Fire, 1994). ceh-22 was identified by the ability of its product to bind the enhancer of the pharyngeal-specific myosin heavy chain gene, myo-2. CEH-22 protein is found in a subset of pharyngeal muscles, prior to the detection of myo-2 product in these cells (Okkema and Fire, 1994).

The diverse roles for NK-2 class homeobox genes in vertebrate and invertebrate mesodermal development led us to search (experimentally and using DNA sequence databases) for additional members of this family in C. elegans. Of three
additional NK-2 family members that were found, only one (designated ceh-24) exhibits mesodermal activity. In this paper, we present molecular and genetic characterization of ceh-24.

MATERIALS AND METHODS

Molecular biology and transgenic strain construction

To identify novel NK homedomains, *C. elegans* genomic DNA was amplified using degenerate primers AF39 – GCTCTAGATTNTTTCNGCNARGT and AF44 – GGAATTCYTRRANCKRTGTRYGTAACCA (IUPAC notation). Both cDNA and genomic phage libraries were screened with the PCR product. Four independent genomic phage inserts and one cDNA were obtained. The genomic clones were used to map ceh-24 to chromosome V between *him-5* and *vmp-2* (Coulson et al., 1986; A. Coulson, personal communication). A *C. briggsae* ceh-24 genomic clone was obtained from screening a Chrom 4 genomic library provided by D. Baille and T. Snutch (Snutch, 1984; Snutch et al., 1988).

Transgenic *C. elegans* were made as described (Mello et al., 1991). The plasmid pRF4 was used as a dominant selectable marker for transgenic animals (Mello et al., 1991).

Whole-mount RNA in situ hybridization

Mixed stage populations of *C. elegans* were processed for in situ hybridization as described by Seydoux and Fire (1995), except that S.T.F. Molecular Biology Grade fixative (Streck Laboratories) was used to fix specimens overnight in place of the 20 minute formaldehyde treatment. Digoxigenin-labeled probe was visualized with alkaline-phosphatase-labeled antibodies. After the addition of coloring reagent, slides were allowed to develop for >5 hours in the dark.

TC1 excision screen

A TC1 insertion in ceh-24 was obtained from K. van der Linden and R. Plasterk. The insertion did not cause a phenotype. A PCR-based strategy was used to screen for imprecise excisions of the TC1 insertion as described by Plasterk (1995). Four independent deletion alleles of ceh-24 were recovered.

To address the activity of the ceh-24 promoter in a ceh-24 mutant background, we injected construct pHB29.1.73. This construct produces full ceh-24 promoter expression in a wild-type background and an identical pattern in the ceh-24 mutant cc539.

RESULTS

A novel NK-2 class homeodomain

We designed a degenerate PCR protocol to identify additional members of the NK-2 class homedomain family in *C. elegans*. The primers used for this protocol [see Materials and Methods] resulted in the identification of genomic sequence for a single novel gene, designated ceh-24. The gene is expressed: we obtained a single cDNA from approximately 1×10⁶ clones in an embryonic cDNA library. Among characterized NK-2 class members, the CEH-24 homedomain is most closely related to the vertebrate thyroid transcription factor TTF-1 (Fig. 1). A lower level of identity (77%) is seen with the *C. elegans* NK-2 homedomain protein CEH-22 (Okkema and Fire, 1994).

The CEH-24 promoter is active in three distinct tissues

To determine the activity pattern of the ceh-24 promoter, we constructed a series of lacZ and gfp translational fusions (Fig. 2). Additional reporter constructs which retained 5′, internal, and 3′ ceh-24 sequences were produced by in-frame insertion of gfp directly into the ceh-24 coding region. All of these chimeras have identical expression patterns. The constructs tested for expression cover a region extending 10.7 kb upstream and 0.7 kb downstream of the ceh-24 coding region.

In adult animals, expression was seen in three distinct areas: in the eight vulval muscles, in 8-10 ventral neurons in the head and in the most posterior pharyngeal muscle cell, m8 (Fig. 3A-C).

Expression in vulval muscle cells was seen from the L4 larval stage onward, starting just before these cells began contractile activity. Vulval muscle cells are born in the late L3 larval stage (Sulston and Horvitz, 1977). Two sex myoblast cells present in

---

**Fig. 1.** NK-2 homedomain comparison. Identical sequences are indicated by ‘*’. The CEH-24 homedomains in *C. elegans* and *C. briggsae* are identical. Sources for homodomain sequences are: *C. briggsae* CEH-24 (this paper); TTF-1/Ntx-2.1 (Guazzi et al., 1990); Dth-2 (Garicia-Fernandez et al., 1993); Nks2-5/Cxx (Lints et al., 1993); Komuro and Izumo, 1993); Nks-2, Nks-3 (Price et al., 1992); XINK-2 (Saha et al., 1993); Lox-10 (Nardelli-Haefliger and Shankland, 1993); NK-2/vnd, tin/NK-4, bag/NK-3 (Kim and Nirenberg, 1989); CEH-22 (Okkema and Fire, 1994); EgHbx3 (Oliver et al., 1992); ceh-27, ceh-28 and ceh-29 are additional related genes that have been sequenced by the ongoing *C. elegans* genome sequencing consortium (Wilson et al., 1994; B. D. Harfe and A. Fire, unpublished results).
Fig. 2. Characterization of the ceh-24 promoter and enhancers. (A) Activity of deletion constructs derived from a ceh-24::gfp translational fusion. All constructs contained the first 5 exons of ceh-24. pBH24.62 was assayed using a lacZ reporter. All other constructs were fused in frame to gfp. For each construct, GFP or LacZ expression was assayed in at least 20 independent F1 animals, and in at least one heritable line. Expression was scored in the vulval muscles, pharyngeal muscle m8, and in head neurons. Reporter-positive cells are indicated by a ‘+’; reporter negative cells are indicated by a ‘−’. Ectopic expression was occasionally seen in gut cells and rarely in anal muscles; this type of ectopic expression has been previously reported to result from juxtaposition of vector sequences (Krause et al. 1994). (B) Delineation of the ceh-24 vulval muscle enhancer sequence. DNA sequences in the vicinity of the putative vulval muscle enhancer were assayed upstream of a truncated pes-10 promoter fragment previously reported to result from juxtaposition of vector sequences (Krause et al. 1994). (B) Delineation of the ceh-24 vulval muscle enhancer sequence. DNA sequences in the vicinity of the putative vulval muscle enhancer were assayed upstream of a truncated pes-10 promoter fragment previously reported to result from juxtaposition of vector sequences (Krause et al. 1994). (B) Delineation of the ceh-24 vulval muscle enhancer sequence. DNA sequences in the vicinity of the putative vulval muscle enhancer were assayed upstream of a truncated pes-10 promoter fragment previously reported to result from juxtaposition of vector sequences (Krause et al. 1994). (B) Delineation of the ceh-24 vulval muscle enhancer sequence. DNA sequences in the vicinity of the putative vulval muscle enhancer were assayed upstream of a truncated pes-10 promoter fragment previously reported to result from juxtaposition of vector sequences (Krause et al. 1994).
the L3 larval stage are the precursors to the eight vulval and eight uterine muscles (Sulston and Horvitz, 1977). No ceh-24 reporter activity was seen in the sex myoblasts or uterine muscles.

Expression in the pharynx was limited to a single cell, the donut-shaped muscle cell m8. This unique cell forms the rear of the pharynx. m8 expression was first seen in the embryo, beginning approximately 400 minutes after the first cell cleavage. This activity continued throughout the life of the animal.

*C. elegans* hermaphrodites contain 302 neurons (White et al., 1986; Chalfie and White, 1988). The precise identity of each neuron can, in principle, be determined by the location of its nucleus and the paths of its nerve processes (White et al., 1986). The ceh-24 reporter construct was active in 8-10 neurons in the ventral head region of *C. elegans*. Using a ceh-24::gfp fusion construct (pBH28.101), the 8-10 ventral head neurons were observed to extend nerve processes into the nerve ring and sublateral nerve cords. Based on this observation, six of these neurons have been identified as SMDVL, SMDVR, SIBVL, SIBVR, SIBDL and SIBDR. The four remaining neurons with ceh-24 promoter activity appear also to belong to the SI/SM class of sublateral neurons.

To localize ceh-24 RNA transcripts in the embryo, we performed RNA in situ hybridization. ceh-24 transcripts were first detected approximately 400 minutes after the first cell cleavage, and were seen in ventral head neurons (Fig. 3F-H). This correlates with the first detection of ceh-24 reporter expression. RNA transcripts could be detected in these cells until hatching. We have not been able to obtain reproducible permeabilization of differentiated pharyngeal or vulval muscles. Difficulty in detecting mRNA in these cells occurs with several unrelated transcripts, likely reflecting a physical barrier to permeabilization of these tissues (G. Seydoux, B. D. Harfe and A. Fire, unpublished results).

**Distinct control elements responsible for the ceh-24 promoter activity pattern**

To locate the control elements that are responsible for the observed ceh-24 promoter activity pattern, we first carried out a unidirectional deletion analysis starting with a ceh-24::gfp translational fusion. Activity was analyzed using both an F1 expression assay and in stable transgenic lines (Okkema et al., 1993).

The unidirectional deletion analysis revealed three distinct DNA elements specifying different aspects of ceh-24 promoter activity (Fig. 2A). The critical elements reside between -2602 bp and -1500 bp upstream of the start of translation, in the order motor neuron element→vulval muscle element→m8 element→promoter.

**Characterization of a vulval muscle enhancer sequence**

The deletion analysis indicated a vulval muscle control element positioned between -2443 and -1989 relative to the translational start. To further narrow down this signal, we carried out enhancement assays using a vector based on the minimal promoter segment from pes-10 (Seydoux and Fire, 1994). The crippled pes-10 promoter in this vector is not intrinsically active, but can be activated in a variety of tissues by juxtaposition to diverse tissue-specific enhancers (B. D. Harfe, A. Fire, S. Xu and G. Seydoux, unpublished observations). A 452 bp sequence from the vulval enhancer region of ceh-24 activated the minimal pes-10 promoter in all eight vulval muscle cells (Fig. 3D). This fragment was capable of enhancement in either orientation. Further deletions identified a core 48 bp sequence that could activate reporter expression in the eight vulval muscles (pBH37.09, Fig. 2B).

Within the core 48 bp vulval enhancer, there are two repeats of the sequence ‘CATATG’ and a consensus ‘GAGA’ site. GAGA sites in other systems have been shown to bind a factor with general ability to aid in enhancer function (Biggin and Tjian, 1988). In constructs with a single copy of the enhancer sequence, the GAGA element is necessary for activity (pBH38.50 and pBH32.78, Fig. 2B). The GAGA site is apparently not the major determinant of tissue specificity, since duplication of an enhancer fragment containing a mutated GAGA sequence can overcome the need for this site (pBH38.56, Fig. 2B).

Assays with duplicated ceh-24 enhancer segments allowed us to define a 22 bp element which is sufficient to specify expression in vulval muscle. The 22 bp minimal vulval muscle enhancer contains two copies of the sequence ‘AACATATG’. We will refer to the sequence ‘CATATG’ as an ‘Nde-box’ (named for the restriction enzyme Ndel, which recognizes this sequence).

Concatamers of the 22 bp Nde-box segment have a broader tissue specificity than the original ceh-24 enhancer. In addition to the eight vulval muscles, expression was seen in many of the remaining single sarcomere body muscles: the two intestinal muscles, the anal depressor muscle and the eight uterine muscles (Fig. 3I-L).

**Characterization of an enhancer active in posterior pharyngeal muscle m8**

The unidirectional deletion analysis suggested a 395 bp region as a candidate for the m8 signal. This region was tested for its ability to activate a minimal (non-active) muscle promoter from myo-2. The truncated myo-2 promoter used in these experiments has no activity by itself but can be activated in a variety of cell types when adjacent to enhancer sequences (Okkema et al., 1993; Jantsch-Plunger and Fire, 1994). A 117 bp fragment from ceh-24 activated expression in the pharyngeal muscle cell m8 (Fig. 3E). Enhancement by this segment was not seen in any other cell.

Further mutational analysis indicated that the 117 bp m8 enhancer contains three distinct elements that are necessary for enhancer activity:

(1) Three GAGA boxes. m8 enhancer activity is lost following deletion of a 50 bp sequence containing two of these sites, or mutation of the third GAGA box (pBH37.28, pBH55.25, Fig. 2C).

(2) A twice-repeated 9 bp sequence (pBH55.28, Fig. 2C). This sequence contains two ‘GATA’ motifs.

(3) An AT-rich region at the promoter-distal end of the enhancer (pBH35.35, Fig. 2C).

**Null mutations of ceh-24 have no detectable phenotype**

A transposon-based strategy has been used to create targeted deletion mutations in a variety of *C. elegans* genes (Plasterk, 1995). Four ceh-24 deletion alleles, designated cc539-cc542, were isolated using a TC1 insertion in ceh-24 obtained from K. van der Linden and R. Plasterk. All of these deletions remove the homeobox and the downstream acidic region (Fig. 4A). The largest, cc539, deletes the entire CEH-24 coding
All four alleles are viable. No phenotypic differences from wild-type were evident. The largest deletion, ceh-24(cc539), was analyzed further for the following properties.

1. Growth – ceh-24(cc539) animals grow at all permissive temperatures at a rate similar to wild-type animals.
2. Visible phenotypes – ceh-24(cc539) animals are not visibly dumpy (Dpy), uncoordinated (Unc) or egg-laying defective (Egl).
3. Mating – A ceh-24(cc539) male stock mated with an efficiency similar to wild-type.
4. Chemotaxis – ceh-24(cc539) animals were found to chemotax normally to three volatile odorants (benzaldehyde, diacetyl and pyrazine) and one non-volatile odorant (Cl⁻ ions) (C. Bargmann, personal communication).
5. Pharyngeal pumping – ceh-24(cc539) animals have a normal pharyngeal activity pattern (electropharyngeogram), with effective muscle activity and pumping (L. Avery, personal communication).
6. VC neuron synapses – the VC neurons were visualized to synapse normally onto the vulval muscles using an unc-4::gfp construct (The unc-4::gfp strain was a gift from H. Hutter and D. Miller; unc-4::gfp is expressed in VC neurons, D. Miller personal communication; Pflugrad et al., 1997).

To address the possibility of autoregulatory phenomena, we introduced a ceh-24::gfp reporter construct into ceh-24(cc539) animals. The observed expression pattern was identical to that of the same DNA construct in wild-type animals. This experiment confirmed the presence in ceh-24(cc539) animals of morphologically normal vulval muscles, sublateral motor neurons and pharyngeal muscle m8. These cells were in their correct positions and their cell shapes appeared to be wild type. The paths of the head neuron processes appeared normal in the

Fig. 3. Micrographs showing ceh-24 reporter expression and RNA localization. Expression patterns for all constructs were assayed both in transgenic lines and in the first generation following microinjection. Identical expression patterns were observed for all constructs in both heritable lines and F1 assays. (A) ceh-24::lacZ expression in an adult animal. Expression is seen in the eight vulval muscles, eight ventral neurons in the head, and in the most posterior pharyngeal muscle, m8. (B) ceh-24::lacZ expression in the eight vulval muscles of an adult animal. (C) ceh-24::lacZ expression in the adult head. Activity is seen in pharyngeal muscle m8 and in 8-10 ventral head neurons. (D) The ceh-24 vulval muscle enhancer driving lacZ expression from a truncated pes-10 promoter in the eight vulval muscles. (E) The ceh-24 m8 enhancer driving lacZ expression from a truncated myo-2 promoter in the pharyngeal muscle m8. (F-H) In situ localization of ceh-24 RNA. Antisense probes give a pattern with ventral head neurons showing strong expression in comma stage (F), two-fold (G), and three-fold (H) embryos. No signal was seen with a comparable ‘sense’ probe (not shown). (I-L) GFP expression driven by concatamersized NdE-boxes. GFP activity is both nuclear and cytoplasmic because of the tendency of the small GFP protein to leak out of the nucleus. (I) Anal depressor muscle (black arrowhead) and the two intestinal muscles (white arrowheads) are shown. Fibers of the intestinal muscles can be seen. (J,K) Higher magnification shows cellular outline of a GFP-positive anal depressor muscle. (L) NdE-box driven GFP expression is seen in vulval and uterine muscles. Scale: Adults are 1 µm long and approximately 100 µm wide; Embryos are 50 µm long.
Fig. 4. Generation of null mutations in ceh-24. (A) Four transposon-catalyzed deletions were isolated (cc539-cc541; see Materials and Methods). Deletion junctions were amplified by PCR and sequenced. Solid lines indicate deleted DNA. ceh-24 genomic organization is shown above the deletions. (B) Southern Blot analysis of deficiency DNA probed with a DNA fragment encoding the CEH-24 homeodomain. No hybridization to the CEH-24 homeodomain probe was detected in either of the two ceh-24 deletion mutations. Marker sizes ‘M’ are in kb. (C) Southern Blot analysis of deficiency DNA probed with the entire ceh-24 gene. There are EcoRI sites flanking ceh-24 and one internal site. Digestion with EcoRI thus yields two wild-type bands. The internal EcoRI site is absent in both ceh-24 deletions, resulting in the detection of just one band. A faint band in EcoRI digested ceh-24(cc539) DNA resulted from partial DNA digestion. A BamHI/Xbal double digestion produced a wild-type 4.5 kb DNA fragment containing the entire ceh-24 gene. A smaller BamHI/Xbal band was observed in each ceh-24 deletion.

In addition to ceh-24, there is just one known C. elegans NK2 family member with mesodermal activity: CEH-22. CEH-22 protein is present in a set of pharyngeal muscle cells that do not show ceh-24 promoter activity (Okkema and Fire, 1994). To determine whether CEH-24 was responsible for the lack of CEH-22 activity in m8, we injected a ceh-24::lacZ construct into ceh-24(cc539) animals (data not shown).

A mutation in ceh-22 has recently been reported (Okkema et al., 1997). ceh-22 mutant animals have a defective pharynx and are occasionally sterile. To investigate a possible role for CEH-22 in ceh-24 transcriptional repression, we examined expression of a ceh-24::gfp construct (pBH29.58, Fig. 2A) in ceh-22 mutant animals. Homozygous ceh-22 mutant animals had normal expression of the ceh-24::gfp construct (data not shown).

The genome of C. briggsae contains a very close relative of C. elegans ceh-24. C. briggsae is a Rhabditid nematode that contains many genes
of similar function and structure to *C. elegans* (Zucker-Aprison and Blumenthal, 1989; Heschl and Baillie, 1990; Kennedy et al., 1993; Krause et al., 1994). Two overlapping phage inserts were obtained from screening a *C. briggsae* genomic library with a *C. elegans* ceh-24 homeobox containing fragment. Within the homeodomain, *C. elegans* CEH-24 and the *C. briggsae* homologue are identical (fig. 1). Outside the homeodomain, the proteins are 79% identical (fig. 5A) and five of six intron positions are conserved. Different NK-2 proteins generally do not have high homology outside the homeodomain. The presence of such high homology outside the homeodomain suggests that we have cloned the *C. briggsae* ceh-24 homologue and not a more distantly related NK-2 class homeodomain.

Injection of a *C. elegans* ceh-24::gfp reporter construct into *C. briggsae* produced GFP expression in *C. briggsae* head neurons, pharyngeal muscle m8 and the eight vulval muscles (data not shown). This is identical to the expression pattern obtained upon injection of this construct into *C. elegans*. We constructed a *C. briggsae* ceh-24 gene tagged with gfp. When injected into *C. elegans*, this construct produced expression identical to the *C. elegans* ceh-24::gfp fusion (data not shown).

Sequences upstream of *C. briggsae* ceh-24 contain motifs identical to the *C. elegans* ceh-24 NDE-boxes and closely related to the m8 enhancer (fig. 5B,C). Outside the NDE-boxes and m8 enhancer, the upstream sequences from the two genes are not conserved. The *C. briggsae* homeologue of the vulval enhancer is apparently a functional homologue; the *C. briggsae* sequence (in either orientation) can drive the pes-10 minimal promoter in all eight vulval muscles in *C. elegans* (data not shown).

**DISCUSSION**

A novel NK-2 class homeodomain factor in *C. elegans*

NK-class homeodomain factors have been implicated in a variety of mesodermal and ectodermal differentiation processes in both invertebrate and vertebrate systems (Azpiazu and Frasch, 1993; Lyons et al., 1995; Okkema et al., 1997). Efforts in this laboratory, combined with the ongoing genome sequencing project, have identified four members of the NK-2 homeodomain family in *C. elegans*: CEH-22, CEH-24, CEH-27 and CEH-28 (Okkema and Fire, 1994; Wilson et al., 1994; this work; B. D. Harfe and A. Fire, unpublished data). Although the various activity patterns of NK-class homeodomains in vertebrate and invertebrate development suggest a diversity of regulatory roles, the conserved nature of the family suggests that certain underlying regulatory mechanisms (and possibly specific targets) could be conserved.

In this paper, we describe the characterization of the *C. elegans* NK class homeodomain factor, CEH-24. The CEH-24 homeodomain shows strong homology with the NK-2 subfamily, and in particular is more closely related to vertebrate thyroid transcription factor (TTF-1) than to any of the other known *C. elegans* NK-2 class homeodomains. In our analysis of ceh-24, we have focused on tissue activity pattern, on the regulation of expression and on the functional consequences of loss of function.

**The ceh-24 gene is active in specific subsets of neural and muscle tissue**

We used a combination of in situ hybridization and reporter gene fusions to analyze the activity pattern of the ceh-24 gene. The set of cells showing activity, a subset of 8-10 sublateral motor neurons and two types of non-striated muscles was somewhat surprising; these groups of cells had not been previously connected by any morphological or functional criteria.

The sublateral motor neurons had been primarily described in terms of their axonal outgrowth and synaptic connectivity. These cells are unique in that they send processes into the sublateral nerve cords, with synapses to body wall muscle cells (White et al., 1986; D. H. Hall and J. B. Rand, unpublished results). The role played by these cells in *C. elegans* behavior has not been elucidated. (Function has yet to be assigned to a large fraction of the neurons in *C. elegans*; e.g., Avery and Horvitz, 1989, Bargmann and Horvitz, 1991). By contrast, the vulval and pharyngeal muscles have been assigned clear roles in the control of egg laying and in pumping of food into the intestine, respectively (Waterston, 1988). The proper physiological regulation of these contractile processes are critical to the fitness of the species. Optimal regulation presumably requires both the proper assembly of the contractile tissue and the ability to respond to a plethora of extrinsic signals modulating muscle contraction. These tissues might thus be expected to utilize a variety of regulatory components for specific aspects of cellular function.

**ceh-24 and ceh-22 define distinct identities within the pharyngeal musculature**

The ceh-24 activity pattern in the pharynx is complementary to that of the previously identified *C. elegans* NK-2 homeodomain factor CEH-22. CEH-22 protein is present in the pharyngeal muscles m1, m3, m4, m5 and m7 (Okkema and Fire, 1994), but is not present in the ceh-24-positive muscle cell m8. Neither gene is active in muscle cells m2 or m6. The activity pattern for ceh-22 is unaffected in a ceh-24 mutant; likewise the ceh-24 activity pattern is unaffected in a ceh-22 mutant. Our analysis thus indicates that the two non-overlapping expression patterns result from independent controls and not from an exclusionary mechanism.

**GAGA and NDE-box elements combine to specify expression in vulval muscles**

In analyzing requirements for the activity pattern of the ceh-24 promoter, we found an underlying piecemeal organization, with each aspect of the expression pattern mediated by a distinct set of cis-acting signals. The vulval muscle enhancer region lies 2104 bp upstream of the mRNA start. The most prominent sequence features are a tandem pair of motifs with the core sequence ‘CATATG’. As a mnemonic, we have designated these sequences ‘NDE-boxes’ (the NdeI restriction enzyme recognizes this sequence). Transcription factors of the bHLH family have been found to recognize sequences with a core ‘CAnnTG’, which would include the NdeI-box. It should be noted that the myogenic bHLH family (including *C. elegans* CeMyoD and vertebrate MyoD, myogenin, Myf-5, and MRF4) recognize a distinct sequence, CAGCTG (Blackwell and Weintraub, 1990; Krause et al., 1997). Hence, these known
myogenic proteins would not be expected to readily bind the NdE-box consensus sequence. Consistent with this hypothesis, *C. elegans* *hlh-1* expression has not been detected in the non-striated vulval muscles (Krause et al., 1990). Instead, it seems likely that activity of the NdE-box element is mediated by a distinct factor, possibly a divergent member of the bHLH family.

In addition to the NdE-boxes, the vulval enhancer contains a ‘GAGA’ sequence that is required for enhancer function. Similar ‘GAGA’ sequences in *C. elegans* have been found to be required for activity of the *unc-54* enhancer (Jantsch-Plunger and Fire, 1994), and the *ceh-24* m8 enhancer (this work). These sites could serve as a binding site for a broadly active factor similar to the *Drosophila* GAGA-binding protein *Trithorax-like* (Farkas et al., 1994; Biggin and Tjian, 1988). This class of DNA-binding proteins have been proposed to create nucleosome-free areas of DNA, allowing the binding of additional transcription factors (Tsukiyama et al., 1994). No putative homologue of the *Drosophila* GAGA gene has yet been identified in *C. elegans*.

The ability to bypass the requirements for the GAGA element by duplicating the NdE-box portion of the enhancer indicates that the GAGA site is not the major determinant of vulval muscle activity. Interestingly, the activity of the dimerized 22-mer included additional non-striated muscles outside of the vulva: the intestinal muscles, uterine muscles and anal depressor muscle. This suggests that the GAGA and/or other sequences in the complete 81 bp vulval enhancer may act to limit expression in these additional muscles. The activity of the concatamerized NdE-box sequence in multiple sets of non-striated muscles could reflect either (1) a single factor binding the NdE-boxes and present in most of the minor muscles, or (2) several distinct NdE-binding factors present in different minor muscle subsets.

Is there a role for *ceh-24*?

Several circumstances could explain the lack of a detectable phenotype for a *ceh-24* null mutation.

1. The function of CEH-24 may be quite subtle, so that a phenotype might not be evident with current techniques. For example, a decrease in a vulval muscle metabolic function or contractile strength would be difficult to quantify.

2. CEH-24 might be serving a redundant function. We do not yet have any clear candidate for a redundant factor. The only previously characterized member of the *C. elegans* NK-2 class of homeodomains, *CEH-22* (Okkema and Fire, 1994), is not present in cells with *ceh-24* activity. The *C. elegans* genome project has recently identified two additional members of the NK-2 family (*ceh-27* and *ceh-28*) and a single member of the closely related brain-specific-homeobox family (Jones and McGinnis, 1993; *ceh-29*). We have begun to characterize these three genes. Preliminary results indicate that none of the three genes are capable of expressing in *ceh-24*-positive cells (B. D. Harfe and A. Fire, unpublished results). In addition, *ceh-27, ceh-28*, and *ceh-29* reporter fusions do not have any ectopic expression in a *ceh-24* null mutation (B. D. Harfe and A. Fire, unpublished results). Genes that could compensate for *CEH-24*, but are not closely related to CEH-24, would not have been identified in this analysis.

3. CEH-24 may function in a behavioral or developmental program that is not required under laboratory conditions. *C. elegans* are grown on a plentiful food source under ideal conditions in the laboratory. In the wild, *C. elegans* is subject to many different environmental signals and stresses. Specific behavioral responses requiring CEH-24 might be important for fitness under these non-laboratory conditions.

The presence of a gene closely related to *C. elegans* *ceh-24* in the rhabditid nematode *C. briggsae* is indicative of an evolutionarily important function for CEH-24. The two genes encode identical homeodomains and yield 79% protein identity over their entire length. *ceh-24* vulval and m8 enhancers are also conserved between *C. elegans* and *C. briggsae*. This conservation is likely to be significant, since the two species are known to have diverged sufficiently to remove any non-selected similarity outside of coding regions (Prasad and Baillie, 1989).

The *C. elegans* and *C. briggsae* control regions showed apparently identical activation patterns when transformed into the heterologous species. This functional equivalence of control regions indicates that the trans-acting factors responsible for *ceh-24* activity are conserved both in their binding specificity and activity pattern. Conservation of regulatory properties in the two species argues strongly for a selective advantage conferred by the *ceh-24* expression pattern.

We thank K. van der Linden and R. Plasterk (funded by grant 5 R01 RR10082-02 of the NIH-NCCR to R. P.) for isolating the Tc1 insertion strain used in this paper. We thank A. Antebi and H. Hutter for help in the identification of neuronal cells, and P. Okkema, M. Montgomery, S. Kostas, J. Hsieh, M. Hsu, H. Rienhoff, and W. Kelly for advice throughout this work. This work was supported by the National Institutes of Health (R01GM37706 to AZF, training grant T32GM07231 to The Johns Hopkins University Department of Biology). Sequences have been submitted to GenBank.

REFERENCES


Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H., Karch, F.


