Development 120, 2175-2186 (1994)
Printed in Great Britain © The Company of Biologists Limited 1994

The *Caenorhabditis elegans* NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in pharyngeal muscle

Peter G. Okkema* and Andrew Fire

Carnegie Institution of Washington, Department of Embryology, 115 W. University Parkway, Baltimore, MD 21210, USA

*Author for correspondence

SUMMARY

The pharyngeal muscles of *Caenorhabditis elegans* are single sarcomere muscles used for feeding. Like vertebrate cardiac and smooth muscles, *C. elegans* pharyngeal muscle does not express any of the known members of the MyoD family of myogenic factors. To identify mechanisms regulating gene expression in this tissue, we have characterized a pharyngeal muscle-specific enhancer from *myo-2*, a myosin heavy chain gene expressed exclusively in pharyngeal muscle. Assaying enhancer function in transgenic animals, we identified three subelements, designated *A*, *B* and *C*, that contribute to *myo-2* enhancer activity. These subelements are individually inactive; however, any combination of two or more subelements forms a functional enhancer. The *B* and *C* subelements have distinct cell type specificities. A duplication of *B* activates transcription in a subset of pharyngeal muscles (m3, m4, m5 and m7). A duplication of *C* activates transcription in all pharyngeal cells, muscle and non-muscle. Thus, the activity of the *myo-2* enhancer is regulated by a combination of pharyngeal muscle-type-specific and organ-specific signals. Screening a cDNA expression library, we identified a gene encoding an NK-2 class homeodomain protein, CEH-22, that specifically binds a site necessary for activity of the B subelement. CEH-22 protein is first expressed prior to myogenic differentiation and is present in the same subset of pharyngeal muscles in which *B* is active. Expression continues throughout embryonic and larval development. This expression pattern suggests CEH-22 plays a key role in pharyngeal muscle-specific activity of the *myo-2* enhancer.

Key words: pharynx, myogenesis, *Caenorhabditis elegans*, NK-2 class homeodomain, myosin gene

INTRODUCTION

Vertebrates contain three muscle types: skeletal muscle, cardiac muscle and smooth muscle. Although these muscle types express many of the same genes, they are morphologically distinct and arise from separate regions of the developing embryo. A group of helix-loop-helix transcription factors, collectively referred to as the MyoD family, has been implicated in skeletal muscle differentiation (for recent review see Emerson, 1993). None of the identified MyoD family members are detected in cardiac and smooth muscle. Therefore differentiation of these muscle types must involve either divergent basic-helix-loop-helix factors or members of other transcription factor families. Several candidates for cardiac muscle differentiation factors have been proposed, including the MEF-2 family of MADS box transcription factors, the homeodomain proteins MHox and *Nkx-2.5/Csx* and the zinc-finger protein HF-1b (Yu et al., 1992; Cserjesi et al., 1992; Lints et al., 1993; Komuro and Izumo, 1993; Zhu et al., 1993).

Like vertebrates, the muscles in the nematode *Caenorhabditis elegans* can be divided into distinct classes: body wall muscle, pharyngeal muscle and several groups of minor muscles (for review see White, 1988). In structure and function, body wall muscle may be most analogous to vertebrate skeletal muscle. Screens for MyoD homologs in *C. elegans* have identified a single family member, designated *hlh-1*, expressed specifically in body wall muscles and their clonal precursors (Krause et al., 1990). No *hlh-1* expression is seen in pharyngeal muscle or the minor muscles, suggesting that differentiation of these muscle types may be analogous to that of vertebrate cardiac or smooth muscle.

We have examined the regulation of myosin heavy chain gene expression as an initial step to analyze myogenesis in *C. elegans*. Two myosin heavy chain genes, *myo-1* and *myo-2*, are specifically expressed in pharyngeal muscle (Miller et al., 1986; Ardizzi and Epstein, 1987). Regulatory regions controlling the expression of *myo-2* have been characterized in most detail (Okkema et al., 1993). The *myo-2* gene contains at least two independent tissue-specific regulatory elements: a promoter sufficient for low level pharyngeal muscle-specific expression is located near the transcriptional start site, and a separable pharyngeal muscle-specific enhancer is located approximately 300 bp upstream of the start site. Neither the promoter nor the enhancer contains consensus binding sites associated with vertebrate muscle-specific genes.

In this paper, we describe the modular structure of the *myo-2* enhancer. At least three subelements contribute to enhancer activity. Two of these display distinct cell type specificities: one is active in a subset of pharyngeal muscles; the second is more generally active in all cell types in the pharynx. In a

screen for factors regulating the *myo-2* enhancer, we identified an NK-2 class homeodomain factor that appears to play a key role in pharyngeal muscle-specificity of the *myo-2* enhancer.

MATERIALS AND METHODS

Plasmids and general methods for nucleic acid manipulation

Standard methods were used to manipulate plasmids DNAs, oligonucleotides and RNA (Ausubel et al., 1990). The parental promoter::*lacZ* fusions used to assay enhancer activity are described in Okkema et al. (1993): pPD26.02 (*myo-3::lacZ*); pPD26.50 (*glp-1::lacZ*); pOK1134 (Δ*myo-2::lacZ*; this construct is similar to pOK5.56 with a *Sty*I restriction site inserted upstream of the *myo-2* sequences to facilitate cloning of oligonucleotides). Oligonucleotides were designed with 5′, non-palindromic overhangs to allow ligation as head-to-tail concatenates and ligation into the *Sty*I site of pOK5.56. Inserts of plasmids containing oligonucleotides from the *B* and *C* fragments were sequenced to determine concatamer number and confirm the integrity of the cloned oligonucleotides. Sequences of plasmid constructs are available from the authors.

When hybridized to an ordered array of YACs spanning the *C. elegans* genome (Coulson et al., 1991), the *ceh-22* cDNA identified four overlapping YACs (Y24E3, Y59H10, Y61D5, Y33F10) located between *her-1* and *act-1,2,3* on chromosome V. *ceh-22* genomic DNA was subcloned from cosmid WB2 (kindly provided by T. Bürglin). To construct the *ceh-22::lacZ* fusion pOK29.02, a 4 kb fragment containing *ceh-22* 5′-flanking sequences and part of the 5′-UTR was subcloned into the *lacZ* expression vector pPD22.04 (Fire et al., 1990).

Handling of nematodes

C. elegans strain Bristol (N2) was grown under standard conditions (Sulston and Hodgkin, 1988). F_1 expression assays were done as previously described (Okkema et al., 1993). Plasmid DNAs were injected at 100 μg/ml into the germ line of adult hermaphrodites (Mello et al., 1991) and F_1 progeny stained for β-gal activity as larvae and adults (Fire, 1993). We have also tested activity of the *B+B* and *C+C* enhancers in transgenic lines (Mello et al., 1991). As has been observed with other enhancer assay constructs, the *B+B* and *C+C* enhancer constructs function poorly in high copy extrachromosomal arrays in transformed lines. We do not know whether the lack of observed enhancer function in these lines is a result of copy number, the co-selected marker gene, or the structure of the transforming DNA. Stable lines expressing *ceh-22::lacZ* were obtained by co-injecting pOK29.02 with pRF4 (Mello et al., 1991); the resulting extrachromosomal array was integrated into a chromosome by X-ray irradiation (Krause et al., 1990).

cDNA library construction and screen for *myo-2* enhancer binding proteins

cDNA libraries in λgt11 were constructed with oligo(dT)-primed cDNA synthesized from poly(A)+ RNA isolated from either embryos or mixed stage animals. cDNAs encoding candidate enhancer binding factors were isolated from the embryo cDNA library by screening with concatenated, double-stranded oligonucleotides (Vinson et al., 1988, Singh et al., 1988) using conditions optimized by V. Jantsch-Plunger (Jantsch-Plunger, 1993). Phage were grown 3-4 hours on *E. coli* Y1090 at 42°C. The plate was overlayed with a nitrocellulose filter soaked in 100 mM IPTG and transfered to 37°C for 6 hours. The filter was removed and a second IPTG soaked filter placed on the plate for 10-12 hours at 37°C. Filter lifts were processed by modification of the procedure described by Vinson et al. (1988). Filters were sequentially agitated in prebinding buffer [10 mM Hepes (pH7.9), 10 mM $MgCl_2$, 50 mM KCl, 1 mM DTT] containing 6 M, 4 M, 2.7 M, 1.3 M, 0.67 M, 0.33 M and 0.17 M guanidine hydrochloride, followed by two washes in prebinding buffer (10 minutes each, 4°C). Filters were blocked in prebinding buffer containing 5% non-fat dry milk (Carnation) (45 minutes, 4°C) and probed in binding buffer [10 mM Hepes (pH7.9), 10 mM $MgCl_2$, 100 mM KCl, 1mM DTT] containing 2% non-fat dry milk, 1.5 μg/ml double-stranded salmon sperm DNA, 3 μg/ml denatured salmon sperm DNA and 10^5 cts/minute/ml nick-translated concatenates of *B207* and *C183* (overnight, 4°C). Filters were washed twice in binding buffer containing 2% non-fat dry milk and twice in TBS (10 minutes, 4°C).

Expression and purification of recombinant CEH-22 protein

Two different recombinant CEH-22 fusion proteins were expressed in *E. coli* BL21. A glutathione-S-transferase::CEH-22 fusion protein (GST::CEH-22) encoding CEH-22 amino acids 1-346 was purified according to Smith and Johnson (1988). A phage T7 gene 10::CEH-22 fusion protein containing a polyhistidine tract (poly-his::CEH-22) encoding CEH-22 amino acids 79-346 and a derivative of poly-his::CEH-22, deleted for CEH-22 amino acids 217-264, were purified by affinity chromatography to immobilized Ni^{2+} under denaturing conditions (Hochuli et al., 1990).

DNAseI protection assay

DNAse protections assays were performed by modification of a protocol described by Ausubel et al. (1990). Poly-his::CEH-22 was bound to an end-labeled *B* fragment probe in 10 mM Hepes (pH 7.9), 10 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, 1.5 μg/ml double-stranded salmon sperm DNA, 3 μg/ml denatured salmon sperm DNA and 100 μg/ml BSA.

Anti-CEH-22 antibody preparation, affinity purification and immunofluorescence

Rabbit polyclonal antibodies were raised separately against GST::CEH-22 (antiserum c184) and poly-his::CEH-22 (antiserum c187). Antibodies that specifically recognize CEH-22 were affinity purified (Harlow and Lane, 1988; N. Patel, personal communication) from c184 by binding to immobilized poly-his::CEH-22 protein and from c187 by binding immobilized GST::CEH-22. Affinity-purified anti-CEH-22 antibodies were used undiluted.

Embryos [isolated by hypochlorite digestion (Sulston and Hodgkin, 1988)] were fixed 30 minutes (in PBS containing 55 mM Pipes (pH 6.95), 1.1 mM EGTA, 0.5 mM $MgSO_4$, 2.3% formaldehyde) under a coverglass on a microscope slide coated with 3% polylysine. The slide was then frozen on an aluminum block cooled in dry ice, separated from the coverglass, dipped for 4 seconds in –20°C methanol and rinsed 3× 5 minutes in TTBS (TBS, 0.1% Tween 20). The embryos were incubated overnight at 4°C under a drop of primary antibody, washed 4× 20 minutes in TTBS, incubated 4 hours at 25°C with a secondary antibody, washed 4× 20 minutes in TTBS and mounted in 70% glycerol, 1 mg/ml phenylenediamine, 0.02% NaN_3 for microscopy.

RESULTS

The *myo-2* enhancer contains multiple subelements that cooperate to activate transcription

In an initial analysis of *cis*-acting sequences regulating myosin gene expression, we identified a 395 bp fragment from *myo-2* that functions as a pharyngeal muscle-specific enhancer (Okkema et al., 1993). This enhancer was defined by its ability to induce pharyngeal muscle expression from a *myo-3::lacZ* fusion which is normally expressed only in body wall muscle (Fig. 1A). For these experiments, we used an F_1 expression

assay to characterize enhancer function (see Okkema et al., 1993 for discussion). Using this assay to delimit the enhancer further, we identified two overlapping fragments that also function as strong pharyngeal enhancers (Fig. 1B, pOK4.50 and pOK3.16). When assayed alone, the region of overlap functions only very weakly (pOK6.22). This analysis suggests the *myo-2* enhancer contains several functional components that cooperate to activate transcription.

Based on the deletion analysis, we provisionally divided the *myo-2* enhancer into three overlapping fragments, designated *A*, *B* and *C* (Fig. 1B). These fragments are individually inactive, but segments spanning *A*+*B* or *B*+*C* function as strong pharyngeal enhancers. A plausible working model is that the *A*, *B* and *C* fragments each contain a discrete subelement, with two or more subelements necessary for activity. Consistent with this model, we found that duplications of *A*, *B* or *C* function as pharyngeal enhancers (Fig. 1C). Thus, each of these fragments contains sufficient information to independently activate transcription when duplicated. A combination of the *C* and *A* fragments also functions as a pharyngeal enhancer (Fig. 1C), indicating that the three fragments are mutually synergistic.

To test promoter requirements for the individual subelements, we assayed these duplications and combinations of fragments upstream of two additional promoters (data not shown). The *glp-1* promoter fragment used shows a background of rare sporadic staining with no bias towards muscle, while a deleted *myo-2* promoter (Δ*myo-2*) shows essentially background activity (Okkema et al., 1993). Both of these promoter segments are sensitive to transcriptional enhancement in a variety of tissues (Okkema et al., 1993; P. O., V. Jantsch-Plunger and A. F., unpublished data). Both the *B* and *C* fragments function identically with all promoters tested: a single copy of either displays little or no enhancer activity; while two copies induce abundant pharyngeal expression. The *B*+*C* enhancer likewise activates the *glp-1* and Δ*myo-2* promoters. In contrast, *A*+*A*, *C*+*A* and *A*+*B* are unable to enhance expression from either the *glp-1* or Δ*myo-2* promoters, although each is able to activate *myo-3*. Therefore, function of the *A* fragment with the *myo-3* promoter appears to require specific enhancer-promoter interactions.

The *B* and *C* fragments have distinct cell type specificities

The pharyngeal muscles of *C. elegans* can be grouped into 8 types (m1-m8) based on cell morphology and position (described in detail in Albertson and Thomson, 1976). The muscles are arranged in layers along the anterior-posterior axis that are three-fold rotationally symmetric, each containing 1-3 cells of a single type (Fig. 2A). The m3-m7 muscles are large and define the overall contour of the pharynx. Smaller m1, m2 and m8 muscles are located at the anterior or posterior ends of the pharynx. The pharynx also contains epithelial cells, neurons, specialized marginal cells and secretory gland cells.

The cell type specificity of various enhancer constructs was analyzed by identification of individual cells expressing enhancer driven *lacZ* fusions. The F_1 expression assay used for this analysis generates animals that exhibit mosaic expression of the transforming DNA. We have used the frequency of staining as a measure of enhancer strength in each cell type (e.g., Weintraub, 1988). A *myo-2::lacZ* fusion containing the

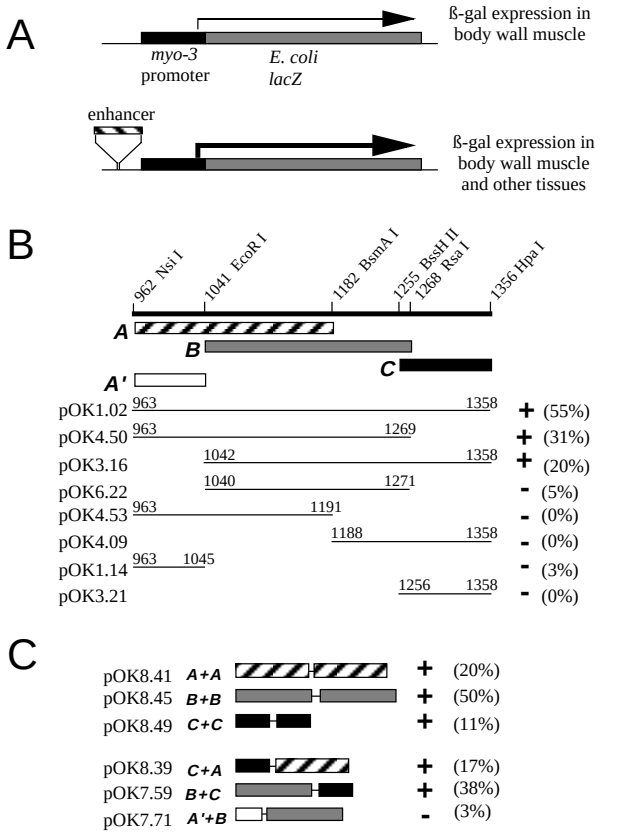
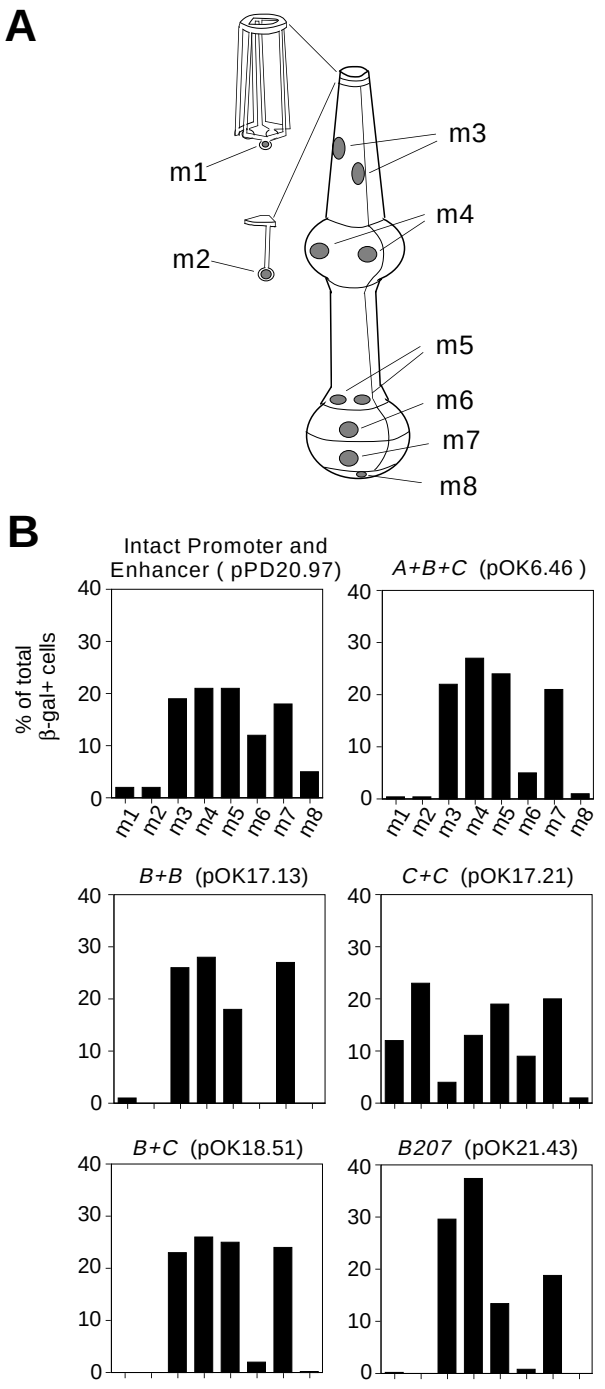


Fig. 1. The *myo-2* enhancer contains multiple elements that cooperate to activate transcription. (A) To characterize the *myo-2* enhancer, DNA fragments are cloned upstream of a *myo-3::lacZ* fusion, which is normally expressed only in body wall muscle. The resulting plasmids are injected into the germline of adult hermaphrodites (Mello et al., 1991) and F_1 progeny are stained to determine if β-galactosidase activity is induced in pharyngeal muscle. Most of the transformed F_1 animals are mosaic for β-gal expression, yet numerous such animals are produced from each injection. Thus, each expression pattern is determined by observing a large number of transformed animals. We have found that the frequency with which expression is observed is indicative of enhancer strength (Okkema et al., 1993). (B) The thick line at top is a partial map of the enhancer indicating restriction sites used for deletions (numbered according to Dibb et al., 1983). Thin lines indicate the precise extent of segments tested for enhancer activity upstream of the *myo-3::lacZ* fusion (Okkema et al., 1993). Boxes drawn below the enhancer map indicate the *A*, *B* and *C* regions. *A′* indicates a segment also tested for transcriptional activity. Pharyngeal muscle enhancer strength is reported as the percentage of total β-gal-positive animals expressing in pharyngeal muscle. The construct was scored as '–', if pharyngeal muscle expression was seen in 5% or less of the animals expressing in body wall muscle. Total numbers of β-gal-positive animals scored were: pOK1.02 [20]; pOK4.50 [49]; pOK3.16 [60]; pOK6.22 [75]; pOK4.53 [95]; pOK4.09 [74]; pOK1.14 [71]; and pOK3.21 [95]. (C) Combining any two subelements creates a functional enhancer. Combinations of enhancer fragments tested for activity upstream of *myo-3::lacZ* are shown schematically. Total numbers of β-gal-positive animals scored were: pOK8.41 [35]; pOK8.45 [20]; pOK8.49 [27]; pOK8.39 [66]; pOK7.59 [88]; and pOK7.71 [112]. The *A*, *B*, *C* and *A′* fragments include base-pairs 963-1191, 1040-1271, 1256-1358 and 963-1045 respectively.

entire enhancer and promoter region is expressed in all pharyngeal muscle cell types (pPD20.97, Figs 2B, 3A; Okkema et al., 1993). Expression of this fusion is most frequent in muscle cells m3-m7, with less frequent expression in m1, m2 and m8. The *A+B+C* enhancer gives a similar expression pattern when assayed with Δ*myo-2* or *glp-1* promoters, although the relative frequency of expression is somewhat reduced in m1, m2, m6 and m8 (Figs 2B, 3B).

The *B* and *C* fragments activate transcription in distinct sets of cells in the pharynx. The *B+B* enhancer (pOK17.13) activates frequent expression only in pharyngeal muscles m3, m4, m5 and m7, with occasional expression in m1 (Figs 2B, 3C). No expression has been observed in m6 or m2. In contrast, the *C+C* enhancer activates frequent expression in all pharyngeal muscles (Figs 2B, 3D). This distinction between the *C* and *B* fragments is particularly apparent in m1, m2 and m6, in which *C+C* is very active, while *B+B* is almost completely inactive. Unlike the other enhancers assayed, the *C+C* enhancer also activates expression in non-muscle cells in the pharynx. Fig. 3D shows an animal expressing β-gal in the e1 and e2 epithelial cells, as well as muscles m1, m2 and m4. We have observed expression induced by the *C+C* enhancer in all pharyngeal cell types including gland cells, neurons and marginal cells as well as muscles. These results define *C* as an organ-specific subelement.

Given the differences between the expression patterns induced by the *B+B* and *C+C* enhancers, it was of interest to determine the expression pattern of the *B+C* enhancer. This pattern is indicative of a restrictive rather than an additive interaction. The *B+C* (pOK18.51) enhancer is active only in cells where both *B* and *C* can be active (m3, m4, m5 and m7; Figs 2B, 3E).



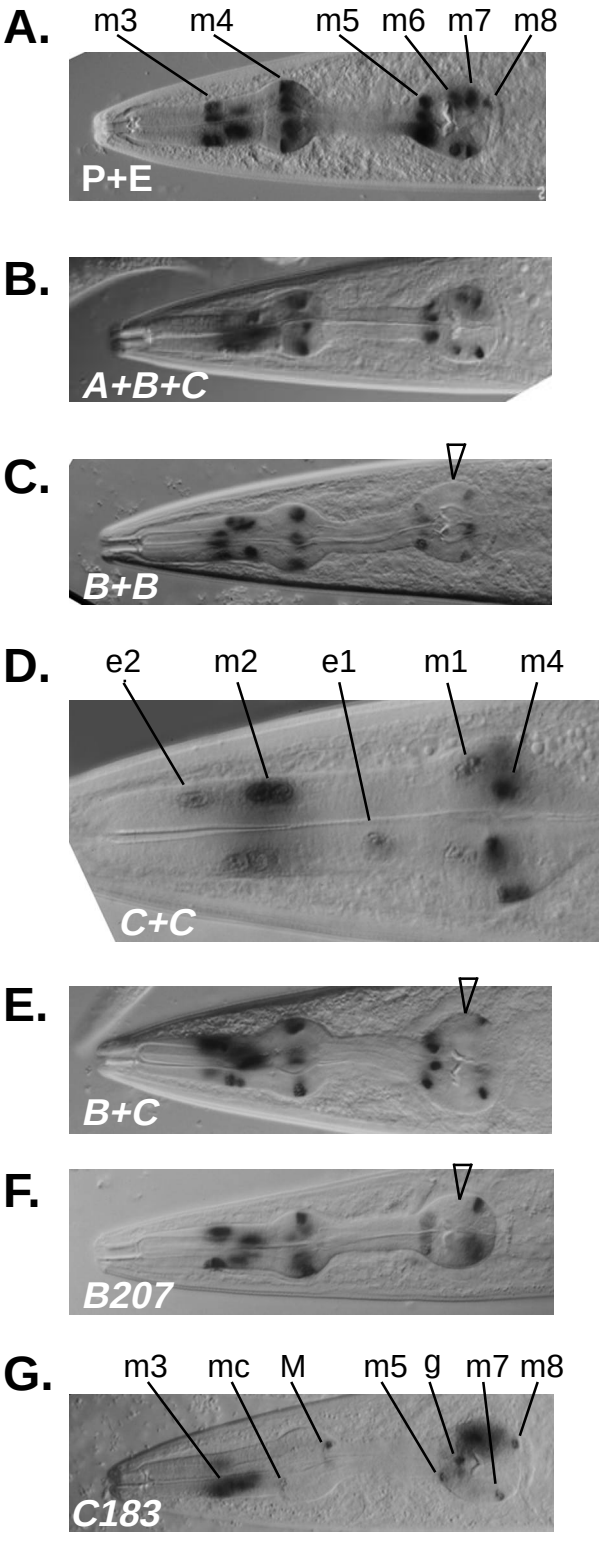
Discrete subelements within the *B* and *C* fragments are sufficient for cell-type-specific expression

To map precisely the subelements within the *B* and *C* fragments, we synthesized a set of overlapping double-stranded oligonucleotides spanning *B* and *C* (Fig. 4). The oligonucleotides were individually ligated to form head-to-tail multimers that were assayed for enhancer function upstream of the Δ*myo-2::lacZ* fusion (Table 1).

We might expect to find both general and cell-type-specific elements using this assay. Indeed, three of the oligonucleotides appear to contain general transcriptional activator elements. *B201*, *B203* and *B205* activate expression at low frequency in a variety of tissues and cell types (Table 1). Given our goal of understanding cell-type-specific regulation of *myo-2*, we have not investigated these general activities further.

Fig. 2. Duplications of the *B* and *C* fragments activate transcription in distinct sets of pharyngeal muscle cells. (A) Drawing of pharyngeal muscles m1-m8 (redrawn from Albertson and Thompson, 1976). Anterior is at top; posterior is at bottom. The approximate shape of individual muscle cells is indicated; shaded circles indicate positions of pharyngeal muscle nuclei. The shape of m2 is reversed from the original (L. Avery, personal communication). Muscle cells m2, m3, m4 and m5 are binucleate. m1 is a single cell encircling the entire pharynx that contains 3 pairs of nuclei located in posterior bulges. Pairs of m1 nuclei stain independently and were counted as separate cells. (B) Distribution of pharyngeal muscle cell types expressing β-gal from our original *myo-2::lacZ* transcriptional fusion (pPD20.97; Okkema et al., 1993) or various enhancer constructs assayed using Δ*myo-2::lacZ*. Identical results have been obtained in assays using *glp-1::lacZ* (data not shown). For each pharyngeal muscle type, the number of cells expressing β-gal is reported as the percent of the total β-gal-positive pharyngeal muscle cells. The number of β-gal-positive cells scored were: pPD20.97, 642 cells in 61animals; pOK6.46, 771 cells in 92 animals; pOK17.13, 430 cells in 60 animals; pOK17.21, 160 cells in 35 animals; pOK18.51, 272 cells in 28 animals; and pOK21.43, 487 cells in 93 animals. The identity of pharyngeal cells expressing β-galactosidase was determined by nuclear position, cell morphology and nuclear number. pOK6.46 contains an *Nsi*I-*Hpa*I fragment spanning *A*, *B* and *C*; pOK18.51 contains an *Eco*RI-*Hpa*I fragment spanning *B* and *C*, respectively.

Multimers of a single oligonucleotide from the *B* fragment (*B207*) activate pharyngeal expression in a pattern very similar to that observed with the duplicated *B* fragment, with expression predominately in pharyngeal muscles m3, m4, m5 and m7 (Table 1, Figs 2B, 3F). Unlike the larger *B* fragment, the *B207* oligonucleotide occasionally activates additional expression in cells other than pharyngeal muscle (Table 1).



Multimers of either of two overlapping oligonucleotides from the *C* fragment exhibit enhancer activity identical to that of the duplicated *C* fragment. *C181* and *C183* induce frequent expression in both muscle and non-muscle cells in the pharynx (Table 1; Figs 2B, 3G). Like the intact *C* fragment, these oligonucleotides activate expression only in the pharynx.

To identify regions in the *B* and *C* oligonucleotides necessary for transcriptional activation, we assayed a set of mutated oligonucleotides for enhancer activity (Fig. 5; Table 1). Mutations near each end of the *B* oligonucleotide *B207* (*Bmut1*, *Bmut2* and *Bmut4*) and each end of the *C* oligonucleotide *C183* (*Cmut1*, *Cmut2* and *Cmut4*) eliminate activity in the pharynx. In contrast, internal mutations in *B207* and *C183* (*Bmut3* and *Cmut3*, respectively) have little effect, suggesting that the activities of both the *B* and *C* subelements might require multiple binding sites. These analyses also suggest that the left end of *B* plays a critical role in specificity, since a mutation in this region (*Bmut1*) eliminates pharyngeal muscle expression without affecting the occasional non-pharyngeal expression (Table 1). Mutations to the right (*Bmut2* and *Bmut4*) drastically reduce both pharyngeal and non-pharyngeal activity.

A factor that specifically binds the *B* subelement

We used the *B207* and *C183* oligonucleotides to identify candidate genes that regulate *myo-2* enhancer activity. Phage plaques from a λgt11 cDNA library expressing *C. elegans* proteins were immobilized on nitrocellulose and probed with multimers of *B207* and *C183* (Singh et al., 1988; Vinson et al., 1988). Screening approximately 4×10^5 recombinant clones, we isolated three related cDNAs whose products specifically bind *B207*. Complete sequencing of two of these clones and restriction analysis of the third indicates that they are co-linear cDNAs encoded by a single gene that we have named *ceh-22* (*ceh* = ***C. e**legans* **h**omeobox; see below). In addition to the three *ceh-22* clones, two unrelated clones with binding specificities distinct from that of *ceh-22* were isolated. Analysis of the latter cDNAs is in progress and will be described elsewhere.

ceh-22 encodes an NK-2 class homeodomain protein

The predicted CEH-22 protein contains a homeodomain DNA-binding motif belonging to the phylogenetically conserved

Fig. 3. β-galactosidase expression induced by various *myo-2* enhancer constructs. (A) A construct containing the intact *myo-2* promoter and enhancer (pPD20.97; Okkema et al., 1993) expressed in pharyngeal muscle types m3-m8. β-gal activity is predominantly localized to cell nuclei because the fusion protein contains the SV40 nuclear localization sequence (Fire et al., 1990). (B-G) Expression induced by the *A+B+C* enhancer (B; pOK6.46) in pharyngeal muscle m3, m4, m5, m6 and m7; by the *B+B* enhancer (C; pOK17.13) in m3, m4, m5 and m7; by the *C+C* enhancer (D; pOK17.21) in muscles m1, m2, m4 and pharyngeal epithelial cells e1 and e2; by the *B+C* enhancer (E; pOK18.51) in pharyngeal muscles m3, m4, m5 and m7; by the concatenated *B207* oligonucleotide (F; pOK21.43) in pharyngeal muscles m3, m4, m5 and m7; and by the concatenated *C* subelement oligonucleotide *C183* (G; pOK19.62) in pharyngeal muscles m3, m5, m7, m8, a marginal cell (mc), a motor neuron (M) and gland cells (g). The positions of unstained m6 cells are indicated by arrowheads.

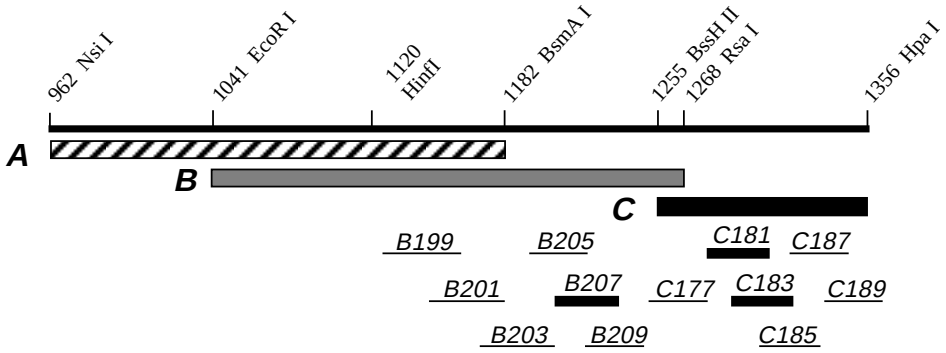


Fig. 4. Discrete sequences are sufficient for cell-type-specific activity of *B* and *C*. Partial restriction map of the *myo-2* enhancer indicating the extent of *A*, *B* and *C*. Numbered lines below indicate double-stranded oligonucleotides tested for enhancer activity when concatenated in 3-4 copies upstream of Δ*myo-2::lacZ*. Thick lines indicate the oligonucleotides *B207*, *C181* and *C183* that activate pharyngeal expression. To minimize the number of oligonucleotides, the left end of the *B* fragment was further defined by deletions not shown in Fig. 1.

Table 1. Enhancer activity of concatenated oligonucleotides from the *B* and *C* fragments

					Number of F_1 transformants expressing β-galactosidase in indicated tissue						
oligo	Location[a]	Clone (# inserts)	Tissue specific activity[b]	Total # F1 β-gal+	Pharyngeal muscle	Pharyngeal marginal cells	Other pharyngeal cells	Body wall muscle	Gut	Hypo-dermis	Neurons
Wild type											
None		pOK1134 (0)	no	0	0	0	0	0	0	0	0
B199	1120-1160	pOK20.81 (3)	no	1	1	0	0	0	0	0	0
B201	1147-1184	pOK20.82 (3)	no	25	1	4	0	3	8	2	7
B203	1171-1208	pOK21.14 (3)	no	8	1	2	0	3	2	0	2
B205	1194-1222	pOK21.24 (4)	no	17	3	3	1	1	0	1	4
B207	1208-1238	pOK21.43 (4)	yes	93	93	9	0	2	7	0	0
B209	1220-1250	pOK21.59 (4)	no	3	1	0	0	2	0	0	0
C177	1251-1278	pOK19.37 (3)	no	1	1	0	0	0	0	0	0
C181	1279-1306	pOK19.55 (3)	yes	>68	>68[c]	>68[c]	>68[c]	0	0	0	0
C183	1293-1320	pOK19.62 (4)	yes	>66	>66[c]	>66[c]	>66[c]	0	0	0	0
C185	1307-1335	pOK19.69 (4)	no	3	1	0	0	1	0	0	1
C187	1321-1348	pOK19.76 (4)	no	2	2	0	0	0	0	0	0
C189	1333-1362	pOK19.81 (4)	no	0	0	0	0	0	0	0	0
Mutants											
Bmut1		pOK22.08 (4)	no	25	5	3	2	4	13	0	0
Bmut2		pOK22.12 (4)	no	4	1	1	1	1	0	0	1
Bmut3		pOK22.21 (4)	yes	34	33	4	0	2	0	0	0
Bmut4		pOK22.31 (4)	no	2	1	1	0	0	0	0	0
Cmut1		pOK20.17 (3)	no	0	0	0	0	0	0	0	0
Cmut2		pOK20.25 (3)	no	0	0	0	0	0	0	0	0
Cmut3		pOK20.33 (3)	yes	>35	>35[c]	>35[c]	>35[c]	0	0	0	0
Cmut4		pOK20.52 (3)	no	3	1	0	3	0	0	0	0

[a]The nucleotide location of oligonucleotides in *myo-2* genomic DNA sequence according to Dibb et al. (1989).
[b]For each clone 8-12 adult hermaphrodites were injected; progeny were stained for β-galactosidase activity 3-4 days after injection.
[c]In many animals we were unable to unambiguously identify all pharyngeal cell types expressing β-galactosidase. Therefore, the total number of F_1 animals staining in the pharynx is included in each pharyngeal cell category.

NK-2 family (Fig. 6A). The CEH-22 homeodomain contains 8 of 9 conserved residues that define the NK-2 family and is most similar (87% identical) to the homeodomain of the *Drosophila* NK-2 gene (Kim and Nirenberg, 1990; Nardelli-Haefliger and Shankland, 1993). The CEH-22 homeodomain is 60% identical to those of *tin*/NK-4 and *bag*/NK-3 (Kim and Nirenberg, 1990; Bodmer et al., 1990; Azpiazu and Frasch, 1993). By contrast, CEH-22 shares only 27-50% identity with published *C. elegans* homeodomain sequences (data not shown). CEH-22 differs from other NK-2 class proteins in containing a serine rather than a conserved glutamine at homeodomain position 22 and an alanine rather than a conserved histidine at position 33. These amino acids are located in helices 1 and 2 of the homeodomain, respectively, and are not predicted to contact DNA (Kissinger et al., 1990).

Outside the homeodomain, CEH-22 shares no significant identity with any of the NK-2 family members. In particular, CEH-22 lacks both a conserved 17 amino acid peptide found downstream of the homeodomain and a decapeptide found upstream of the homeodomain in several family members (Price et al., 1992; Azpiazu and Frasch, 1993; Lints et al., 1993; Saha et al., 1993). An acidic region is located just upstream of the CEH-22 homeodomain (Fig. 6B). Highly acidic regions are also found upstream of the NK-1, NK-2 and *Dth-1* homeodomains (Kim and Nirenberg, 1990; Garcia-Fernàndez et al., 1991).

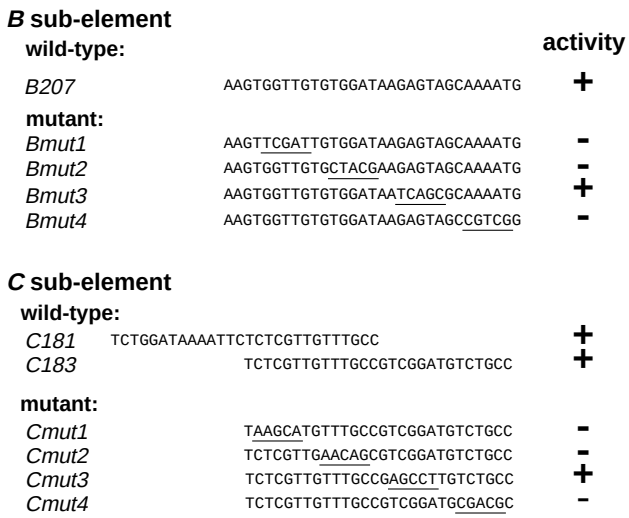


Fig. 5. Sequence of wild-type and mutant *B* and *C* oligonucleotides. Sequence of wild-type and mutant *B* subelement oligonucleotides (top) and *C* subelement oligonucleotides (bottom) tested for enhancer activity. Transcriptional activity of each is indicated by '+' or '–'. The nucleotides altered in the mutants are underlined. For the *C* subelement, we have focused on *C183*. The mutational analysis suggests a separate binding site at the left end *C181* can also contribute to *C* subelement activity.

Using reagents provided by the *C. elegans* genome project (Coulson et al., 1991), we mapped *ceh-22* to chromosome V, between *her-1* and the actin gene cluster. In a hybridization screen for homeoboxes, Bürglin and co-workers identified a hybridization signal in this region (locus 29; Bürglin et al., 1989). Further analysis of a cosmid from the region indicates the *ceh-22* homeobox indeed corresponds to locus 29 (data not shown; T. Bürglin, personal communication). A genomic fragment from this cosmid was sequenced, revealing that the *ceh-22* cDNA is derived from 7 exons spanning 3.2 kb (Fig. 6B). The homeobox is split by an intron within codon 53. Although this intron is absent in the NK-2 class homeoboxes for which genomic DNA sequence is available [NK-1,-2, *bag*/NK-3, *tin*/NK-4, *Dth-1* and *Dth-2* (Kim and Nirenberg, 1989; Garcia-Fernàndez et al., 1993)], an intron in this position is present in at least one non-NK-2 class homeobox in *C. elegans* (*ceh-19*; Bürglin, 1993).

ceh-22 encodes two RNAs that are most abundant in embryos and present in decreasing amounts throughout development (Fig. 7). The relative abundance of the two *ceh-22* RNAs is modulated during development: the 1.5 kb RNA is present at higher levels than the 1.45 kb RNA in embryos; while the two are present at roughly equal levels in late larvae. Structural or functional differences between the two RNAs have not yet been determined. A low level of *ceh-22* RNA is detected in adults. At least a fraction of this adult expression is in the soma (data not shown), since the RNA is detected in mutant *glp-4(bn2)* animals that contain very few germ cells (Beanan and Strome, 1992).

To define the CEH-22-binding site, we used recombinant CEH-22 protein purified from *E. coli* to generate a DNAseI footprint on the *B* fragment of the *myo-2* enhancer (Fig. 8). At high concentrations, recombinant CEH-22 protects the sequence TAAAGTGGTTGTGTG, which overlaps the 5′ end of *B207* by 13 bp. This sequence contains a TNNAGTG which is present in consensus binding sites for the NK-2 class homeoproteins TTF-1 and NK-2 (Guazzi et al., 1990; M. Nirenberg, personal communication). The mutation *Bmut1*, which affects this consensus sequence, eliminates *B* subelement activity in vivo (Fig. 5). An identically prepared protein containing a deletion removing helix 3 of the homeodomain fails to footprint the *B* fragment (data not shown).

CEH-22 is expressed in pharyngeal muscles m3, m4, m5 and m7

Immunostaining was used to determine the temporal and spatial expression pattern of CEH-22. Antibodies were raised separately against two CEH-22 fusion proteins purified from *E. coli* by different protocols and affinity purified (see Methods). Identical staining patterns were observed with both antisera. Staining is limited to nuclei within the pharynx and is detected from the beginning of morphogenesis onwards (Fig. 9A-C). When CEH-22 is first detected (approximately 330 minutes after fertilization; the lima bean stage), all 37 pharyngeal muscle nuclei are present (Sulston et al., 1983). At this stage, CEH-22 is detected in 11-14 pharyngeal nuclei (Fig. 9A). Positive identification of the CEH-22 containing nuclei is difficult; however, their positions suggest they are pharyngeal muscles. As the embryo elongates to the $1\frac{1}{2}$-fold stage, the number of CEH-22-positive nuclei in the pharynx increases to 14-23 (Fig. 9B). The wide range of staining nuclei suggests that, within a relatively short time period, a number of nuclei begin to accumulate CEH-22. CEH-22-positive cells were identified in these animals as muscles m3, m4, m5 and m7 by double staining with the monoclonal antibody 3NB12 (data not shown); 3NB12 had previously been shown to recognize a surface antigen in this set of pharyngeal muscles (Priess and Thompson, 1987). In embryos that have completed elongation (the pretzel stage), CEH-22-positive nuclei can be recognized by their characteristic positions as m3, m4, m5 and m7 (Fig. 9C). CEH-22 is also detected in 6 additional pharyngeal nuclei, which we believe are the m1 muscles. After hatching, CEH-22 remains detectable in m1, m3, m4, m5 and m7, but remains absent in m6 and m2 (data not shown).

We have also examined expression of a *ceh-22::lacZ* fusion in transgenic *C. elegans*. This construct contains approximately 4 kb of *ceh-22* 5′-flanking DNA fused to *lacZ* within the *ceh-22* 5′-UTR (see Fig. 6B). The timing and distribution of β-galactosidase expression in transgenic animals containing this fusion is identical to the staining pattern observed using anti-CEH-22 antibodies (Fig. 9D-F).

DISCUSSION

Modular structure of the *myo-2* enhancer

We have identified three regions of the *myo-2* enhancer, which we call *A*, *B* and *C*, that function together to activate transcription. Within the *B* and *C* regions, we have identified short oligonucleotides with activities virtually identical to the intact fragments. These results support the model that discrete subelements are combined to form the *myo-2* enhancer.

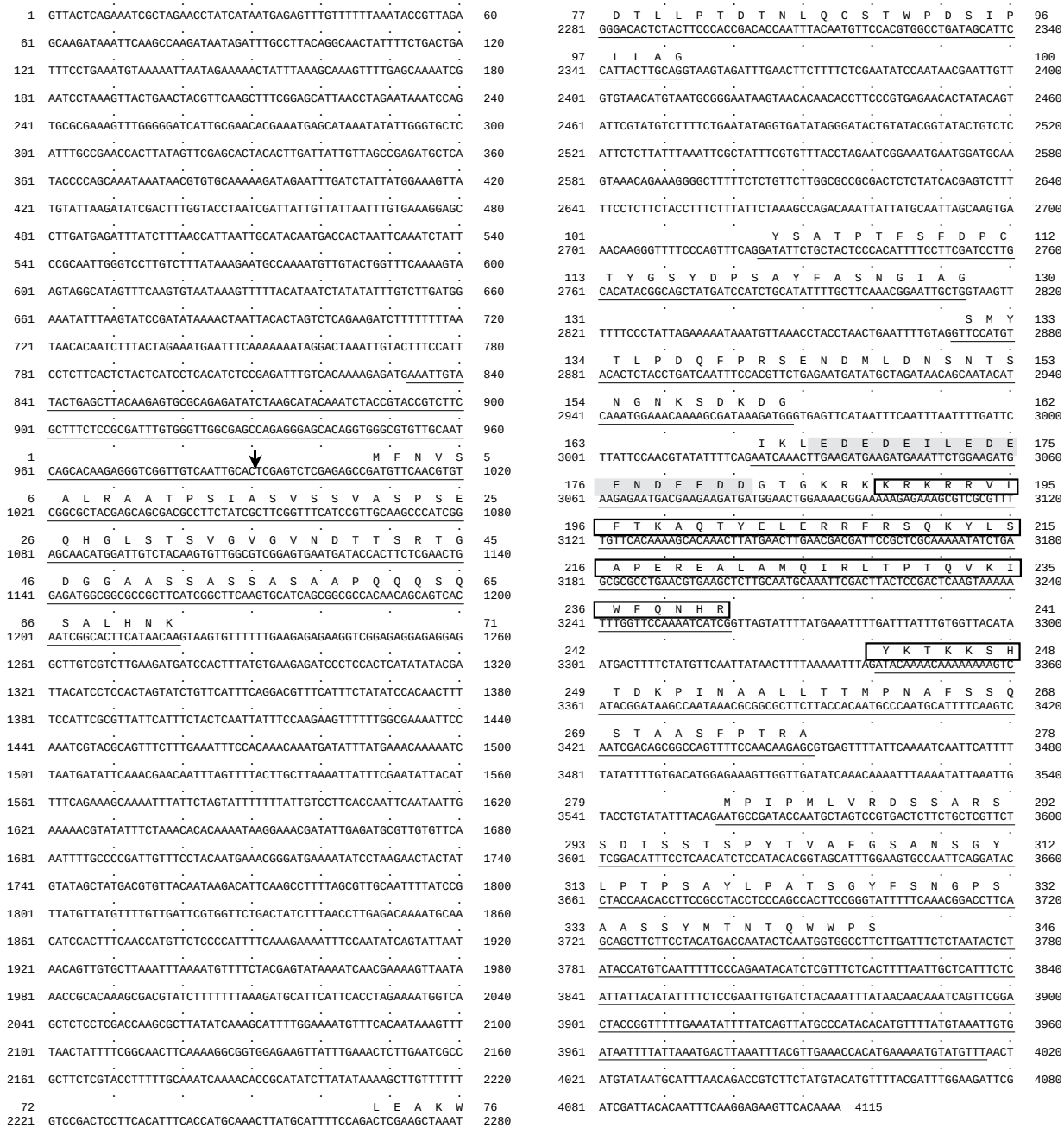
The *B* and *C* subelements appear to contain multiple sites necessary for activity. Mutations at each end of these subele-

A

```
ceh-22                  KRKRRVLFTK AQTYELERRF RSQKYLSAPE REALAMQIRL TPTQVKIWFQ NHRYKTKKSH
                              *       *    *       *     *    *    *          * *
Drosophila NK-2         ---------- ---------- -Q-R------ --H--SL--- ---------- -------RAQ  87
mouse Nkx-2.2           --------S- ---------- -Q-R------ --H--SL--- ---------- -----M-RAR  83
Xenopus XlNK-2          --------S- ---------- -Q-R------ --H--SL--- ---------- -----M-RAR  83
leech Lox-10            R----I--SQ --I------- -Q-------- --H--TF-G- ---------- ---------K  82
rat TTF-1               R--------SQ --V------- KQ-------- --H--SM-H- ---------- -----M-RQA  77
planarian Dth-2         R----I--SQ --I------- KQ-------- --H--NL-N- ---------- -----C-R-Q  77
planarian Dth-1         --------S- K-IL----H- -QK------- --H--NL-G- S--------- -----M-RA-  75
mouse Nkx-2.5/Csx       R--P----SQ --V------- KQ-R------ -DQ--SVLK- -S-------- -R---C-RQR  67
mouse Nkx-2.3           R--P----SQ --VF------ KQ-R------ --H--SSLK- -S-------- -R---C-RQR  67
flatworm EgHbx3         QS------N- F-ISQ--K-- -K-R--T-Q- -QE--HT-G- ---------- --A--M-RLF  63

Drosophila tin/NK-4     ---P----SQ --VL---C-- -LK---TGA- --II-QKLN- SA-------- -R---S-RGD  60
Drosophila bag/NK-3     -KRS-AA-SH --VF------ AQ-R---G-- -SEM-KSL-- -E-------- -R-----RKQ  60
```

B



ments eliminate activity, whereas mutations nearer the center have no effect. Thus two levels of organization apparently exist in the *myo-2* enhancer: subelements separated by approximately 100-200 bp are each composed of a cluster of binding sites for interacting transcription factors.

Similar modular organization is a common feature of transcriptional enhancers. The structure of the SV40 enhancer has been analyzed in detail (reviewed in Wildeman, 1988). It is composed of several elements of differing specificity that cooperate to activate transcription (Herr and Clarke, 1986; Ondek et al., 1987). The elements each in turn contain multiple sites necessary for activity (Ondek et al., 1988). This type of organization has also been demonstrated for a vertebrate muscle-specific enhancer from the creatine kinase gene (Cserjesi et al., 1992).

Muscle-type-specific and organ-specific pathways converge to activate the *myo-2* enhancer

The *B* and *C* subelements of the *myo-2* enhancer exhibit distinct cell type specificities. The *B* subelement is primarily active in the pharyngeal muscles m3, m4, m5 and m7. The *C* subelement is active in all pharyngeal cells. Thus distinct muscle-type-specific and organ-specific pathways converge to activate the intact enhancer.

What benefits might be realized by constructing the *myo-2* enhancer from muscle-specific and organ-specific subelements? The *B+C* combination is active in the same cells as the duplicated *B* subelement. Thus, cooperation with *C* does not refine cell type specificity of *B*. Perhaps the combination of the *B* and *C* subelements coordinates timing of pharyngeal muscle differentiation with that of other cell types in the pharynx. Proper formation of the pharynx might require differentiation of many cell types to be synchronized. The *C* subelement could be a target of a factor that synchronizes differentiation of the entire organ.

Independent evidence exists for a pathway specifying organ identity in the pharynx: the genes *pha-1* (Schnabel and Schnabel, 1990) and *pha-4* (S. Mango, E. Lambie and J. Kimble, personal communication) are specifically required for differentiation of many cell types in the pharynx. These genes might participate in pharyngeal differentiation by directly or indirectly activating regulatory sequences similar to the *myo-2 C* subelement.

CEH-22 activation of *myo-2* expression in pharyngeal muscle

We have identified a new homeobox-containing gene, *ceh-22*, which appears to be a key factor in activating *myo-2* expression through the *B* subelement. This conclusion is based upon both in vitro DNA-binding specificity and the in vivo expression pattern of CEH-22.

The *ceh-22* gene was initially identified by screening a mixed tissue cDNA expression library for clones encoding proteins that specifically bind the *B* subelement. No attempt was made to enrich the library for pharyngeal muscle cDNAs. After screening approximately 4×10^5 recombinant phage, we

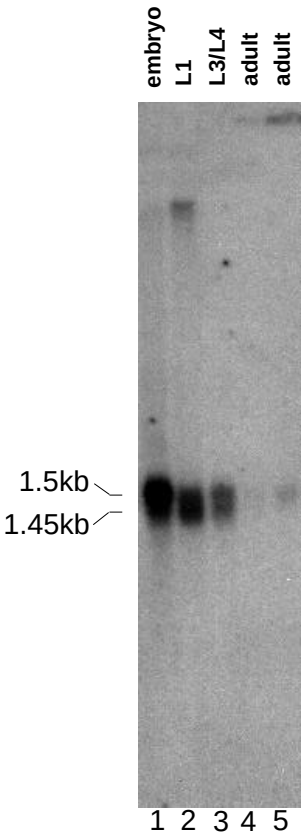


Fig. 7. *ceh-22* mRNA expression is highest during embryogenesis and decreases throughout development. Northern blot of poly(A)+ RNA isolated from developmentally synchronized populations of wild-type (N2) hermaphrodites probed with a *ceh-22* cDNA. Lanes 1-4 contain 1 μg poly-adenylated RNA from embryos, L1s, a mixed population of L3s and L4s, and adults, respectively. Lane 5 contains 2 μg adult poly(A)+ RNA to demonstrate the low level of *ceh-22* expression in adults. The blot was stripped and rehybridized with a probe from *unc-54* to verify that equal amounts of RNA were loaded in each lane (data not shown). The high molecular weight band in the L1 lane results from contaminating genomic DNA (data not shown).

Fig. 6. Homeodomain comparison and *ceh-22* genomic DNA sequence. (A) Comparison of the CEH-22 homeodomain (top) to NK-2 class (middle) and other related homeodomains (bottom). Identical amino acids are indicated by '–'. The position of conserved amino acids that define the NK-2 class are indicated by '*' (Nardelli-Haefliger and Shankland, 1993). Two amino acids in the CEH-22 homeodomain (position 22 and 33) that differ from those conserved in the other NK-2 class homeodomains are underlined. Sources for homeodomain sequences are: NK-2, NK-1, *tin*/NK-4, *bag*/NK-3 (Kim and Nirenberg, 1989); *Nkx-2.2*, *Nkx-2.3* (Price et al., 1992); *Nkx-2.5/Csx* (Lints et al., 1993; Komuro and Izumo, 1993); XlNK-2 (Saha et al., 1993); *Lox10* (Nardelli-Haefliger and Shankland, 1993); TTF-1 (Guazzi et al., 1990); *Dth-1* and *Dth-2* (Garcia-Fernàndez et al., 1991); EgHbx3 (Oliver et al., 1992). The published DNA sequences of *Nkx-2.5* and *Csx* appear identical with a frameshift mutation that shifts the reading frame downstream of the homeodomain. (B) Genomic DNA sequence of *ceh-22*. Nucleotides contained within the *ceh-22* cDNA are underlined. The predicted CEH-22 protein sequence is indicated above the DNA. The homeodomain sequence is boxed and the highly acidic region upstream of the homeodomain is shaded. The longest *ceh-22* cDNA contains a 1486 bp insert with a 3′ poly(A) tail. Based on the size of *ceh-22* mRNAs measured on northern blots, this cDNA is approximately full length. A 1215 bp open reading frame extends from the 5′ end of the cDNA and is fused in frame to the *E. coli lacZ* gene in the original λgt11 clone. In the *ceh-22* genomic sequence, a TGA stop codon blocks this reading frame immediately upstream of the cDNA. Because no potential splice acceptors that might be joined to an upstream exon are present in the genomic sequence, translation likely initiates at the first ATG 174 bp downstream of the cDNA 5′ end and would produce a 346 amino acid protein. Of the three *ceh-22* cDNA clones, two (clones 25 and 28) contain the indicated insert and appear identical by restriction mapping. The third (clone 27) contains a 0.9 kb insert that is co-linear with this cDNA, but is truncated at the 5′ and 3′ ends. This cDNA spans nucleotides 2287-3321 of the genomic sequence. The arrowhead at position 990 indicates the *Xho*I site used to construct the *ceh-22::lacZ* fusion pOK29.02.

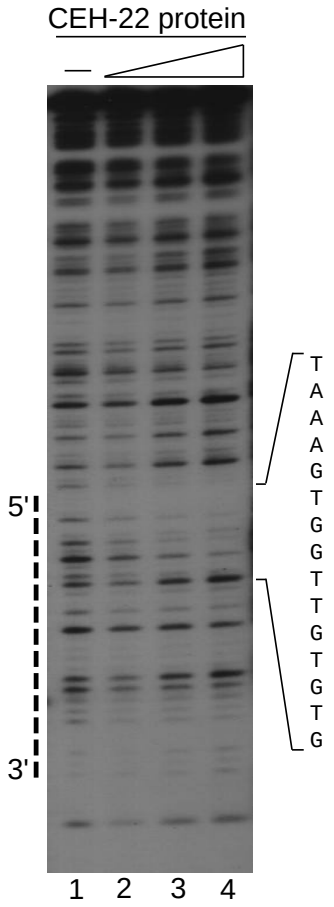


Fig. 8. Recombinant CEH-22 protein binds sequences essential for *B* subelement transcriptional activity. DNAseI footprint generated on the *B* fragment by recombinant CEH-22 protein. The probe was the *Eco*RI-*Bss*HII fragment labeled at the 3′ end on the top strand. The dashed line indicates the extent of the *B*207 oligonucleotide within the *B* fragment. The nucleotide sequence protected by CEH-22 is indicated at right. Lanes 1, 2, 3 and 4 contain 0 μg, 2.3 μg, 11.3 μg and 22.5 μg poly-his::CEH-22 protein, respectively (see Materials and Methods).

identified three clones encoding CEH-22 (of which at least two are independent). Recombinant CEH-22 protein footprints sequences necessary for in vivo activity of the *B* subelement. Moreover, the mutation *Bmut1* in this region, which eliminates in vivo activity of *B*, also decreases CEH-22 binding (unpublished observations). Thus a functional CEH-22-binding site appears necessary for transcriptional activity of the *B* subelement.

We have examined the *ceh-22* expression pattern using antibodies against CEH-22 protein and examining expression of a *ceh-22::lacZ* fusion. Both analyses indicate *ceh-22* is expressed in pharyngeal muscle prior to the onset of *myo-2* expression. To date, *ceh-22* expression is the earliest known marker of pharyngeal muscle differentiation. Strikingly, *ceh-22* expression is limited to the same subset of pharyngeal muscle cells in which the *B* subelement is active. This correlation is most apparent in the major muscles m3-m7: both *ceh-22* expression and activity of the *B* subelement are limited to m3, m4, m5 and m7; neither *ceh-22* expression nor frequent *B* subelement activity has been observed in m6. We expect that expression of other genes might be specifically activated by CEH-22 in m3, m4, m5 and m7; one candidate for this regulation is the gene encoding the antigen recognized by the monoclonal antibody 3NB12, which is expressed in this subset of pharyngeal muscles (Priess and Thompson, 1987).

Potential CEH-22-binding sites are present at additional locations in *myo-2* and in the other pharyngeal myosin gene, *myo-1*. In *myo-2*, one site (located at position 1504) is within a 33 bp segment necessary for activity of the pharyngeal muscle-specific promoter (Okkema et al., 1993). In *myo-1*, a site (located at position 3742) is within a 138 bp segment necessary for activity of the pharyngeal muscle-specific

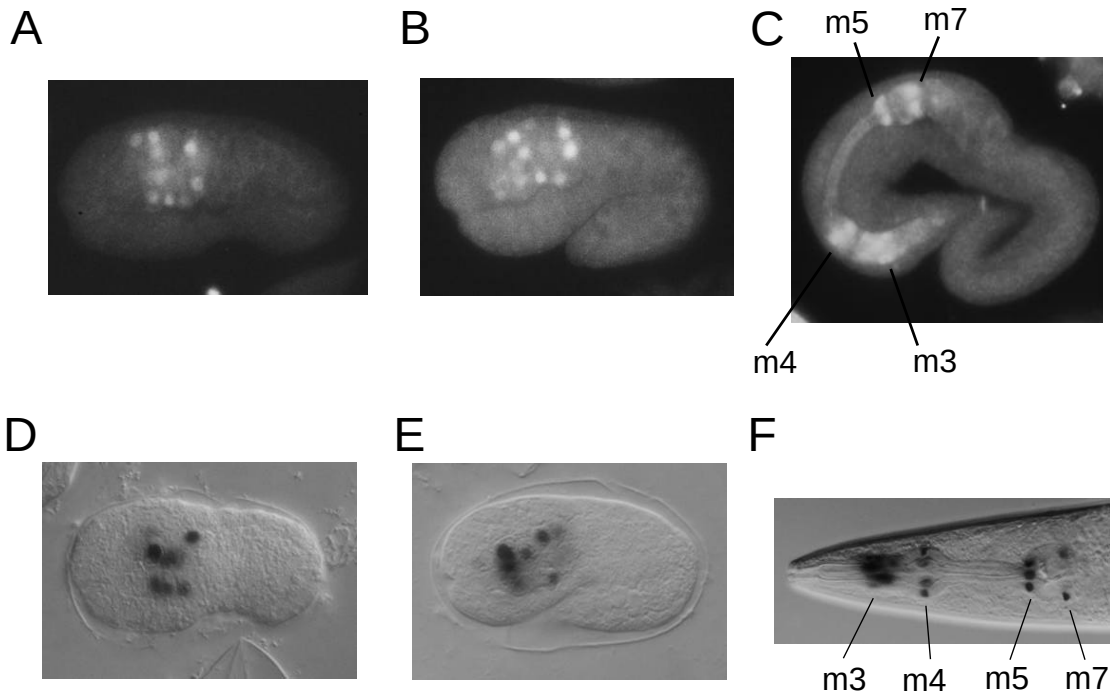


Fig. 9. Endogenous CEH-22 protein and a *ceh-22::lacZ* gene fusion are expressed in developing pharyngeal muscle. (A-C) Immunofluorescent localization of endogenous CEH-22 protein using affinity-purified c187 anti-CEH-22 antibodies. Examples of a lima bean stage embryo (A), a 1½-fold embryo (B) and a pretzel stage embryo (C) are shown. The developmental stage of stained embryos is inferred from the degree of morphogenesis: lima bean embryos have just completed the bulk of cell divisions; 1½-fold and pretzel stage are progressive stages of embryonic elongation. (D-F) Expression of a *ceh-22::lacZ* fusion detected by in situ staining with X-gal. Examples of β-gal expression in a bean stage embryo (D), 1½-fold embryo (E) and adult (F) are shown. In the pretzel stage embryo (C) and the adult (F), the positions of CEH-22-positive or β-gal-positive m3, m4, m5 and m7 nuclei are indicated. The *ceh-22::lacZ* fusion is integrated into a chromosome; however β-gal expression in many animals is mosaic. In a large number of stained animals expressing this *ceh-22::lacZ* fusion, no activity outside of pharyngeal muscle has been observed.

enhancer (Okkema et al., 1993). We have not yet tested whether these sites can bind CEH-22.

The fact that *B* subelement activity correlates very well with *ceh-22* expression suggests that cell type specificity of this subelement is largely determined by the expression pattern of *ceh-22*. While CEH-22 binding appears necessary for *B* subelement transcriptional activity, it is not sufficient. Three oligonucleotides (*B205*, *Bmut2* and *Bmut4*), which contain the CEH-22-binding site and bind CEH-22 protein in vitro, fail to activate transcription in vivo (unpublished observations; Fig. 5). These observations suggest a second factor binding to the *B* subelement is essential for activity. We have not yet identified that factor but we are currently screening for proteins that bind this second critical site.

CEH-22-independent activation of *myo-2* in m6

Previous workers have reported morphological and immunological differences between pharyngeal muscle m6 and other pharyngeal muscles. Albertson and Thompson (1976) observed vesicles in m6 that were absent in other pharyngeal muscle types. They suggested m6 may be specialized to secrete the pharyngeal grinder, a cuticular structure associated with this cell type. Priess and Thompson (1987) noted the monoclonal antibody 3NB12 recognizes a surface antigen in all the large pharyngeal muscle cells except m6.

Our results indicate the m6 cells also activate *myo-2* expression differently than other pharyngeal muscles, via a CEH-22-independent mechanism. How then is *myo-2* expression induced in m6? The complete *myo-2* enhancer can activate expression in all large pharyngeal muscle cells including m6, whereas an enhancer composed of the *B* and *C* subelements activates transcription only in m3, m4, m5 and m7. Thus sequences in the *A* subelement may contribute to expression in m6, perhaps in combination with the *C* subelement.

ceh-22 is a member of the NK-2 class of homeobox-containing genes

Comparison with known homeodomain sequences identifies CEH-22 as a member of NK-2 family of homeodomain factors (Kim and Nirenberg, 1990; Nardelli-Haefliger and Shankland, 1993). Members of this family are expressed in a variety of tissues, suggesting they play diverse roles in development. The NK-2 class homeodomains most similar to that of CEH-22 (77-87% identical) are preferentially expressed in the central nervous system (CNS). The *Drosophila* NK-2, mouse *Nkx-2.2* and TTF-1, *Xenopus* XlNK-2 and leech *Lox10* genes are all expressed in the developing CNS (M. Nirenberg, personal communication; Price et al., 1992; Saha et al., 1993; Nardelli-Haefliger and Shankland, 1993). In addition to the nervous system expression, *Lox10* and NK-2 are expressed in midgut (Nardelli-Haefliger and Shankland, 1993; M. Nirenberg, personal communication), while TTF-1 is also expressed in thyroid and lung (Guazzi et al., 1990). The planarian *Dth-1* and *Dth-2* genes, whose homeodomains are approximately 75% identical to that of CEH-22, are expressed in intestine and unidentified peripheral parenchymal cells, respectively (Garcia-Fernàndez et al., 1993).

In contrast, three homeodomains that are somewhat less similar to that of CEH-22 (60-67% identical) are expressed in muscle tissues similarly to CEH-22. The mouse *Nkx-2.5/Csx* gene is expressed in cardiac muscle progenitors prior to myogenic differentiation and continues throughout development (Lints et al., 1993; Komuro and Izumo, 1993). *Nkx-2.5/Csx* expression is also detected in a subset of pharyngeal endoderm adjacent to the cardiac mesoderm, tongue muscle, visceral muscle in the stomach and spleen. The *Drosophila* gene *tinman* (*tin*) is initially expressed throughout the presumptive mesoderm and becomes restricted to cardiac and visceral muscle (Bodmer et al., 1990; Azpiazu and Frasch, 1993). Loss of *tin* activity results in absence of cardiac and midgut visceral muscle, and defects in a subset of dorsal body wall muscles (Azpiazu and Frasch, 1993; Bodmer, 1993). Likewise, the *Drosophila* gene *bagpipe* (*bag*) is expressed in a segmented pattern in visceral muscle and in a subset of cardiac muscles (Azpiazu and Frasch, 1993). Loss of *bag* activity results in segmental gaps in midgut visceral muscle (Azpiazu and Frasch, 1993).

The similarity in expression patterns between *ceh-22*, *Nkx-2.5/Csx*, *tin* and *bag* suggests that the function of these genes may be conserved. Interestingly, the developmental programs in vertebrate cardiac muscle, *Drosophila* cardiac and visceral muscle, and *C. elegans* pharyngeal muscle all occur without myogenic factors related to MyoD (Emerson, 1993; Michelson et al., 1990; Krause et al., 1990). An attractive hypothesis from the analysis of *ceh-22* and homologs in other species is that these homeodomain factors function in a phylogenetically conserved myogenic pathway occurring in muscle types that do not utilize the MyoD family.

We are indebted to V. Jantsch-Plunger for advice throughout this work, and to J. Ahnn, L. Avery, T. Bürglin, L. Chen, S. Dymecki, R. Jehan, W. Kelly, N. Patel, A. Pinder, G. Seydoux and C. Thompson for their help and suggestions. Sequence comparisons were done at the NCBI using the BLAST network service. This work was supported by the NIH (HD07532-03 to P. G. O.; R01-GM37706 to A. F.), the Carnegie Institution of Washington. A. F. is a Rita Allen Foundation scholar.

REFERENCES

Albertson, D.G. and Thomson, J.N. (1976). The pharynx of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **275**, 299-325.

Ardizzi, J.P. and Epstein, H.F. (1987). Immunochemical localization of myosin heavy chain isoforms and paramyosin in developmentally and structurally diverse muscle cell types of *C. elegans*. *J. Cell Biol.* **105**, 2763-2770.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1990). *Current Protocols in Molecular Biology*. John Wiley and Sons, New York.

Azpiazu, N. and Frasch, M. (1993). *tinman* and *bagpipe*: two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*. *Genes Dev.* **7**, 1325-1340.

Beanan, M.J. and Strome S. (1992). Characterization of a germ-line proliferation mutation in *C. elegans*. *Development* **116**, 755-66.

Bodmer, R., Jan, L.Y. and Jan, Y.N. (1990). A new homeobox containing gene, *msh-2*, is transiently expressed early during mesoderm formation of *Drosophila*. *Development* **110**, 661-669.

Bodmer, R. (1993). The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila*. *Development* **118**, 719-729.

Bürglin, T.R., Finney, M., Coulson, A. and Ruvkun G. (1989). *Caenorhabditis elegans* has scores of homeobox-containing genes. *Nature* **341**, 239-243.

Bürglin, T.R. (1993). A comprehensive classification of homeobox genes. In *A Guidebook to Homeobox Genes*, pp. 25-71, (ed. D. Duboule) Oxford: Oxford University Press.

Coulson, A., Kozono, Y., Lutterbach, B., Shownkeen, R., Sulston, J. and Waterston, R. (1991). YACs and the *C. elegans* genome. *BioEssays* **13**, 413-417.

Cserjesi, P., Lilly, B., Bryson, L., Wang, Y., Sassoon, D.A. and Olson, E.N. (1992). MHox: a mesodermally restricted homeodomain protein that binds an essential site in the muscle creatine kinase enhancer. *Development* **115**, 1087-1101.

Dibb, N.J., Maruyama, I.N., Krause, M. and Karn, J. (1989). Sequence analysis of the complete *Caenorhabditis elegans* myosin heavy chain gene family. *J. Mol. Biol*. **205**, 603-613.

Emerson, C.P. (1993). Skeletal myogenesis: genetics and embryology to the fore. *Curr. Opin. Genetics Dev.* **3**, 265-274.

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

Fire, A. (1993). Histochemical techniques for locating *Escherichia coli* ß-galactosidase activity in transgenic organisms. *Gene Analysis Techniques and Applications* **9**, 151-158

Garcia-Fernàndez, J., Baguñà, J. and Saló, E. (1991). Planarian homeobox genes: cloning, sequence analysis, and expression. *Proc. Natn. Acad. USA* **88**, 7338-7342.

Garcia-Fernàndez, J., Baguñà, J. and Saló, E. (1993). Genomic organization and expression of the planarian homeobox genes *Dth-1* and *Dth-2*. *Development* **118**, 241-253.

Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, M.-G. and Di Lauro, R. (1990). Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. *EMBO J.* **9**, 3631-3639.

Harlow, E. and Lane, D. (1988). *Antibodies, a Laboratory Manual*. Cold Spring Harbor NY: Cold Spring Harbor Laboratory.

Herr, W. and Clarke, J. (1986). The SV40 enhancer is composed of multiple functional elements that can compensate for one another. *Cell* **45**, 461-470.

Hochuli E., Döbeli, W. and Schacher, A. (1990). Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *BioTechnology* **6**, 1321-1325.

Jantsch-Plunger, V. (1993). Gene expression and regulation of the *unc-54* myosin heavy chain gene in *Caenorhabditis elegans*: characterization of an enhancer element and enhancer binding factors. Ph.D. thesis, University of Vienna.

Kim, Y. and Nirenberg, M. (1990). *Drosophila* NK-homeobox genes. *Proc. Natn. Acad. Sci. USA* **86**, 7716-7720.

Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8A resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**, 579-590.

Komuro, I. and Izumo, S. (1993). *Csx*: a murine homeobox-containing gene specifically expressed in the developing heart. *Proc. Natn. Acad. Sci. USA* **90**, 8145-8149.

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

Lints, T.J., Parsons, L.M., Hartley, L., Lyons, I. and Harvey, R.P. (1993). *Nkx-2.5*: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development* **119**, 419-431.

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**, 3959-70.

Michelson, A.M., Abmayr, S.M., Bate, M., Arias, A.M. and Maniatis, T. (1990). Expression of a MyoD family member prefigures muscle pattern in *Drosophila* embryos. *Genes Dev.* **4**, 2086-2097.

Miller, D.M., Stockdale, F.E. and Karn, J. (1986). Immunological identification of the genes encoding the four myosin heavy chain isoforms of *C. elegans. Proc. Natn. Acad. Sci. USA* **83**, 2305-2309.

Nardelli-Haefliger, D. and Shankland, M. (1993). *Lox10*, a member of the NK-2 homeobox gene class, is expressed in a segmental pattern in the endoderm and in the cephalic nervous system of the leech *Helobdella*. *Development* **118**, 877-892.

Okkema, P.G., Harrison, S.W., Plunger, V., Aryana, A. and Fire, A. (1993). Sequence requirements for myosin gene expression and regulation in *C. elegans*. *Genetics* **135**, 385-404.

Oliver, G., Vispo, M., Mailhos, A., Martínez, C., Sosa-Pineda, B., Fielitz, W. and Ehrlich, R. (1992). Homeoboxes in flatworms. *Gene* **121**, 337-342.

Ondek, B., Shepard, A. and Herr, W. (1987). Discrete elements within the SV40 enhancer region display different cell-specific enhancer activities. *EMBO J.* **6**, 1017-1025.

Ondek, B., Gloss, L. and Herr, W. (1988). The SV40 enhancer contains two distinct levels of organization. *Nature* **333**, 40-45.

Price, M., Lazzaro, D., Pohl, T., Mattei, M.-G., Rüther, U., Olivo, J.-C., Duboule, D. and Di Lauro, R. (1992). Regional expression of the homeobox gene *Nkx-2.2* in the developing mammalian forebrain. *Neuron* **8**, 241-255.

Priess, J. and Thompson, N. (1987). Cellular interactions in early *Caenorhabditis elegans* embryos. *Cell* **48**, 241-250.

Saha, M.S., Michel, R.B., Gulding, K.M. and Grainger, R.M. (1993). A *Xenopus* homeobox gene defines dorsal-ventral domains in the developing brain. *Development* **118**, 193-202.

Singh, H., LeBowitz, J.H., Baldwin, A.S. and Sharp, P.A. (1988). Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. *Cell* **52**, 415-423.

Schnabel, H. and Schnabel, R. (1990). An organ-specific differentiation gene, *pha-1*, from *Caenorhabditis elegans. Science* **250**, 686-688.

Smith, D.B. and Johnson, K.S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-S-transferase. *Gene* **67**, 31-40.

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

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

Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschulz, W.H. and McKnight, S.L. (1988). In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Dev.* **2**, 801-806.

Weintraub, H. (1988). Formation of stable transcription complexes as assayed by analysis of individual templates. *Proc. Natn. Acad. Sci. USA* **85**,5819-5823.

White, J. (1988). The Anatomy. In *The Nematode Caenorhabditis elegans* (ed. W.B. Wood). pp. 81-122, Cold Spring Harbor New York: Cold Spring Harbor Laboratory.

Wildeman, A.G. (1988). Regulation of SV40 gene expression. *Biochem. Cell Biol.* **66**, 567-577.

Yu, Y.-T., Breitbart, R.E., Smoot, L.B., Lee, Y., Mahdavi, V. and Nadal-Ginard, B. (1992). Human myocyte-specific enhancer factor 2 comprises a group of tissue restricted MADS box transcription factors. *Genes Dev.* **6**, 1783-1798.

Zhu, H., Nguyen, V.T.B., Brown, A.B., Pourhosseini, A., Garcia, A.V., van Bilsen, M., Chien, K.R. (1993). A novel, tissue-restricted zinc-finger protein (HF-1b) binds to the cardiac regulatory element (HF-1b/MEF-2) in the rat myosin light-chain2 gene. *Mol. Cell. Biol.* **13**, 4432-4444.

(Accepted 4 May 1994)

Note added in proof

The *ceh-22* cDNA and genomic DNA sequences have been submitted to GenBank (accession numbers, U10080 and U10081, respectively).