Copyright © 1993 by the Genetics Society of America

Sequence Requirements for Myosin Gene Expression and Regulation in *Caenorhabditis elegans*

Peter G. Okkema, Susan White Harrison,[1] Verena Plunger, Arash Aryana[2] and Andrew Fire

Carnegie Institution of Washington, Department of Embryology, Baltimore, Maryland 21210

Manuscript received April 30, 1993
Accepted for publication June 17, 1993

ABSTRACT

Four *Caenorhabditis elegans* genes encode muscle-type specific myosin heavy chain isoforms: *myo-1* and *myo-2* are expressed in the pharyngeal muscles; *unc-54* and *myo-3* are expressed in body wall muscles. We have used transformation-rescue and *lacZ* fusion assays to determine sequence requirements for regulated myosin gene expression during development. Multiple tissue-specific activation elements are present for all four genes. For each of the four genes, sequences upstream of the coding region are tissue-specific promoters, as shown by their ability to drive expression of a reporter gene (*lacZ*) in the appropriate muscle type. Each gene contains at least one additional tissue-specific regulatory element, as defined by the ability to enhance expression of a heterologous promoter in the appropriate muscle type. In rescue experiments with *unc-54*, two further requirements apparently independent of tissue specificity were found: sequences within the 3′ non-coding region are essential for activity while an intron near the 5′ end augments expression levels. The general intron stimulation is apparently independent of intron sequence, indicating a mechanistic effect of splicing. To further characterize the myosin gene promoters and to examine the types of enhancer sequences in the genome, we have initiated a screen of *C. elegans* genomic DNA for fragments capable of enhancing the *myo-2* promoter. The properties of enhancers recovered from this screen suggest that the promoter is limited to muscle cells in its ability to respond to enhancers.

MUSCLE differentiation is characterized by coordinated production of a large group of proteins that subsequently assemble to form contractile filaments and their associated structures. The nematode *Caenorhabditis elegans* provides advantages for dissecting muscle-specific gene regulation. *C. elegans* genes encoding many muscle structural components have been characterized in both genetic and molecular detail. Techniques for genetic transformation of *C. elegans* allow rapid assays for gene expression in a whole organism (FIRE 1986; MELLO *et al.* 1991); in particular, the availability of recessive mutations and cloned DNA for a given gene allows rescue of mutant phenotypes to be used as an assay to determine requirements for gene expression (WAY and CHALFIE 1988; FIRE and WATERSTON 1989).

The myosin heavy chain genes have been the focus of extensive characterization by several laboratories, and were thus natural candidates for this analysis. Four genes encode muscle myosin heavy chain proteins in *C. elegans* (DIBB *et al.* 1989). Two of these genes, *myo-1* and *myo-2* are expressed only in the pharyngeal muscles, which are used by the animal to eat; the other two genes, *unc-54* and *myo-3*, are expressed in all of the muscle cells outside of the pharynx (ARDIZZI and EPSTEIN 1987; MILLER, STOCKDALE and KARN 1986). We use the term "body muscle" to refer as a group to the non-pharyngeal muscles. The most numerous of these are the 95 body-wall muscle cells used for locomotion. The body musculature also includes several "minor" muscle groups: four intestine-linked muscles that function in defecation, hermaphrodite-specific vulval and uterine muscles used in egg laying, and a set of male-specific muscles used for mating (WHITE 1988). The somatic sheath covering the hermaphrodite gonad is a myoepithelial cell which has also been shown to express *myo-3* and *unc-54* (ARDIZZI and EPSTEIN 1987).

The four myosin genes are located in different chromosomal regions (EPSTEIN, WATERSTON and BRENNER 1974; ALBERTSON 1985). These genes have similar structures (Figure 1) with 7–12 exons separated by relatively short introns (38–1069 bp); the entire coding regions are contained within 6.5–7.5 kb of genomic DNA (KARN, BRENNER and BARNETT 1983; DIBB *et al.* 1989). As has been found with other metazoans, virtually all *C. elegans* genes contain introns (BLUMENTHAL and THOMAS 1988; KLASS, AMMONS and WARD 1988; FIELDS 1990).

Of the four myosin genes, the *unc-54* gene is the most abundantly expressed and has been the focus of the bulk of previous analysis (WATERSTON 1988). Mu-

[1] Present address: Tufts University Sackler School of Biomedical Sciences, Boston Massachusetts.
[2] Present address: Johns Hopkins University School of Public Health, Baltimore, Maryland.

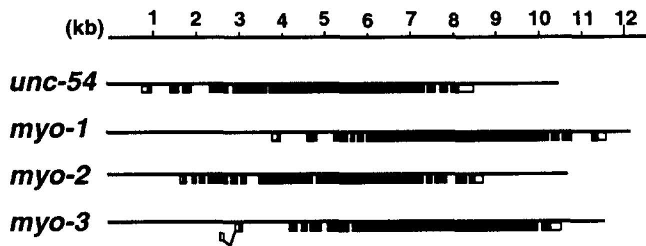


FIGURE 1.—Structure of the four *C. elegans* myosin heavy chain genes. Intron/exon structure of the myosin genes. Top line indicates the nucleotide numbering system used throughout (KARN, BRENNER and BARNETT 1983; DIBB *et al.* 1989). Filled boxes represent protein coding regions; open boxes represent untranslated regions present in the mRNAs. *myo-3* is trans-spliced to SL1 (see text). Diagrams are oriented 5′ (left) to 3′ (right).

tations eliminating *unc-54* function result in animals that are paralyzed but which are still viable (BRENNER 1974; EPSTEIN, WATERSTON and BRENNER 1974). This has allowed isolation of a large number of spontaneous and induced mutations at the locus (BRENNER 1974; EPSTEIN, WATERSTON and BRENNER 1974; MACLEOD, KARN and BRENNER 1981; ANDERSON and BRENNER 1984; EIDE and ANDERSON 1985a,b,c; BESJOVIC and ANDERSON 1990). All of the point mutations so far sequenced result from either nonsense or missense changes directly affecting the coding region (DIBB *et al.* 1985; BEJSOVIC and ANDERSON 1990). In addition, a variety of deletion and insertion mutations affecting the coding region have also been described (EIDE and ANDERSON 1985a,c). One result of the mutational analysis was those mutations causing premature termination of translation lead to instability of the message (MACLEOD *et al.* 1979; DIBB *et al.* 1985). Thus, steady state myosin mRNA levels in different mutants cannot be assumed to reflect transcription rates. Only three mutations with no evident effect on the coding region have been described. Two, *r104* and *r109*, are large duplications of sequences spanning the 5′ end of the gene (EIDE and ANDERSON 1985b). The third, *r293*, is a 256-bp deletion eliminating most of the 3′-untranslated region (PULAK and ANDERSON 1988). Unfortunately, the analysis of spontaneous and induced mutations has not readily yielded an understanding of signals responsible for tissue specific regulation.

Transcription of *unc-54* begins 85 bp upstream of the ATG used for initiation of translation (DIBB *et al.* 1989). Although many *C. elegans* genes possess standard eukaryotic transcription signals (TATAA and CAAT elements) upstream of their start sites (KLASS, KINSLEY and LOPEZ 1984; SPIETH *et al.* 1985; KAY *et al.* 1986), these signals are not evident in the region just upstream of *unc-54* (KARN, BRENNER and BARNETT 1983).

In situ hybridization studies have shown that *unc-54* mRNA accumulation occurs just in body muscle cells (KRAUSE 1986; G. SEYDOUX and A. FIRE, unpublished). This is consistent with (but does not prove) transcriptional regulation as the primary control mechanism mediating tissue specificity. Quantitative measurements of transcription, mRNA, and protein levels at different developmental stages have provided evidence for both transcriptional and post-transcriptional modulation of body wall myosin expression levels (HONDA and EPSTEIN 1990).

In this study we have used DNA transformation techniques to analyze the signals responsible for expression of myosin in *C. elegans*. Two types of elements have been discovered in this analysis. The first set are general signals which allow high levels of expression of the myosin message without any obvious contribution to tissue specificity. The second set are promoter and enhancer elements which appear to act at the transcriptional level to generate the observed tissue specificity of the genes.

MATERIALS AND METHODS

Genetic stocks and procedures: Genetic stocks used in this study were as follows:

CB190[*unc-54(e190)*I],
PD2071[*smg-1(e1228) unc-54(e1092)*I],
PD2073[*unc-54(r259)*I;*smg-6(r896)*III],
TR1325[*smg-1(r861) unc-54(r293)*I],
TR1326[*smg-2(r863) unc-54(r293)*I],
TR1328[*unc-54(r293)*I; *smg-6(r896)*III].

The latter three TR strains and *unc-54(r259)* were obtained from R. PULAK and P. ANDERSON (EIDE and ANDERSON 1985a; PULAK and ANDERSON 1988; HODGKIN *et al.* 1989). All other alleles were originally from the MRC (Cambridge) collection. *unc-54(e190)* is a 401-bp out-of-frame deletion in the *unc-54* coding region. *unc-54(e1092)* is an ochre termination codon at position 5343 (DIBB *et al.* 1985). *unc-54(r259)* (EIDE and ANDERSON 1985a) is a large (>17 kb) deletion removing several kilobases of 5′-flanking DNA and most of the coding region, and *unc-54(r293)* is a 256-bp deletion 3′ to the coding region (PULAK and ANDERSON 1988). Each of these mutations eliminates synthesis of *unc-54* product in an otherwise wild-type genetic background (EPSTEIN, WATERSTON and BRENNER 1974; MCCLEOD, WATERSTON and BRENNER 1977; FIRE and WATERSTON 1989; this work).

Construction of plasmids (FIRE, HARRISON and DIXON 1990), plasmid bookkeeping (MARCK 1988), microinjection of oocytes (FIRE 1986), microinjection of the syncytial gonad (MELLO *et al.* 1991), immunofluorescence (MILLER, STOCKDALE and KARN 1986; FIRE and WATERSTON 1989) and histochemical staining for β-galactosidase (FIRE 1993) were as described.

All experiments testing rescue of *unc-54* mutations were done by injection of oocytes (FIRE and WATERSTON 1989). The parent plasmid construct for the *unc-54* deletions used in rescue assays (pUNK54) was constructed by ITHIRO MARUYAMA and is described in FIRE and WATERSON (1989). The *lacZ* vectors utilized have been described previously (FIRE, HARRISON and DIXON 1990). *LacZ* fusion experiments were performed by both oocyte injection and syncytial gonad injections. Semiquantitative assessments are always based on injections performed using the same injection protocol. In the course of this work, *lacZ* promoter fusions

to each gene as well as a variety of enhancer test constructs have been injected using both types of injection protocol. A higher number of transgenic animals (approximately 5-fold) was obtained from syncytial injections of comparable numbers of animals, but staining patterns were identical for the two injection protocols.

Although most of the data described were obtained by directly staining the progeny of injected animals, similar results have been obtained using integrative transformation. *lacZ* fusions to *unc-54*, *myo-3* and *myo-2* were integrated using the marker *sup-7*. The resulting lines gave appropriate *lacZ* expression patterns, with one exception: while five independent transgenic lines derived from *unc-54::lacZ* stained just in body-wall tissue, a sixth line shows additional staining in pharyngeal muscle (data not shown). The behavior of this aberrant line could result from an integration site near to a chromosomal enhancer segment active in pharyngeal muscle.

Some of the *lacZ* promoter fusions (the *unc-54* fusions pPD19.64 and pPD18.42, and the *myo-2* fusions pPD18.56 and pPD20.97; see Figure 4) have also been introduced at high copy number into transformed lines as described (FIRE *et al.* 1991; MELLO *et al.* 1991). These constructs behave as expected in their staining patterns, although the staining in these high copy number lines is often much more mosaic than the inheritance of the injected DNA (similar observations have been made by M. MACMORRIS, J. SPIETH, and T. BLUMENTHAL, personal communication). In the case of the *unc-54::lacZ* fusion construct pPD18.42, a dominant effect of high copy numbers of the DNA was observed; animals carrying arrays of pPD18.42 show an uncoordinated phenotype characteristic of muscle defective animals (EPSTEIN, WATERSTON and BRENNER 1974).

Analysis of myosin mRNA termini: The 5′ ends of the myosin mRNAs were mapped by S1 nuclease protection (CALZONE, BRITTEN and DAVIDSON 1987). Briefly, 5′ ^{32}P-labeled, single-stranded DNA probes spanning the myosin gene transcription start sites were hybridized (42° overnight) with 25 μg total *C. elegans* RNA (prepared as described in ROSENQUIST and KIMBLE 1988) in 12 μl 52% formamide, 0.4 M NaCl, 40 mM PIPES (pH 6.5), 1 mM EDTA. This mixture was diluted on ice with 300 μl 0.5 M NaCl, 30 mM potassium acetate (pH 4.5), 2 mM $ZnSO_4$, 5% glycerol, 25 μg/ml each double-stranded and single-stranded herring sperm DNA. Five hundred to 2000 units S1 nuclease (Boehringer Mannheim) were added and the reactions incubated 90 min at 15–25°. Digestion was stopped by the addition of 80 μl 4 M ammonium acetate, 20 mM EDTA, 40 μg/ml tRNA. The reactions were extracted with phenol:chloroform, ethanol precipitated, electrophoresed on a 6% polyacrylamide sequencing gel, and visualized by autoradiography. Identically labeled dideoxy sequencing ladders were used as size standards. Probes spanning bp 2199–3948 of *myo-1*, 1256–1765 of *myo-2*, and 1504–2999 of *myo-3* were generated as described (CALZONE, BRITTEN and DAVIDSON 1987).

The 5′ ends of *myo-2* and *myo-3* were also mapped by polymerase chain reaction (PCR) amplification of primer extension products (FROHMAN, DUSH and MARTIN 1988; YAOITA, SHI and BROWN 1990). First strand cDNA was synthesized using gene specific oligo I (see below). Following 3′ oligo(dA) addition, second strand cDNA was synthesized using oligo P13. The products were amplified by PCR using oligo P0 and gene-specific oligo II (see below), cloned using restriction sites within the oligonucleotides, and sequenced.

Myosin heavy chain gene poly(A) addition sites were mapped by a modification of the procedure used by FROHMAN, DUSH and MARTIN (1988) *C. elegans* mixed stage total RNA was annealed to oligonucleotide P13 (YAOITA, SHI and BROWN 1990). Following extension with reverse transcriptase, the resulting cDNA was selectively amplified by PCR with a gene-specific sense oligonucleotide and an oligo covering the unique linker portion of P13 (P0). Products of these reactions were cloned using restriction sites in the oligonucleotides. A fraction of the resulting clones were spurious and unrelated to myosin genes, while the remainder were clones of myosin cDNAs, each containing a poly(A) tail. Poly(A) sites were determined from the junction sequences.

The *unc-54* 3′ end was confirmed by RNase protection assays (MELTON *et al.* 1984) performed under conditions described by FIRE *et al.* (1991). A uniformly ^{32}P-labeled antisense RNA probe covering bases 8953 to 8130 of the gene was synthesized using T3 RNA polymerase from a bluescript subclone of the corresponding DNA fragment. This probe protected a single prominent band of 340 ± 5 nucleotides (nt) in a mixed-stage RNA population. This band is close to the size expected (337 nt) for poly(A) addition at the site determined by PCR analysis.

Oligonucleotides: Gene specific oligos for mapping 5′ termini were:

myo-2 I(GCGAAGGTACTTCCATCCTGGGTCG),
myo-2 II(cggatcCGTTTTCGTAATCCATTTCTGTGT),
myo-3 I(GGATTTCCAGACATTTCTAGATGG),
myo-3 II(cggatccGGATCTAGTGGTCGTGGG).

Gene-specific oligos for mapping poly(A) sites were:

unc-54 (cgggatccATCAGATCGCCATCTCGCG),
myo-1 (cgggatccAAATTCCGACAGATCCAAC),
myo-2 (cggatccaTCTTAGCAAGTACAGAACC),
myo-3 (cggatccTGTCAAAGATGCGTAACAAG).

Myosin gene sequences are in uppercase.
Common oligos (YAOITA, SHI and BROWN 1990):

P13 (GTCGACATCGATCTCGAGT$_{18}$),
P0 (GTCGACATCGATCTCGAG).

Genomic library construction: To screen random genomic fragments for enhancer activity, *C. elegans* genomic DNA (prepared as described; SULSTON and HODGKIN 1988) was digested either with *Pst*I or with a combination of *Sph*I and *Sal*I. Fragments in the ranges 2.5–3.0 kb and 4.5–5.0 kb were excised from an agarose gel and cloned into the appropriate sites upstream of a *myo-2::lacZ* fusion (the parent construct was pPD18.56; Figure 4E). Eighteen fragments were analyzed; of these at least 17 were different as judged by *Hin*dIII digestion patterns.

RESULTS

Structure of the myosin heavy chain transcripts: The *unc-54* transcription start has been previously mapped to 79 bp upstream of the ATG translational initiation site (DIBB *et al.* 1989). We mapped the 5′ end of the *myo-1*, *myo-2* and *myo-3* mRNAs by S1 nuclease protection and sequencing PCR amplified primer extension products (data not shown). *myo-1* and *myo-2* start 136 and 83 bp, respectively, upstream of their ATG initiation codons (Figure 2a).

Consistent with previous observations (DIBB *et al.*, 1989), no consensus TATA is found upstream of either *unc-54*, *myo-1* or *myo-2*. The pentamer TTATC surrounding the *myo-1* and *myo-2* 5′ ends is similar to transcriptional start sites for a variety of abundantly

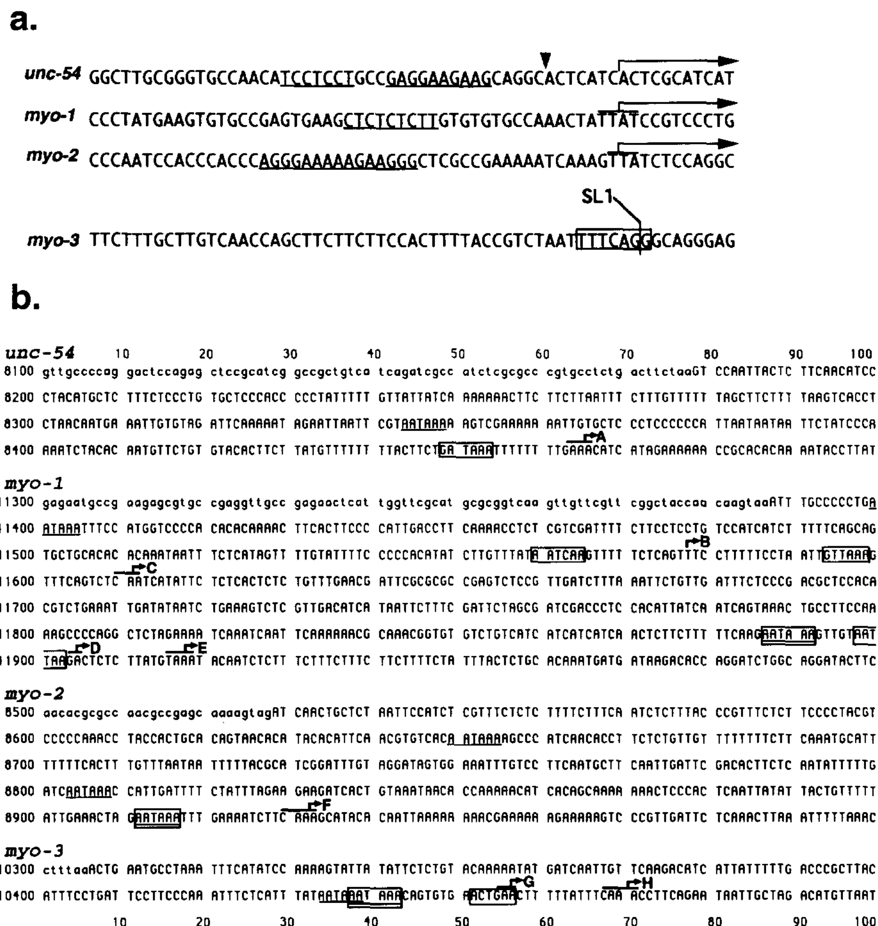


FIGURE 2.—5′ and 3′ ends of myosin heavy chain gene mRNAs. Genomic DNA sequence indicating the positions of 5′ and 3′ termini, determined by sequencing of PCR products and/or direct analysis of mRNA. (**a**) 5′ Termini. The major transcription start sites of *unc-54*, *myo-1* and *myo-2* are indicated by arrows. A minor start site for *unc-54* (arrowhead) has also been mapped (DIBB *et al.* 1989; A. FIRE unpublished). Ambiguities of plus or minus 1–2 bp for *myo-1* and *myo-2* are indicated by a line above the sequence. For *myo-1*, a cluster of bands at the indicated position were present in S1 nuclease protection assays; these bands could result either from clustered transcription start sites or incomplete protection from nuclease digestion. For *myo-2* adjacent T residues in the genomic sequence prevented precise determination of the transcription start using our protocol for PCR amplification of primer extension products (see MATERIALS AND METHODS). The *myo-3* trans-splice acceptor is boxed. Polypurine/polypyrimidine tracts upstream of the *unc-54*, *myo-1* and *myo-2* start sites are underlined. (**b**) 3′ Termini. For *unc-54* a single poly(A) addition site (labeled **A** in the figure) was identified by RNase protection assays and three independent clones amplified by PCR show the same 3′ end. Because the genomic sequence encodes three A residues at the point of poly(A) addition, there is a nonresolvable ambiguity of four adjacent base pairs, any of which might be the site of poly(A) addition. For the three other genes, several independent 3′ poly(A)-anchored PCR products were sequenced. For *myo-1*, a single clone with a junction at point **B**, two at point **C**, three at **D** and one at **E** were found in seven sequenced products. For *myo-2*, only a single clone was recovered, with a poly(A) junction at **F**. For *myo-3*, two poly(A) junctions were identified in three sequenced termini, **G** (2 clones) and **H** (1 clone). Some independent confirmation of *unc-54* and *myo-3* poly(A) addition sites comes from sequence studies of random cDNA clones (MCCOMBIE *et al.* 1992). One *unc-54* cDNA with a poly(A) junction at **A** and two *myo-3* cDNAs with poly(A) junctions at **G** were obtained in those studies. Canonical polyadenlyation signals (AATAAA) (WICKENS 1990) are underlined. Boxed hexanucleotide sequences 18–19 bp upstream of the identified poly(A) addition sites show a 4–6-bp match to the canonical signal. Myosin coding sequences are indicated by lowercase.

expressed *C. elegans* genes (*e.g.*, KRAUSE and HIRSH 1987; S. SPRUNGER and P. ANDERSON, personal communication). Although no striking conservation of the sequences upstream of the transcription start sites was evident, polypurine and/or polypyrimidine stretches are found preceding each of these myosin genes. In contrast to the other myosin genes, we found that *myo-3* is trans-spliced: the SL1 spliced leader (KRAUSE and HIRSH 1987) is added at an acceptor site 44 bp upstream of the *myo-3* ATG. The 5′ end of the *myo-3* primary transcript was not mapped.

The 3′ ends of the myosin mRNAs were mapped by PCR amplification of oligo(dT) primed cDNA (FROHMAN, DUSH and MARTIN 1988). We found a single polyadenylation site each for *unc-54* and *myo-2*, two sites for *myo-3*, and four sites for *myo-1* (Figure 2b). Nuclease protection experiments confirmed independently that the site identified by PCR was indeed the predominant poly(A) site in *unc-54* (data not shown); nonetheless, there could be additional poly(A) sites for both *unc-54* and the other myosin heavy chain genes.

The *myo-3* 3′ ends are located downstream of the first AATAAA polyadenylation site. Somewhat sur-

prisingly, *myo-1*, *myo-2* and *unc-54* contain 1–2 additional AATAAA sequences within their predicted 3'-UTRs (abbreviation: UTR = *U*n*T*ranslated *R*egion). Degenerate versions of the conserved AATAAA hexamer are found upstream of each of the 3' ends observed. This suggests, as has been observed in other systems (WICKENS 1990), that sequences distinct from the AATAAA consensus contribute significantly to the choice of 3' end.

Transformation assays: The transformation system in *C. elegans* allows gene expression to be assayed in transiently transformed animals or in transgenic lines (FIRE *et al.* 1991; MELLO *et al.* 1991). DNA microinjected directly into developing oocytes or into the syncitial germline of adult hermaphrodites is incorporated and expressed in many F_1 progeny, providing a fast and reliable assay for gene activity. These transiently transformed animals are generally mosaic, with only a subset of cells detectably expressing the injected DNA. In a population of transiently transformed animals, individuals vary in extent of mosaicism: some exhibit expression in just a few scattered cells of the positive tissue(s) while others show expression in all cells of a positive tissue. It is therefore necessary to assay expression in many transformed animals to determine which cell types can express a given construct. At relatively low frequency, lines can be isolated containing the transforming DNA in semistable extrachromosomal arrays or integrated into a chromosome. These lines can be used to generate large numbers of transformed animals suitable for antibody or histochemical staining.

We have used transformation assays with intact myosin genes and myosin gene::*lacZ* fusions to identify sequences responsible for expression. Neither transient transformants nor transformed lines yield precise quantitative measures of gene expression. We report results from injected constructs as negative or positive. In most cases we also give a qualitative measure of observed activity [either rescue of a mutant phenotype or staining for β-galactosidase (abbreviation = β-gal)]. These qualitative measures are based on both frequency and intensity of staining (for *lacZ* fusions) and degree of phenotypic rescue (for myosin coding constructs). When reproducible differences in transgene activity were observed, these are reported using a rough qualitative scale (+, ++, +++). It should be noted that we have not based any major conclusions solely on such differences. In several cases, negative results with individual constructions are critical. These negative results have been confirmed by assaying independently constructed plasmids.

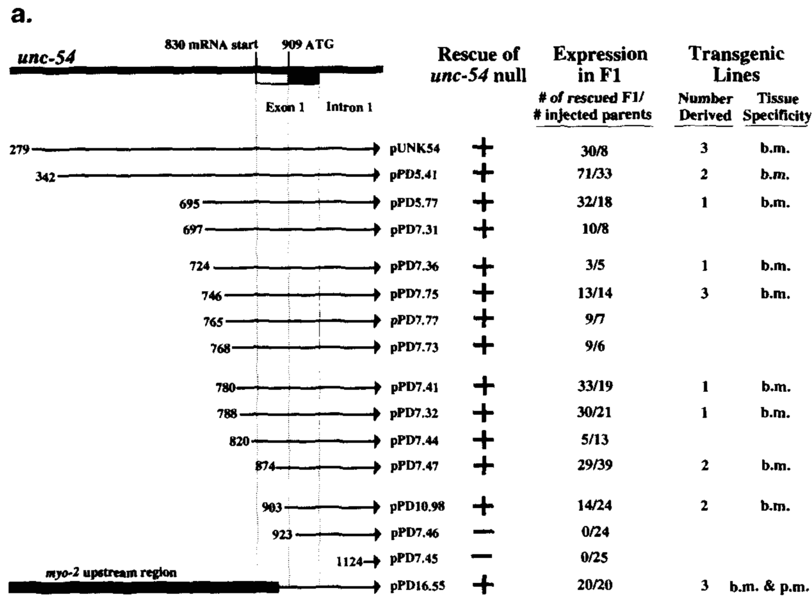
Transformation rescue of *unc-54* mutants: Null mutants in *unc-54* have markedly defective muscle function: the animals move slowly as larvae and are paralyzed as adults (BRENNER 1974; EPSTEIN, WATERSTON and BRENNER 1974). The adults fail to lay eggs due to paralysis of the muscles in the vulva (TRENT, TSUNG and HORVITZ 1983; ARDIZZI and EPSTEIN 1987). We have previously demonstrated that microinjection of the *unc-54* gene can yield phenotypic rescue of both paralysis and egg laying phenotypes, and that *unc-54* myosin is expressed in the appropriate muscles (FIRE and WATERSTON 1989). We began our analysis of sequence requirements for expression by testing clones with engineered 5' and 3' deletions for ability to rescue *unc-54(e190)* mutants.

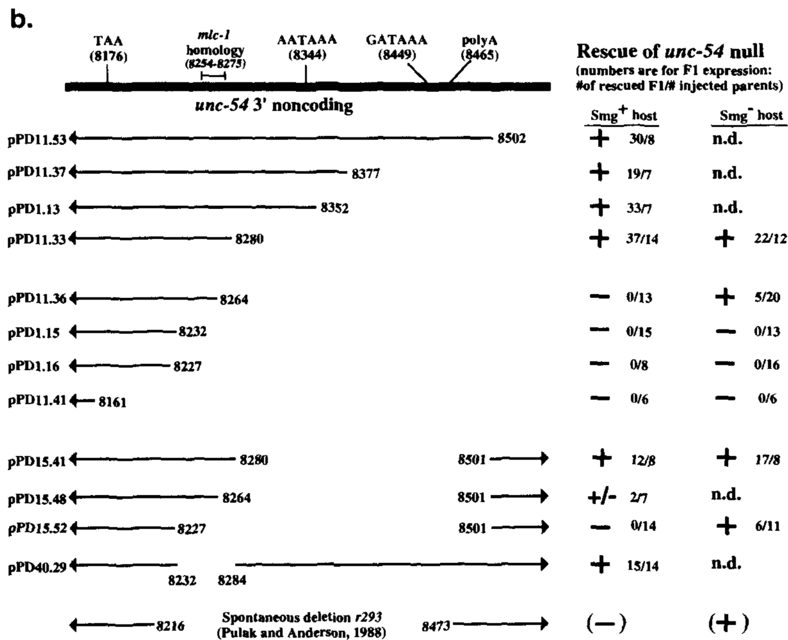
5'-Flanking sequences are not necessary for proper *unc-54* expression: We constructed a series of deletions removing *unc-54* 5'-flanking DNA (Figure 3a). Surprisingly, deletions replacing almost all DNA upstream of the ATG translation start could still rescue *unc-54(e190)* mutants. In particular, replacement of all sequences to 6 bp upstream of the translation start site resulted in a construction active in the mutant rescue assay (pPD10.98). As expected, deletions which extended beyond the ATG completely eliminated gene function.

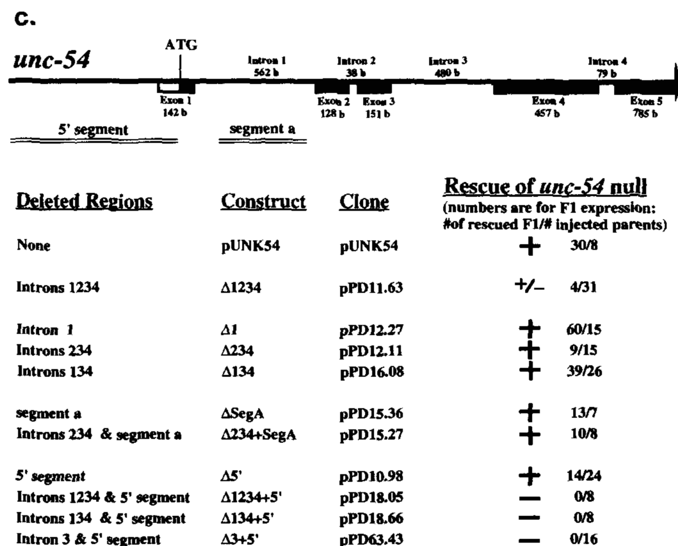
In several active deletions, the normal *unc-54* transcription start site had been replaced. Mapping of 5' ends for the resulting mRNAs revealed multiple start sites within the newly juxtaposed plasmid sequences. Cryptic starts were also observed in a deletion which retained the original transcriptional start site but retain only 42 bp of upstream sequence (data not shown).

We could assess the expression patterns of the reintroduced *unc-54* deletions both by analysis of rescued phenotypes and antibody staining. Each of the active 5' deletion constructs rescued both motility and egg laying defects seen in the *unc-54* mutant, reflecting expression both in body-wall striated muscles and in the vulval muscles. To directly assess tissue specificity, lines were stained with an *unc-54* specific monoclonal antibody, using an antibody to *myo-3* myosin as a control (MILLER, STOCKDALE and KARN 1983). The host strain *unc-54(e190)* shows no staining of the *unc-54* product, but exhibits abundant body muscle staining with the *myo-3* antibody (FIRE and WATERSTON 1989). In each of the rescued lines, *unc-54* staining was restored in essentially the same pattern as wild-type: *myo-3* and *unc-54* staining are co-localized to body-wall and vulval muscle, with no staining seen in pharynx or non-muscle tissue (Figure 3a).

These observations demonstrate that 5' sequences can be deleted from *unc-54* with no loss in tissue specificity. We did, however, observe a subtle quantitative effect in some weakly rescued lines transformed with the most proximal deletions (pPD7.32, pPD7.47 and pPD10.98; Figure 3a). In these lines the *unc-54* specific monoclonal antibody detected abundant expression in egg-laying muscles with somewhat lower







Deleted Regions	Construct	Clone	Rescue of unc-54 null (numbers are for F1 expression: #of rescued F1/# injected parents)	
None	pUNK54	pUNK54	+	30/8
Introns 1234	Δ1234	pPD11.63	+/–	4/31
Intron 1	Δ1	pPD12.27	+	60/15
Introns 234	Δ234	pPD12.11	+	9/15
Introns 134	Δ134	pPD16.08	+	39/26
segment a	ΔSegA	pPD15.36	+	13/7
Introns 234 & segment a	Δ234+SegA	pPD15.27	+	10/8
5' segment	Δ5'	pPD10.98	+	14/24
Introns 1234 & 5' segment	Δ1234+5'	pPD18.05	—	0/8
Introns 134 & 5' segment	Δ134+5'	pPD18.66	—	0/8
Intron 3 & 5' segment	Δ3+5'	pPD63.43	—	0/16

expression in body-wall muscles. Phenotypically, these lines exhibited wild-type egg-laying with only partial motility.

To test whether the *unc-54* expression obtained in the absence of 5′-flanking DNA resulted fortuitously from neighboring plasmid sequences, we replaced these flanking plasmid sequences with sequences upstream of the pharyngeal myosin gene *myo-2*. This construct also rescued *unc-54(e190)*, suggesting that expression in the 5′ deletions was not simply a result of the juxtaposed vector sequence. Lines rescued with the *myo-2::unc-54* chimera expressed *unc-54* myosin in both non-pharyngeal and pharyngeal muscles (data not shown). The *unc-54* myosin ectopically produced in the pharynx was present in radially oriented filaments, suggesting it could assemble in pharyngeal muscle cells. All of the lines transformed with this construct exhibited slow growth, which could conceivably result from the pharyngeal *unc-54* expression. These results indicate that ectopic expression of *unc-54* could be detectable using these antibodies, at least in pharyngeal muscle, and that body-wall and pharyngeal muscle regulatory sequences can independently function on the same gene.

Sequences within the *unc-54* 3′-UTR are necessary for efficient mutant rescue: The genetically isolated mutation *unc-54(r293)* removes 246 bp of the *unc-54* 3′-UTR and eliminates gene activity (PULAK and ANDERSON 1988). We made two sets of deletions to identify sequences within the *unc-54* 3′-UTR necessary for expression (Figure 3b).

Terminal deletions removing all 3′-flanking DNA, the polyadenylation site, and sequences beyond 102 bp downstream of the stop codon remained capable of rescuing *unc-54(e190)* (Figure 3b). However, rescue was eliminated by further deletions to 87 and 50 bp downstream of the stop codon (pPD11.36 and pPD1.15). These deletions affect a region of similarity between *unc-54* and the myosin light chain gene *mlc-1* (CUMMINS and ANDERSON 1988).

Internal deletions also implicated this conserved region in *unc-54* expression (*e.g.*, comparing pPD15.41, pPD15.48 and pPD15.52). Although some role for the *mlc-1* homology is evident, other downstream sequences can apparently function in an analogous manner, since a construct lacking most of the *mlc-1* homology region can be made active by providing sequences downstream (pPD40.29).

HODGKIN *et al.* (1989) have shown *unc-54(r293)* is suppressed by mutations in the *smg* genes which result

FIGURE 3.—Deletion analysis of *unc-54* function. (**a**) 5′ Deletion analysis. Each construct is a complete copy of the *unc-54* gene which has been deleted at the 5′ end, with nematode sequences replaced by sequences from the plasmid vector (*myo-2* sequences for construct pPD16.55). Only the 5′ region of each construct is diagrammed. Deletion endpoints are indicated by base numbers; arrows indicate that the construct described continues further into *unc-54* sequences. Numbering of deletion endpoints is from KARN, BRENNER and BARNETT (1983), with the parent construction [pUNK54; (FIRE and WATERSTON 1989)] extending from positions 279 to 9410. Expression in the F_1 reflects the ability of the clone to rescue the *unc-54* null mutation *e190* in the first generation following injection of oocytes (FIRE and WATERSTON 1989). For each construct, a number of *unc-54(e190)* parent animals were injected, each in 5–12 oocytes, and the resulting progeny scored for phenotypic rescue of movement and egg laying defects. In general about 30% of animals with rescued movement also were capable of laying eggs. Transgenic lines (much rarer than examples of F_1 expression) were obtained from these injections for a fraction of the positive plasmids. Transformed lines were analyzed by staining with a mixture of monoclonal antibodies to *unc-54* and *myo-3* products (MILLER, STOCKDALE and KARN 1983; FIRE and WATERSTON 1989). As a control for staining, all lines stained well with the antibody to *myo-3* product. "Tissue Specificity" describes the staining pattern observed with the antibody to *unc-54* product. "b.m." indicates a pattern of *unc-54* expression essentially the same as wild type: stain is readily observed and is limited to muscles of the body, including both body-wall striated and vulval muscles; no attempt was made to assess staining in the minor body muscles. "p.m." indicates additional staining in pharyngeal muscle. The most proximal deletions showed a quantitative difference from wild type in tissue distribution: lines derived from pPD7.32, pPD7.47 and pPD10.98 gave most intense stain in vulval muscles while wild-type animals show most intense stain in body-wall striated muscles. (**b**) 3′ Deletion analysis. Deletions at the 3′ end of the *unc-54* gene were assayed for rescue of *unc-54(e190)* as in **a**. Sequence features shown at the top are from KARN, BRENNER and BARNETT (1983) and this work. A homology to the *mlc-1* gene was noted by CUMMINS and ANDERSON (1988). For each construct, the number of rescued animals/the number of injected parents is shown. From these injections, heritable transformed lines have been obtained for pPD11.37, pPD1.13, pPD11.33 and pPD15.48. All were stained with the mixture of antibodies described in **a**, and each exhibits a normal pattern of *unc-54* expression. Deletion constructs were also tested for rescue of *unc-54* in a Smg⁻ mutant background. The strains used were *unc-54(e1092)smg-1(e1228)* and *unc-54(r259)smg-6(r896)*; comparable results were seen with these two strains. A single transgenic line, designated PD107, was obtained following injection of pPD11.36 into *unc-54(e1092)smg-1(e1228)*. The transgene rescued both movement and egg-laying defects. This rescuing activity was dependent on a Smg⁻ background: when a *smg-1+* allele was introduced by the appropriate genetic crosses, the PD107 transgene was unable to rescue either *unc-54* defect. Antibody staining was used to assess the pattern of *unc-54* expression by the PD107 transgene in the Smg⁻ background; in addition, we examined expression of *unc-54* product by the spontaneous deletion *unc-54(r293)* in three Smg⁻ backgrounds [*smg-1(r861)*, *smg-2(r863)* and *smg-6(r896)*]. In each case the observed expression pattern was indistinguishable from that of wild type. (**c**) Intron deletion analysis. Deletions of *unc-54* were assayed as described in **a**. "Segment a" (1109–1496) and 5′ segment (279–903) are deleted where noted; introns were deleted using the cDNA construct pMyo561 (MITCHELL *et al.* 1989). Each negative construct was confirmed by checking equivalent plasmids obtained independently. The precise intron deletions have been confirmed by DNA sequencing (MITCHELL *et al.* 1989). Plasmids pPD12.27 and pPD12.11 contain between them all pieces of the cDNA clone used for constructing pPD11.63. This rules out any functional defects in pPD11.63 resulting from the use of pMyo561 as a source of cDNA. Transgenic lines have been obtained for pPD12.11, pPD12.27, pPD15.27 and pPD11.63. The latter line exhibits only marginal rescue, while the other lines rescue to near wild-type phenotype. Antibody staining revealed that each of these lines expresses *unc-54* product in a wild-type pattern.

in stabilization of aberrant mRNAs. We found likewise that several of the inactive deletions within the *unc-54* 3′-UTR could rescue *unc-54(e190)* in a Smg⁻ genetic background (Figure 3b). In these suppressed animals, *unc-54* myosin was present in the normal distribution: only in body muscle cells. This indicates that the essential sequence element within the *unc-54* 3′-UTR, including the *mlc-1* homology, is not necessary for proper tissue-specific expression.

Two types of intron requirement: a non-sequence-specific stimulation by an intron and a specific activator element in *unc-54* intron 3: Two types of intron requirements have been described for optimal expression of eukaryotic genes. Nonspecific requirements for at least one intron within the transcribed sequences have been proposed (*e.g.*, CHOI *et al.* 1991). Specific enhancer elements can also be located within introns (*e.g.*, SCHULTZ *et al.* 1991). Given that we had observed proper regulation of *unc-54* in the absence of 5′-flanking sequence, we examined introns near the 5′ end of the gene for a role in gene expression. We have observed both a nonspecific requirement for splicing and a specific requirement for intron 3 for optimal expression of *unc-54* constructs.

The *unc-54* gene has eight introns (KARN, BRENNER and BARNETT 1983). Six of these are relatively short (38–79 bp) while two [intron 1 (562 bp) and intron 3 (480 bp)] are considered long for *C. elegans* introns (BLUMENTHAL and THOMAS 1988). We first deleted various introns in the presence of 5′-flanking DNA, using a synthetic cDNA clone (MITCHELL *et al.* 1989) kindly provided by J. KARN (Figure 3c). The precise deletion of the first four introns essentially eliminates mutant rescue (Δ1234). However, any one of the first four introns can be deleted without eliminating gene function (Figure 3c). In particular, constructs containing either intron 1 alone (Δ234) or intron 2 alone (Δ134) are functional. Furthermore, a construct removing 388 bp of internal sequences from intron 1 as well as introns 2, 3 and 4 (Δ234 + segment a) is sufficient for expression. These data suggested a requirement for splicing rather than specific sequences within intron 1.

Although Δ134, Δ234 and Δ234 + segment a are active, they exhibit a subtle but reproducible reduction in their ability to rescue. This reduction is observed as a decreased frequency and quality of rescue, particularly the ability to rescue the egg-laying defect (data not shown). These observations suggested the possibility of a special contribution by the third or fourth intron to full activity.

In the absence of 5′-flanking DNA, a construct deleting introns 1, 3 and 4 (Δ134 and 5′ segment) is unable to rescue *unc-54(e190)*. Because a similar construct containing 5′ sequences is functional, introns 1, 3 or 4 must contain regulatory information redundant with that in the 5′-flanking DNA. The critical element is apparently in intron 3: a deletion removing only intron 3 and 5′-flanking sequence is inactive. These results suggested the working model that two different sequence elements in *unc-54*, one upstream of the transcription start site and the other in intron 3, act as independent positive activators allowing gene expression in body-wall muscles.

***lacZ* fusion assays:** Assays which require the function of the myosin gene greatly constrain the analysis of regulatory sequences internal to the gene. In addition, these assays can only be done in a mutant background. For these reasons, we used myosin::*lacZ* fusion constructions (Figure 4a) to further define sequences responsible for tissue-specific regulation. We used vectors incorporating a nuclear localization peptide at the N terminus of β-gal; this leads to predominant staining in the nuclei of expressing cells, facilitating cell identification (FIRE, HARRISON and DIXON 1990). A 3′-non-coding sequence is placed downstream of *lacZ*. Initial *myo::lacZ* fusions (not shown) containing the SV40-early 3′ region expressed very inefficiently in *C. elegans*. This inefficiency and the results with 3′ deletions described above in Figure 3b led us to make and utilize vectors which incorporate the *unc-54* 3′ region (FIRE, HARRISON and DIXON 1990). Because splicing also appeared necessary for efficient *unc-54* expression, we included a synthetic intron sequence based on consensus splice sites (EMMONS 1988) in many of our constructs. This intron has no homology to the myosin genes outside of the *C. elegans* splice consensus. These vectors (carrying the *unc-54* 3′ end and synthetic intron) can express β-gal in nearly all *C. elegans* tissues (*e.g.*, FIRE, HARRISON and DIXON 1990) and have no apparent bias toward expression in muscle.

Myosin gene::*lacZ* fusions are properly expressed: We constructed both translational and transcriptional fusions for each of the four myosin genes (Figure 4). The *unc-54::lacZ* and *myo-3::lacZ* fusions were specifically expressed in body muscle (Figure 5, A and B). This expression mimicked expression of the parent genes both in the abundant striated body-wall muscles used for locomotion, and in several sets of specialized "minor" body muscles: intestine-associated, vulva-associated, and uterine sheath muscle (not shown). The *myo-1::lacZ* and *myo-2::lacZ* fusions likewise mimicked expression of their endogenous counterparts, expressing only in pharyngeal muscle (Figure 5, C, D and E).

In several subsequent experiments, we have used modified *lacZ* vectors as follows to confirm that the observed expression patterns reflect promoter activity and not specific features of the vectors (not shown). Elimination of the nuclear localization signal results in constructions expressed cytoplasmically in the same

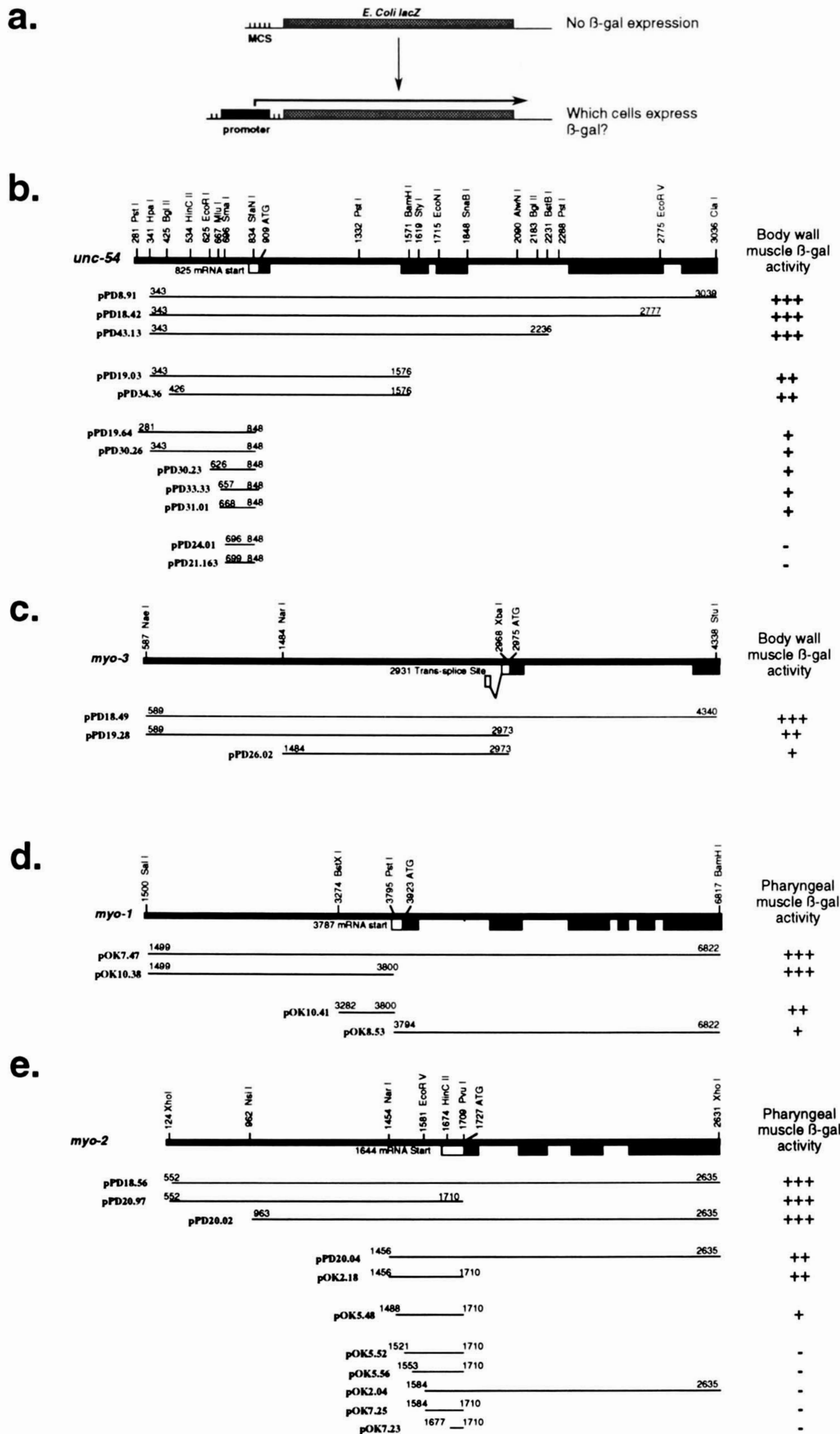


FIGURE 4.—Promoter::*lacZ* fusions. (**a**) General strategy for assaying promoters using fusions to *lacZ*. A DNA fragment is cloned into a multiple cloning site (MCS) upstream of *lacZ* (FIRE, WHITE HARRISON and DIXON 1990) creating either a transcriptional fusion within the 5′-UTR or an in frame fusion within the protein coding sequence. Expression from constructs transformed into worms is assayed by *in situ* staining for *β*-galactosidase. Segments of *unc-54* (**b**), *myo-3* (**c**), *myo-1* (**d**) and *myo-2* (**e**) tested for promoter activity. A partial restriction map, transcription start or trans-splice sites, and intron/exon structure are indicated for each gene. Lines with clone names indicated to the left indicate the precise extent of fragments tested for promoter activity (numbered according to KARN, BRENNER and BARNETT 1983; DIBB *et al.* 1989). Fragments were generally manipulated using the restriction enzymes noted above the fragment endpoints. In some cases nuclease Bal31 exonuclease or PCR were used to generate fragments. Activity levels are represented as − (no observed activity), and +, ++, +++ (increasing levels of staining; the scale is arbitrary).

cells. Elimination of synthetic intron sequences results in reduced expression in some cases (see below), but no change in tissue specificity. Replacement of *unc-54* 3′ sequences downstream of *lacZ* with those from a non-muscle gene (*glp-1*) causes no change in expression pattern. Replacement of *lacZ* sequences with those of the *Escherichia coli* gene *uidA* results in expression of the *uidA* product [*β*-glucuronidase (JEFFERSON

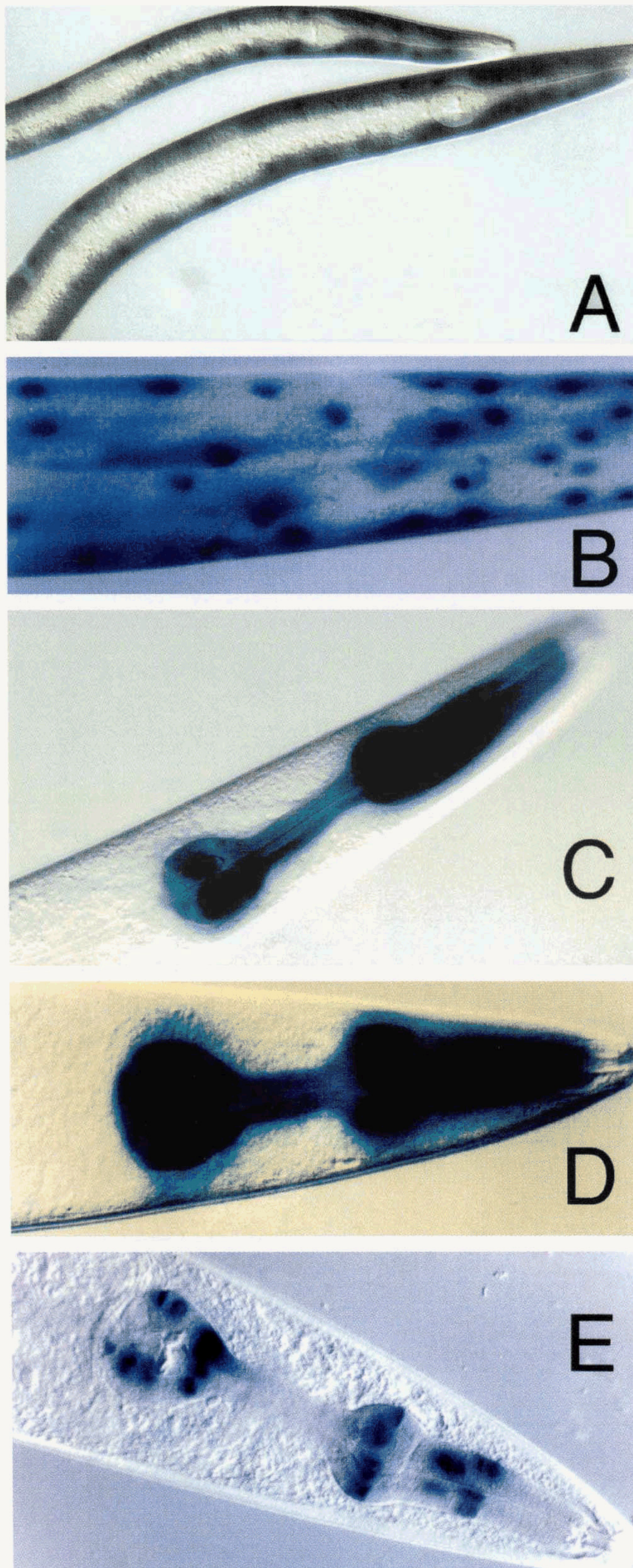


FIGURE 5.—Expression of myosin::*lacZ* fusions. Expression patterns of *lacZ* fusions of the *C. elegans* myosin heavy chain genes. Animals shown are histochemically stained for X-gal (FIRE 1993). Cell identifications were confirmed by co-staining (not shown) with 4,6-diamidino-2-phenylindole (DAPI) (to label nuclei) and rhodamine phalloidine (which primarily labels muscle cells). (**A**) *unc-54* translational fusion; animal from transformed line PD56, which has a single integrated copy of the *unc-54::lacZ* fusion construction pPD18.42. The stained nuclei are from body-wall muscles. Cytoplasmic staining is seen with longer incubation. Similar staining is seen in F_1 animals following injection of pPD18.42. In each of the other cases the staining shown is of F_1 animals following injection of the described fusion plasmids. Although each stained cell was of the expected muscle class, not all cells of a given type were stained. This presumably results in part from the mosaicism due to segregation of the injected DNA. Similar results were obtained with injection into oocytes (FIRE and WATERSTON 1989) or into the syncytial germline (MELLO *et al.* 1991). (**B**) *myo-3* translational fusion, pPD18.49. Staining in body-wall muscle (primarily nuclear). (**C**) *myo-1* translational fusion, pOK7.47. Staining in pharyngeal muscles. (**D**) *myo-2* translational fusion, pPD18.56. Staining of pharyngeal muscles (nucleus + cytoplasm). (**E**) *myo-2* transcriptional fusion, pPD20.97. Staining of pharyngeal muscle nuclei. Similar transcriptional fusions to *myo-1* show an identical pattern.

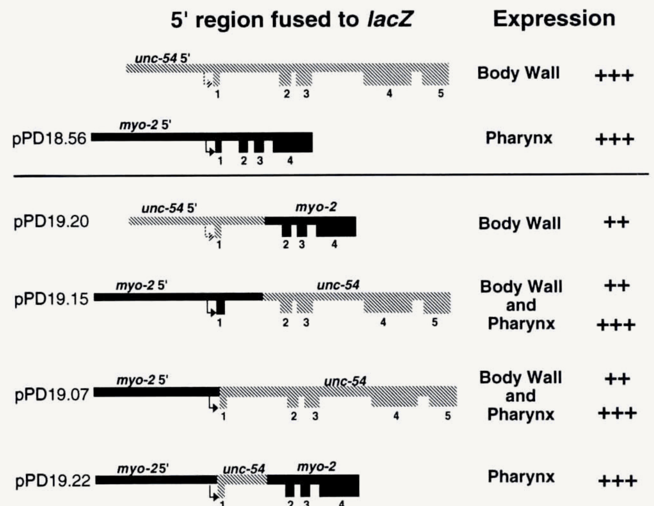


FIGURE 6.—Chimeric *lacZ* fusions. At the top are diagrammed the initial *unc-54::lacZ* and *myo-2::lacZ* translational fusions, pPD18.42 and pPD18.56. The activities of these are described in Figures 4 and 5. In this schematic, high bars represent coding exons; lower regions are introns and 5′-flanking sequence. Unique *Hind*III sites in the first intron of each gene (1382 in *unc-54* and 1832 in *myo-2*) were used in making chimeras pPD19.20 and pPD19.15. The *myo-2::unc-54* fusion junction in pPD19.07 was made by joining a *myo-2 Pvu*I end (1709 in *myo-2*) to a Bal-31-created *unc-54* terminus (874 in *unc-54*). pPD19.22 caries both junctions (*Hind*III and *Pvu*I*::Bal-31*). Each construct was assayed in the F_1 following injection as in Figure 4. The expression patterns observed are described at the right.

et al. 1987)] in an identical pattern to the initial *lacZ* fusion.

Sequence requirements for promoter activity: To delineate upstream sequences necessary for expression in each muscle type, we assayed promoter activities for a set of deletions based on each myosin::*lacZ* transcriptional fusion (Figure 4, b, c, d and e). For *unc-54* and *myo-2*, minimal active promoter regions of 180 and 222 bp, respectively, were defined by this analysis [*unc-54*: Figure 4b (pPD31.01); *myo-2*: Figure 4E (POK5.48)]. Less detailed deletion analysis of the *myo-1* and *myo-3* genes identified functional promoter regions of 518 and 1489 bp, respectively [*myo-1*: Fig-

ure 4D (pOK10.41); *myo-3*: Figure 4C (pPD26.02)]. In each case, the minimal active promoter fusions are less efficient than fusions including more extensive 5'-flanking DNA. The differences in activity suggest that sequence elements upstream of each minimal promoter augment transcription. This assay does not, however, address whether these activating sequences play a role in tissue specificity.

Characterization of regulatory sequence requirements within the genes is less straightforward using simple deletions. Two constructions with different extents of internal sequence will produce mRNA and protein products with different structures, and hence with potentially different activity and stability. Each of the four translational fusions was somewhat more active than similar transcriptional fusions (Figure 4 and data not shown). These differences could thus reflect either a higher transcriptional activity, more efficient translation or enzyme function, or increased stability conferred by the bona fide N-terminal and 5'-untranslated sequences.

Chimeric *lacZ* fusions identify determinants of tissue specificity: Several chimeric *lacZ* fusions containing different parts from *unc-54* and *myo-2* were constructed (Figure 6). By assaying for the tissue localization of expression (*i.e.*, body-wall *versus* pharyngeal) from these constructs, we hoped to identify the sequences responsible for the observed difference in tissue specificity. In particular we wanted to test whether sequences inside the genes played a role in tissue specificity.

The most informative chimeric fusions were pPD19.07 and pPD19.15. Results with these constructs delineate a sequence element within the *unc-54* transcribed region capable of activating the *myo-2* promoter in body muscle. These experiments taken with results above demonstrating the activity of *unc-54::lacZ* transcriptional fusions containing only upstream *unc-54* sequence [*e.g.*, Figure 4b (19.64)], show that two different elements, one upstream and the other downstream of the *unc-54* initiation site, are independently capable of directing gene expression to body muscles. For *myo-2*, no indication for an internal control element was found, although it should be noted that sequences downstream of the *Xho*I site in exon 4 of *myo-2* have not been examined.

Body-muscle enhancer sequences: The chimera experiments above suggested a sequence within the *unc-54* gene might act as a tissue specific enhancer to induce body muscle expression from the *myo-2* promoter region. To test this, and to further characterize such elements, we cloned fragments of *unc-54* upstream of the *myo-2::lacZ* translational fusion pPD18.56 (Figure 7a). Normally this construct is expressed very specifically in the pharynx; hence any element that can induce the *myo-2* promoter to express outside the pharynx could be detected.

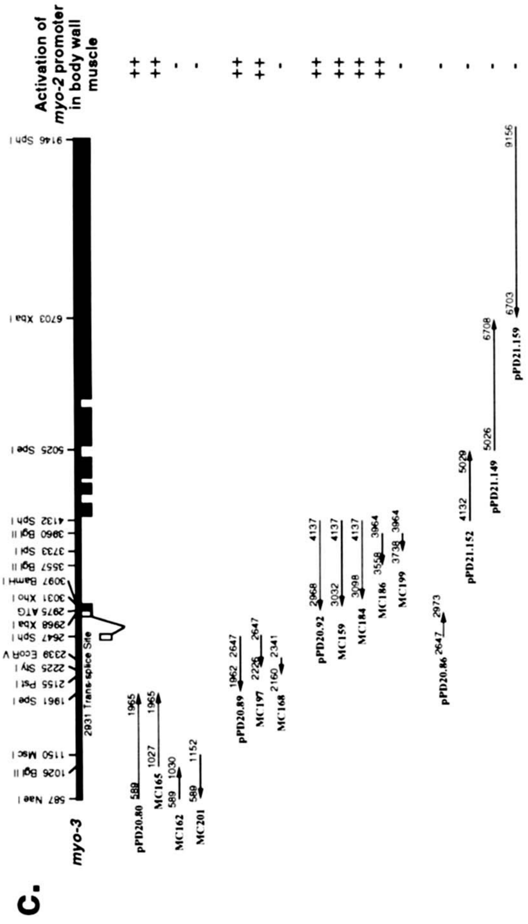
We initially found a 962-bp restriction fragment within the *unc-54* gene that induced body-wall expression when placed in either orientation upstream of *myo-2::lacZ* [Figure 7b (pPD19.84, pPD19.93); Figure 8, A and B)]. Deletion analysis of this fragment revealed that enhancer activity is contained within the third intron of the *unc-54* gene (Figure 7b). Sequences covering the second intron of *unc-54* exhibit no enhancer activity. The ability of the element within *unc-54* intron 3 to induce body wall-muscle expression (Figures 6 and 7b) in different positions and orientations proximal to the *myo-2* promoter indicated this element was indeed a tissue-specific transcriptional enhancer.

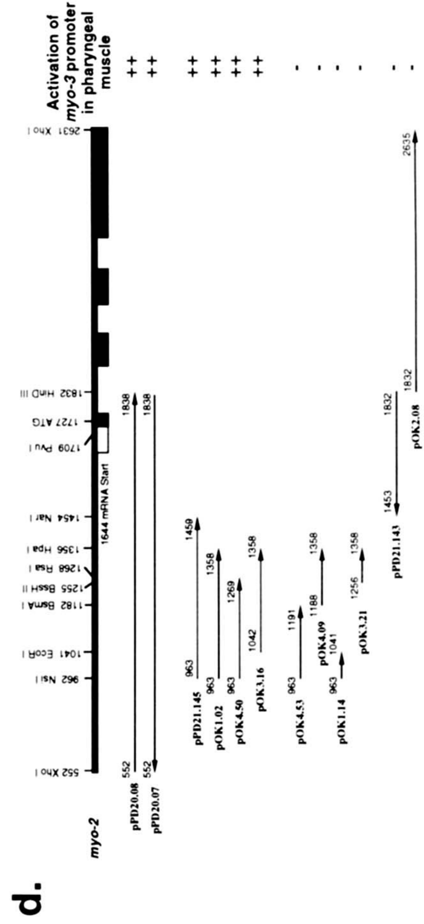
Other regions from *unc-54* and segments from *myo-3* were tested similarly for enhancer function. A fragment spanning the *unc-54* promoter exhibits weak body-wall muscle enhancer activity in both orientations [Figure 7b (pPD19.72, pPD19.76)]. In addition, three enhancer sequences around the *myo-3* gene were identified. Two of the *myo-3* enhancer segments lie upstream of the gene [Figure 7c (MC165, MC197)], while the third is within the first intron [Figure 7c (MC186)].

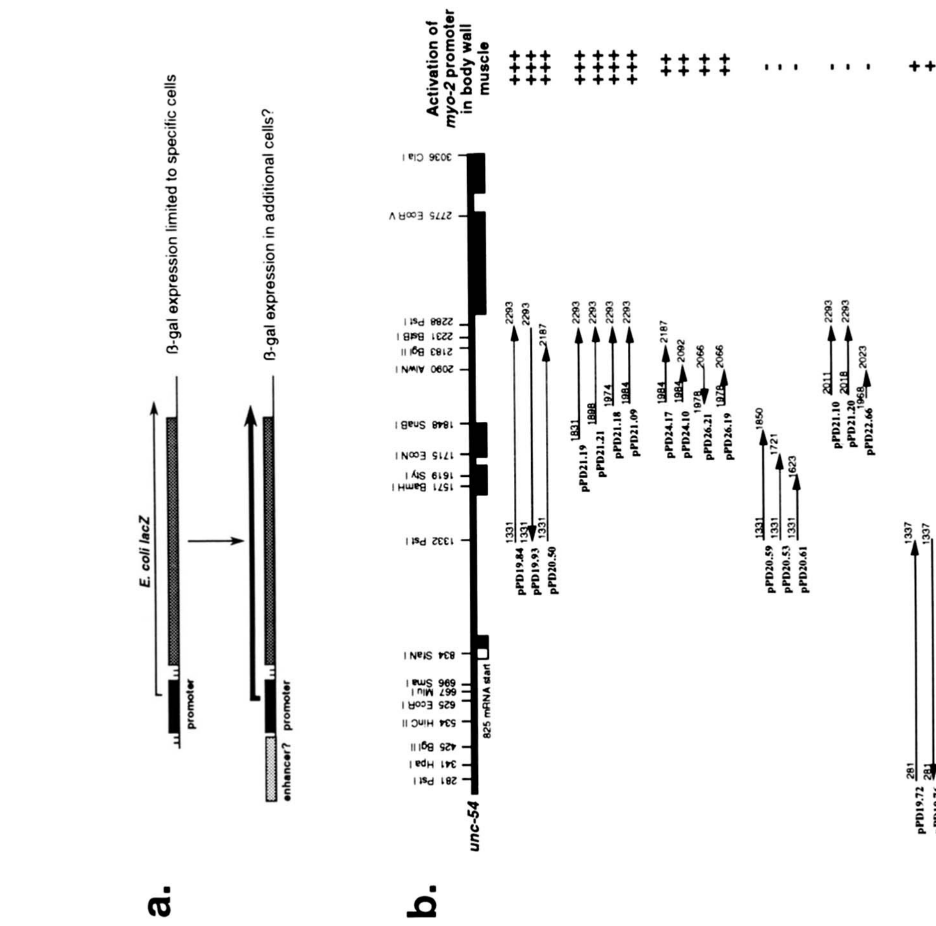
Pharyngeal muscle enhancer sequences: In experiments analogous to those identifying body muscle-enhancer activities, we looked for pharyngeal-specific enhancer activities in *myo-2* by inserting segments upstream of a *myo-3::lacZ* fusion construct. We identified a single enhancer approximately 0.4 kb upstream of the *myo-2* transcription start site (Figure 7d). Two overlapping 0.3-kb fragments within the region are sufficient for enhancer activity [Figure 7d (pOK4.50, pOK3.16)].

myo-1 contains at least two enhancers: a weak distal enhancer is located at least 0.5 kb upstream of the transcription start and a strong proximal enhancer is located nearer the start site [Figure 7e (pOK8.66, pOK9.05)]. Terminal deletions of the promoter proximal enhancer identify a 375-bp fragment sufficient for enhancer activity (Figure 7e (pOK9.05)]. A third tissue-specific element is apparently located downstream of the transcription start in *myo-1*, since a fusion starting at 8 nucleotides downstream of the start is active in pharyngeal muscle [Figure 4d (pOK8.53)]. We were not able to identify an enhancer corresponding to this activity, suggesting that activity of the corresponding activator element(s) may be dependent on positioning relative to the promoter.

Tissue specificity of myosin enhancers: The patterns of expression from specific enhancer-promoter combinations such as those just described presumably reflect the underlying specificity of the promoter as well as the enhancer. We have therefore tested en-


c.
myo-3
Activation of
myo-2 promoter
in body wall
muscle
2931 *trans*-splice Site
pPD20.80
MC165
MC162
MC201
pPD20.89
MC197
MC168
pPD20.92
MC159
MC184
MC186
MC199
pPD20.86
pPD21.152
pPD21.149
pPD21.159


d.
myo-2
Activation of
myo-3 promoter
in pharyngeal
muscle
1644 mRNA Start
pPD20.08
pPD20.07
pPD21.145
pOK1.02
pOK4.50
pOK3.16
pOK4.53
pOK4.09
pOK1.14
pOK3.21
pPD21.143
pOK2.08


a.
E. coli lacZ
promoter
β-gal expression limited to specific cells
enhancer?
β-gal expression in additional cells?
b.
unc-54
825 mRNA start
Activation of
myo-2 promoter
in body wall
muscle
pPD19.84
pPD19.93
pPD20.50
pPD21.19
pPD21.21
pPD21.18
pPD21.09
pPD24.17
pPD24.10
pPD26.21
pPD26.19
pPD20.59
pPD20.53
pPD20.61
pPD21.10
pPD21.20
pPD22.66
pPD19.72
pPD19.76

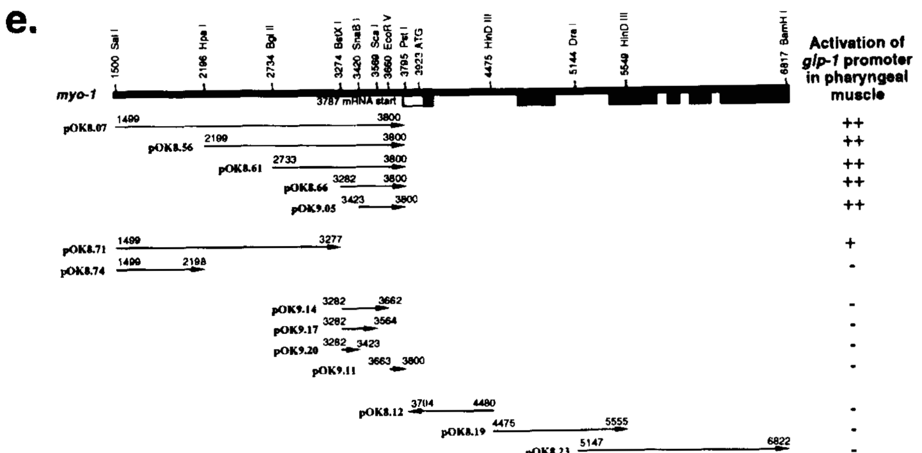


FIGURE 7.—Enhancer assays. (**a**) General strategy for identifying enhancers. DNA fragments are assayed for the ability to activate a promoter::*lacZ* fusion, normally active only in certain cells, in additional cell types. *unc-54* and *myo-3* enhancers were identified by their ability to induce body-wall muscle expression from a normally pharyngeal *myo-2::lacZ* fusion (*myo-3* and *unc-54*). All constructs to assay body muscle enhancement were derived from the *myo-2::lacZ* translational fusion pPD18.56 (Figure 4d), except for pPD24.17 and pPD24.10, which were derived from the *myo-2::lacZ* transcriptional fusion pPD20.97. The *myo-2* enhancer was identified by its ability to induce pharyngeal muscle expression from a normally body wall *myo-3::lacZ* fusion (pPD20.08, pPD20.07 and pOK2.08 were assayed upstream of fusion pPD19.28 (Figure 4c); the remainder were assayed upstream of fusion pPD26.02 (Figure 4c). *myo-1* enhancers were identified by their ability to induce pharyngeal expression from the *glp-1::lacZ* fusion pPD26.50 (see text). Panels **b–e** show segments of *unc-54* (**b**), *myo-3* (**c**), *myo-2* (**d**) and *myo-1* (**e**) tested for enhancer activity. Features of the genes are diagrammed as in Figure 4. Construct names are shown to the left of fragments tested for enhancer activity. Right arrows indicate fragment tested in a plus orientation relative to the promoter::*lacZ* fusion; left arrows indicate constructs in which the putative enhancer fragment was tested in inverted orientation. Activity levels are represented as – (no observed activity), and +, ++, +++ (increasing levels of staining; the scale is arbitrary).

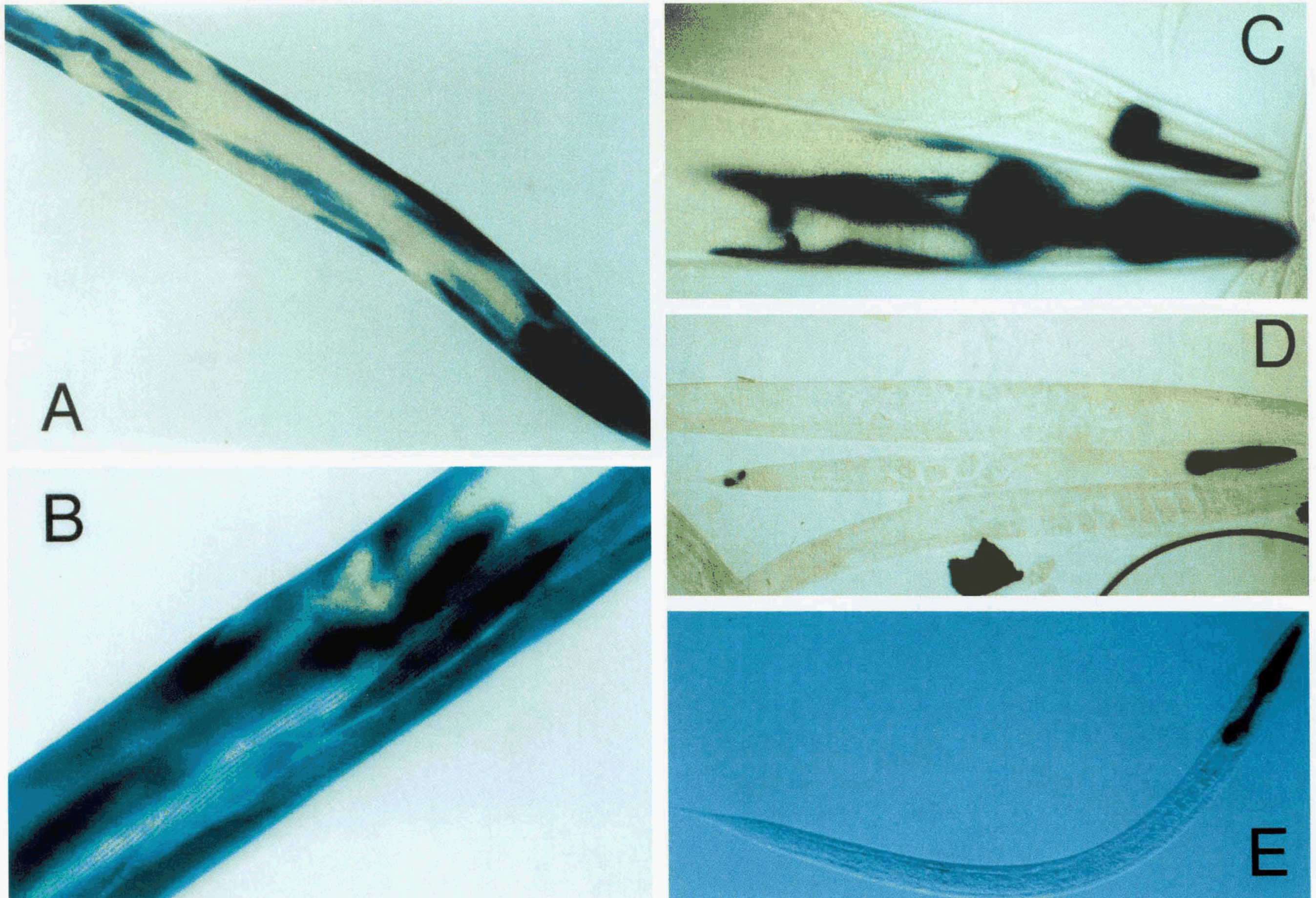


FIGURE 8.—Expression of enhancer assay constructs. Staining is shown of F_1 animals ("transient transformants") following injection of each plasmid. (**A**--**D**) These enhancer assay constructions were derived from *myo-2::lacZ* translational fusion 18.56 (normally expressed in the pharynx only: see Figure 5D). Note that the nuclear localization signal is not completely functional in the fusion protein produced so that staining was often cytoplasmic. The staining follows the muscle filament pattern, but it is not clear whether this represents the distribution of the fusion protein or precipitation of the stain. (**A** and **B**) Expression following injection of plasmid pPD19.84 (Figure 7B). This is a derivative of the *myo-2::lacZ* fusion with a 962-bp *Pst*I fragment internal to *unc-54* inserted upstream of *myo-2* sequences. (**C** and **D**) Expression from plasmids pPD21.61 (**C**) and pPD21.63 (**D**) (insertions of random 5-kb *Sal*I-*Sph*I fragments from the *C. elegans* genome upstream of the *myo-2* segment in pPD18.56). Note the intense staining of two body-wall muscle cells in **C**. The extensions from the two cells presumably represent "muscle arms" (processes sent out by nematode muscle cells to synapse to the ventral and dorsal nerve cords (WHITE 1988). Tail cells expressing β-galactosidase after injection of pPD21.63 (*e.g.*, **D**) were identified as the intestine-associated anal depressor and sphincter muscles. Expression of this construct in other body muscle cells has also been observed (not shown). (**E**) Expression of pOK4.81, which contains the *myo-2* enhancer (bases 963–1358) upstream of *glp-1::lacZ* fusion pPD26.50.

hancers from the *unc-54*, *myo-1* and *myo-2* genes for their patterns of activation with a promoter having no preference for muscle. *glp-1* is a gene involved in signal transduction during embryonic and larval development (AUSTIN and KIMBLE 1987; PRIESS, SCHNABEL and SCHNABEL 1987; YOCHEM, WESTON and GREENWALD 1988; LAMBIE and KIMBLE 1991). A *glp-1::lacZ* fusion in the first exon of *glp-1* (pPD26.50) gives rare expression in posterior gut cells, posterior hypodermis and occasionally cells of other types (not shown). The enhancer regions for *myo-1*, *myo-2* and *unc-54* (positions 3423–3800, 963–1358 and 1332–2293, respectively) were placed upstream of the *glp-1* promoter in pPD26.50. The resulting constructs showed expression in pharyngeal muscle for the *myo-1* and *myo-2* enhancers and body wall muscle for the *unc-54* enhancer (Figure 7e and data not shown). In experiments with other enhancer segments from a variety of other genes upstream of the *glp-1* promoter (V. PLUNGER, W. WADSWORTH, P. OKKEMA, A. FIRE, S. XU and S. HARRISON, unpublished), we found that the *glp-1::lacZ* could be enhanced in a variety of non-muscle tissues including body hypodermis, neurons, glial cells, gut, and non-muscle pharyngeal tissue. These experiments strongly suggest that the *unc-54*, *myo-1* and *myo-2* enhancer segments carry complete information sufficient to specify activity in just the appropriate muscle cells.

Screening the genome for enhancer-like sequences: We have begun to screen random segments of the nematode genome for enhancer activities using assays similar to those described above. Although our

Intron Segment	**Promoter Segment**				
	unc-54 pro (279-848)	***unc-54 enh+pro*** (1832-2293 & 279-848)	***myo-2*** (125-1710)	***myo-3*** (589-2973)	***myo-3*** (1486-2973)
	Average number of stained F1 cells per injected adult (numbers in parentheses are total adults injected)				
A. Synthetic Intron I	**29 (8)**	**99 (8)**	**>90 (8)**	**108 (8)**	**13 (25)**
B. Synthetic Intron II	**19 (16)**		**84 (15)**		
C. Precise Deletion	**0 (11)**		**16 (15)**	**68 (8)**	**9 (24)**
D. Synthetic Intron II*	**0 (24)**		**5 (16)**		
E. PpuMI deletion	**0 (10)**	**5 (8)**	**3 (8)**		

A. Synthetic Intron I	AGGACCCAAAGgtatgtttcgaatgatactaacataacatagaacattttcagGAGGACCC
B. Synthetic Intron II	AGGACCCAAAGgta**a**gtttcgaat**C**atactaacataacatagaacattttcagGAGGACCC
C. Precise Deletion	AGGACCCAAAG--GAGGACCC
D. Synthetic Intron II*	AGGACCCAAAG**A**TA**A**GTTTCGAAT**C**ATACTAACATAACATAGAACATTTTC**G**GGAGGACCC
E. PpuMI deletion	AGGA---CCC

FIGURE 9.—Stimulatory activities of synthetic introns. (**Top**, **A–E**) Five different myosin::*lacZ* transcriptional fusions have been tested with a variety of intron-like and non-intronic sequences in their 5′-nontranslated regions. In each case the number of stained cells were determined in the F_1 following injection of oocyte nuclei in 8–24 parental animals. The number of injected animals is shown in parenthesis, with the average number of stained muscle cells per injected parent represented for each combination tested. Because of some variability in injections, differences of less than twofold are not considered significant. As an alternate measure of expression, we also observed the intensity of staining: removal of introns from *myo-2* and *unc-54* fusions resulted in much fainter staining in those cells that were stained, while no difference in intensity was observed between the corresponding *myo-3* fusions. (**Bottom**, **A–E**) Sequences of equivalent regions in the 5′-UTR. In the constructs described, intronic sequences or their residues were present between the start of transcription and the start of translation. In each case, the transcriptional start of the myosin gene was retained in the constructs, while translation would be expected to start at the ATG in the nuclear localization cassette attached just upstream of *lacZ* (see Figure 4). Synthetic Intron I is the original intron sequence of FIRE and WATERSTON (1989). For comparison of intron-like sequences with and without functional splice junctions, it was necessary to start with an alternate synthetic intron (Synthetic intron II) which lacks ATG sequences, as these could potentially interfere with translation of the *lacZ* open reading frame located downstream (KOZAK 1991). Uppercase: sequences expected to be retained in the processed mRNA. Lowercase: sequences expected to be spliced out. Bases in outline represent changes from synthetic intron I. The promoter segments used from *unc-54*, *myo-2* and *myo-3* correspond to segments described in Figure 4. The corresponding intron I containing constructs for *unc-54* and *myo-2* were pPD19.64 and pPD20.97, respectively; the long and short *myo-3* promoter constructs with intron I were pPD19.28 and pPD26.02, respectively (Figure 4). The *unc-54* enhancer+promoter constructs contain the indicated intron 3 sequence in a '+' orientation placed upstream of the *unc-54* promoter in construct pPD19.64. All of these constructs were identical outside of the synthetic intron and promoter sequences.

initial screens were designed to survey only a small fraction of the genome, these experiments could yield several types of information relevant to the enhancer and promoter activities described above. First, we should obtain a very rough estimate for the frequency in the genome of enhancer activities as defined by these assays. Second, the different patterns of expression observed with a given promoter should illuminate the underlying specificity of the promoter used, in terms of its ability to respond to enhancement in different tissues. In their normal context, the myosin gene promoter elements need only respond to signals in the muscle types where the corresponding genes are expressed. The data in the preceding sections suggests strongly that these promoter elements can respond to signals for expression in the other muscle types, if such signals are added in *cis*.

To assay response to enhancement in a broad set of tissues, we have carried out a limited screen of genomic DNA for sequences capable of enhancing the *myo-2* promoter. The *myo-2* promoter was chosen since expression in any tissue outside the pharynx would be readily detected in enhancer assays. Random fragments (2–5 kb) of *C. elegans* DNA were inserted upstream of the active *myo-2::lacZ* translational fusion construct used in the assays described above. The resulting constructs were injected individually, and stained animals from the F1 generation were examined for non-pharyngeal β-gal expression (Figure 8, C and D). Six of eighteen fragments tested functioned as enhancers in this assay (approximately 62 kb of genomic DNA was examined). All of the observed expression from these constructs occurs in muscle cells (Figure 8). One explanation for this would be that the *myo-2* promoter segment used is selective in its response to enhancer function, responding predominantly in muscle tissue. Data consistent with this hypothesis come from assays of two of the enhancers identified in this screen for enhancement of the *glp-1* promoter segment. One of these fragments enhances

the *glp-1* promoter in non-muscle cells (in the pharyngeal-intestinal valve and a variety of other cells), while a second weakly enhances *glp-1* in muscle (data not shown). The first fragment thus appears to be capable of enhancement in both muscle and non-muscle tissues, however the *myo-2* promoter segment responds only in muscles.

Intron requirements revisited: To address the nature of requirements for intron sequences, we compared transcriptional fusions of the myosin genes to *lacZ* with and without the synthetic intron present. To control for potential effects of 5′ leader sequences on gene expression which are independent of a functional intron, we have generated three sets of constructs to compare intron effects. In the first set, the synthetic intron is deleted by excision of a *Ppu*MI restriction fragment, which creates a 12-bp deletion in the resulting mRNA. In the second, the intron is precisely deleted, so the primary transcript would be identical to the spliced message from the intron-containing construct. The third set of constructs contains a synthetic intron which has been mutated at the splice junctions, so that the transcripts are otherwise identical.

For *myo-2* and *unc-54*, similar results were obtained for the three sets of constructs: the lack of an intron resulted in decreased expression as assayed both by the number of stained cells and the intensity of the staining (Figure 9). For *myo-3*, no significant difference in expression was observed between intron-containing and intron-lacking constructs. *myo-3* is the only one of the four genes to be trans-spliced at the 5′ end. These experiments initially suggested that high level intron-independent expression might correlate with trans-splicing. However, preliminary results with two other genes suggest that the actual situation could be more complex: for *hlh-1*,which is trans-spliced to SL1 (KRAUSE *et al.* 1990), we observed strong stimulation of a transcriptional fusion following addition of synthetic intron I; for *myo-1*, which is not trans-spliced, we observed no difference between a transcriptional fusion containing synthetic intron I and an equivalent intron-lacking construct (data not shown).

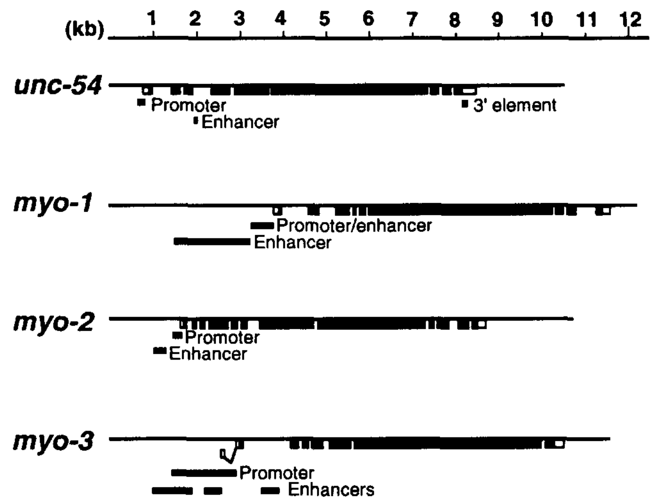


FIGURE 10.—Summary of promoter/enhancer assay results. These maps show the current delineation of functionally important regions for the four myosin heavy chain genes studied. The **Enhancer** sequences are the minimum sequences which enhance in the assays described in Figure 7. **Promoter** segments are likewise the smallest regions tested which show expression when fused to *lacZ*. Bars indicate the minimal extent of each element as defined by our deletion analysis.

DISCUSSION

We have characterized sequences required for myosin gene expression in transgenic *C. elegans*. These experiments have identified regulatory elements controlling tissue-specific expression of the myosin genes as well as general features of one of these genes which are necessary for activity. In discussing the results we will describe first the individual elements and then describe their interactions in the context of a complete functional gene.

Promoter regions are sufficient for muscle type-specific expression: For each of the myosin genes, a segment upstream of the coding region is sufficient to drive expression of the *E. coli lacZ* coding region in the appropriate muscle type (summarized in Figure 10). Minimal segments sufficient for expression have been identified in *unc-54* and *myo-2* (Figure 4): a fragment containing 168 bp of *unc-54* 5′-flanking DNA can drive *lacZ* expression specifically in body muscle, while a fragment containing 156 bp of *myo-2* 5′-flanking DNA can drive *lacZ* specifically in pharyngeal muscle. While these promoters are sufficient for tissue-specific expression, additional sequences upstream and/or downstream of the promoters are necessary for full transcriptional activity.

An underlying second level of promoter specificity was evident in experiments with the *myo-2* promoter. A given promoter might be expected to have a range of tissues in which it could respond to enhancement. A *myo-2* promoter segment has been analyzed by testing a small number of random genomic fragments from *C. elegans* placed upstream for enhancer function. The observed activation in this experiment was limited to muscle cells. Some of the active fragments were capable of enhancing in non-muscle tissues, as evidenced when a promoter with no known muscle preference (*glp-1*) was used to assay for enhancer function. This suggests that *myo-2* promoter segment is poised to express and sensitive to nearby enhancers in all types of muscle, but that the same promoter segment is relatively insensitive to enhancement in non-muscle tissue.

Tissue-specific enhancers augment expression: Muscle type-specific enhancers, distinct from the minimal promoters defined by simple deletion analysis,

were identified for each of the myosin genes (summarized in Figure 10). The enhancers from the body wall myosin gene *unc-54* and the pharyngeal muscle myosin gene *myo-2* have been examined in greatest detail. The major *unc-54* enhancer is located within the third intron and activates transcription in body-wall muscle and other non-pharyngeal muscles. The *myo-2* enhancer is located upstream of the promoter and activates transcription specifically in pharyngeal muscle. The *unc-54* and *myo-2* enhancers activate transcription in the appropriate muscle type when assayed upstream of both muscle and non-muscle promoters, indicating that tissue specificity is an inherent property of the enhancer and not dependent on a particular promoter.

Two types of intron contribution: intron resident enhancers and mechanistic effects of splicing: Two types of requirements for intervening sequences were found in this analysis. The first are requirements for specific sequences inside of introns. These sequences (in the first intron of *myo-3* and the third intron of *unc-54*) act as enhancers. The presence of relatively large introns near the 5′ ends of these genes provides a venue to place activating enhancers downstream as well as upstream of the promoter; this added flexibility could facilitate the evolution of highly regulated genes.

A second type of intron effect apparently reflects a mechanistic need in some circumstances for an mRNA splicing event. A stimulatory effect of splicing is particularly evident in comparing the activity of *unc-54::lacZ* fusions which have been constructed with and without synthetic intron sequences (we used synthetic introns that have no homology outside of the consensus splice sites to any of the myosin gene introns). A similar stimulation was seen with *myo-2::lacZ* fusions, but analogous experiments with the *myo-1* and *myo-3* promoters failed to detect any increased activity in the presence of the synthetic intron. Previous studies with mammalian tissue culture cells have similarly suggested that splicing-dependent stimulation is a property of the promoter (*e.g.*, NEUBERGER and WILLIAMS 1988).

Although the synthetic introns dramatically stimulated expression of some gene fusions, no changes in tissue specificity were seen. There is thus no indication of a tissue specificity related to the mechanistic requirement for splicing.

***unc-54* 3′ sequences are required for expression but not tissue specificity:** A requirement for sequences at the 3′ end of *unc-54* was evident from both whole gene assays (Figure 3b) and *lacZ* fusions (not shown). A short region of homology to the gene *mlc-1* plays some role, although other downstream sequences can apparently substitute.

The 3′-untranslated sequences required for *unc-54* expression do not appear to contribute to tissue specificity. Mutations in the *smg* genes, which result in stabilization of aberrant mRNAs, allow *unc-54* expression in the absence of otherwise essential 3′ sequences. The tissue specificity of the resulting expression is normal, indicating that these 3′ sequences are not necessary for tissue specificity. In addition, this region from *unc-54* can be replaced in *lacZ* fusions by the 3′-untranslated leader sequence from the gene *glp-1*, which has no known muscle association. The resulting fusions retain expression in just body wall muscle cells (not shown). The *unc-54* 3′ end can also function in a very wide variety of somatic tissues (*e.g.*, FIRE, HARRISON and DIXON 1990), suggesting this region does not contain sequences which specify body wall muscle expression.

Tissue specificity of myosin gene expression results from multiple positively acting control elements: Our data argue strongly that the myosin heavy chain gene promoters are controlled fundamentally by activators rather than repressors of expression. In particular, none of the promoter deletions that we tested exhibit ectopic expression of *lacZ* or myosin. This could conceivably reflect the biological constraints under which the gene has evolved. Muscle myosin heavy chain gene products must be expressed both at high levels and in very restricted patterns. Any expression in non-muscle cells would likely lead to interference with normal actin based processes. *C. elegans* has apparently evolved a control system in which both promoters and enhancers have intrinsic muscle specificity. This should serve to decrease the possibility of spurious expression in non-muscle cells.

The ability of enhancer and promoter sequences to independently direct proper expression was particularly evident in our assays of mutant rescue by *unc-54* deletions. Constructs with either element alone were capable of rescuing the *unc-54* mutant phenotype. The results with promoter deletions were particularly surprising, given the view that sequences upstream of and around the initiation site direct placement of 5′ ends by RNA polymerase II. Transcripts initiated just upstream of the coding region were observed even in constructs with all the natural 5′-flanking sequence replaced by plasmid sequence. One possibility is that a hypothetical "initiator" element within the coding region helps direct polymerase to start within a measured region just upstream. Alternatively, transcription from the 5′ deletion constructs might initiate at many dispersed sites around the gene, while a post-transcriptional mechanism degrades those RNAs that cannot be properly processed or translated. In either case, the system has apparently evolved to allow both upstream and internal sequences to activate expression.

Why does each myosin gene have more than one

tissue-specific activator element? Although some tissue-specific activation was seen with individual elements, the frequency of observed expression increases when several activator elements are present in a single construct (Figure 4). Multiple control elements for each locus may be necessary to increase the probability of gene expression in a given muscle cell to 100%, and to ensure that expression begins synchronously in all cells (*e.g.*, WEINTRAUB 1988). In this context it should be noted that we examined only sequences within 0.8–3 kb of the myosin promoters. Thus it is quite possible that there could be additional body wall muscle specific enhancers upstream of each of these genes, which function to increase frequency of expression but are not uniquely required for activity.

In other studies, regulatory sequences controlling *C. elegans* gene expression in several other tissues have been analyzed, using methods similar to those described here (AAMODT, CHUNG and MCGHEE 1991; WAY *et al.* 1991; MACMORRIS *et al.* 1992). The vitellogenin genes, which are regulated by activators, behave with broad similarity to the myosin genes (MACMORRIS *et al.* 1992). In contrast, the *ges-1* and *mec-3* genes are apparently controlled by both positive and negatively acting *cis*-acting elements (AAMODT, CHUNG and MCGHEE 1991; WAY *et al.* 1991). We expect that the full spectrum of conceivable regulatory mechanisms will be used in *C. elegans*, as has been seen with other complex organisms (*e.g.*, PTASHNE 1986; LAWRENCE 1992).

One common feature of the analyses to date of *cis*-acting sequences controlling expression in *C. elegans* (AAMODT, CHUNG and MCGHEE 1991; WAY *et al.* 1991; MACMORRIS *et al.* 1992; KRAUSE *et al.* 1990) is that control sequences have been found relatively close to and within the genes. Initial large scale sequencing studies of the *C. elegans* genome (SULSTON *et al.* 1992) indicate a closer packing of genes in the *C. elegans* genome than that seen in mammalian and insect genomes; this may mandate in general a more compact organization of control sequences. The distances over which enhancers exert their effects could indeed be one of the forces driving evolution of genome size.

We are grateful to S. XU, D. MILLER, R. PULAK, P. ANDERSON, J. HODGKIN, R. WATERSTON, D. HARRISON, H. EPSTEIN, M. KRAUSE, J. KARN, D. DIXON, T. BLUMENTHAL, D. BROWN, Y. SHI, M. SEPANSKI, L. CONNAH, J. AHNN, W. KELLY and G. SEYDOUX for their help and suggestions. This work was supported by the National Institutes of Health (postdoctoral fellowship HD07532-03 to P.G.O. and grant R01-GM37706 to A.F.), the Carnegie Institution of Washington, and the Rita Allen Foundation.

LITERATURE CITED

AAMODT, E. J., M. A. CHUNG and J. A. MCGHEE, 1991 Spatial control of a gut-specific esterase during *Caenorhabditis elegans* development. Science **252:** 579–582.

ALBERTSON, D. G., 1985 Mapping muscle protein genes by *in situ* hybridization using biotin-labeled probes. EMBO J. **4:** 2493–2498.

ANDERSON, P., and S. BRENNER, 1984 A selection for myosin heavy-chain mutants in the nematode *C. elegans*. Proc. Natl. Acad. Sci. USA **81:** 4470–4474.

ARDIZZI, J. P., and H. F. EPSTEIN, 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.

AUSTIN, J., and J. KIMBLE, 1987 *glp-1* is required in the germ line for regulation of the decision between mitosis and meiosis in *C. elegans*. Cell **51:** 589–599.

BEJSOVEC, A., and P. ANDERSON, 1990 Functions of the myosin ATP and actin binding sites are required for *C. elegans* thick filament assembly. Cell **60:** 133–140.

BLUMENTHAL, T., and J. THOMAS, 1988 Cis and trans mRNA splicing in *C. elegans*. Trends Genet. **4:** 305–308.

BRENNER, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71–94.

CALZONE F. J., R. J. BRITTEN and E. H. DAVIDSON, 1987 Mapping of gene transcripts by nuclease protection assays and cDNA primer extension. Methods Enzymol. **152:** 611–632.

CHOI, T., HUANG, M., C. GORMAN and R. JAENISCH, 1991 A generic intron increases gene expression in transgenic mice. Mol. Cell. Biol. **11:** 3070–3074

CUMMINS, C., and P. ANDERSON, 1988 Regulatory myosin light-chain genes of *Caenorhabditis elegans*. Mol. Cell. Biol. **8:** 5339–5349.

DIBB, N. J., BROWN, D. M., KARN, J., MOERMAN, D. G., S. L. BOLTEN and R. H. WATERSTON, 1985 Sequence analysis of mutations that affect the synthesis, assembly and enzymatic activity of *unc-54* myosin heavy chain of *C. elegans*. J. Mol. Biol. **183:** 543–551.

DIBB, N. J., MARUYAMA, I. N., M. KRAUSE and J. KARN, 1989 Sequence analysis of the complete *Caenorhabditis elegans* myosin heavy chain gene family. J. Mol. Biol. **205:** 603–613.

EIDE, D., and P. ANDERSON, 1985a The gene structures of spontaneous mutations affecting a *C. elegans* myosin heavy chain gene. Genetics **109:** 67–79.

EIDE, D. J., and P. ANDERSON, 1985b Novel insertion mutation in *C. elegans*. Mol. Cell. Biol. **5:** 1–6.

EIDE, D., and P. ANDERSON, 1985c Transposition of Tc1 in the nematode *C. elegans*. Proc. Natl. Acad. Sci. USA **82:** 1756–1760.

EMMONS, S. W., 1988 The genome, pp. 47–79 in *The Nematode Caenorhabditis elegans*, edited by W. B. WOOD. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

EPSTEIN, H. F., R. H. WATERSTON and S. BRENNER, 1974 A mutant affecting the heavy chain of myosin in *C. elegans*. J. Mol. Biol. **90:** 291–300.

FIELDS, C., 1990 Information content of *Caenorhabditis elegans* splice site sequences varies with intron length. Nucleic Acids Res. **18:** 1509–1512.

FIRE, A., 1986 Integrative transformation of *Caenorhabditis elegans*. EMBO J. **5:** 2673–2680.

FIRE, A., 1993 Histochemical techniques for locating *Escherichia coli* β-galatosidase activity in transgenic organisms. Gene Anal. Tech. Appl. **9:** 151–158.

FIRE, A., and R. H. WATERSTON, 1989 Proper expression of myosin genes in transgenic nematodes. EMBO J. **8:** 3419–3428.

FIRE, A., S. WHITE HARRISON and D. DIXON, 1990 A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. Gene **93**: 189–198.

FIRE, A., ALBERTSON, D., S. WHITE HARRISON and D. MOERMAN, 1991 Production of antisense RNA leads to effective and

specific inhibition of gene expression in *C. elegans* muscle. Development **113:** 503–514.

FROHMAN M. A., M. K. DUSH and G. R. MARTIN, 1988 Rapid amplification of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide. Proc. Natl. Acad. Sci. USA **85:** 8998–9002.

HODGKIN, J., A. PAPP, R. PULAK, V. AMBROS and P. ANDERSON, 1989 A new kind of informational suppression in the nematode *Caenorhabditis elegans*. Genetics **123:** 301–313.

HONDA, S., and H. F. EPSTEIN, 1990 Modulation of muscle gene expression in *Caenorhabditis elegans*: differential levels of transcripts, mRNAs and polypeptides for thick filament proteins during nematode development. Proc. Natl. Acad. Sci. USA **87:** 876–880.

JEFFERSON, R. A., M. KLASS, N. WOLF and D. HIRSH, 1987 Expression of chimeric genes in *Caenorhabditis elegans*. J. Mol. Biol. **193:** 41–46.

KARN, J., S. BRENNER and L. BARNETT, 1983 Protein structural domains in the *C. elegans unc-54* myosin heavy chain gene are not separated by introns. Proc. Natl. Acad. Sci. USA **80:** 4253–4257.

KAY, R. J., R. J. BOISSY, R. H. RUSSNAK and E. P. M. CANDIDO, 1986 Efficient transcription of *C. elegans* heat shock gene pair in mouse fibroblasts dependent on multiple promoter elements which can function bidirectionally. Mol. Cell. Biol. **6:** 3134–3143.

KLASS, M., D. AMMONS and S. WARD, 1988 Conservation in the 5′ flanking sequences of transcribed members of the *Caenorhabditis elegans* major sperm protein gene family. J. Mol. Biol. **199:** 15–22.

KLASS, M. R., S. KINSLEY and L. C. LOPEZ, 1984 Isolation and characterization of a sperm-specific gene family in the nematode *C. elegans*. Mol. Cell. Biol. **4:** 529–537.

KOZAK, M., 1991 Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. **266:** 19867–19870

KRAUSE, M. 1986 Actin gene expression in the nematode *Caenorhabditis elegans*. Ph.D. Thesis, University of Colorado, Boulder.

KRAUSE, M., and D. HIRSH, 1987 A trans-spliced leader sequence on actin mRNA in *C. elegans*. Cell **49:** 753–761.

KRAUSE, M., FIRE, A., WHITE HARRISON, S, J. PRIESS and H. WEINTRAUB, 1990 CeMyoD accumulation defines the body wall muscle cell fate during *C. elegans* embryogenesis. Cell **63:** 907–919.

LAMBIE, E., and J. KIMBLE, 1991 Two homologous regulatory genes, *lin-12* and *glp-1*, have overlapping functions. Development **112:** 231–240

LAWRENCE, P., 1992 *The Making of a Fly: The Genetics of Animal Design*. Blackwell Scientific Publications, Boston.

MACLEOD, A. R., J. KARN and S. BRENNER, 1981 Molecular analysis of the unc-54 myosin heavy-chain gene of *C. elegans*. Nature **291:** 386–390.

MACLEOD, A. R., R. H. WATERSTON and S. BRENNER, 1977 An internal deletion mutant of a myosin heavy chain in *C. elegans*. Proc. Natl. Acad. Sci. USA **74:** 5336–5340.

MACLEOD A. R., WATERSTON R. H., R. M. FISHPOOL and S. BRENNER, 1977 Identification of the structural gene for a myosin heavy-chain in *C. elegans*. J. of Mol. Biol. **114:** 133–140.

MACLEOD, A. R., KARN, J., R. H. WATERSTON and S. BRENNER, 1979 The unc-54 myosin heavy chain gene of *C. elegans*: a model system for the study of genetic suppression in higher eukaryotes, pp. 301–312 in *Nonsense Mutations and tRNA Suppressors* (Proceedings of the EMBO Laboratory Course and Aarhus University 50 Year Anniversary Symposium, July 1978), edited by C. E. CELIS and J. D. SMITH. ACADEMIC PRESS, LONDON.

MACMORRIS, M., BROVERMAN, S., GREENSPOON, S., LEA, K., MADEJ, C., T. BLUMENTHAL and J. SPIETH, 1992 Regulation of vitellogenin gene expression in transgenic *Caenorhabditis elegans*–short sequences required for activation of the *vit-2* promoter Mol. Cell. Biol. **12:** 1652–1662.

MARCK, C., 1988 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16:** 1829–1836.

MCCOMBIE, W. R., M. D. ADAMS, J. M. KELLY, M. G. FITZGERALD, T. R. UTTERBACK, M. KHAN, M. DUBNICK, A. R. KERLAVAGE, J. C. VENTER and C. FIELDS, 1992 *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease homologs. Nature Genet. **1:** 124–131.

MELLO C. C., J. M. KRAMER, D. STINCHCOMB and V. AMBROS, 1991 Efficient gene transfer in *C.elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J. **10:** 3959–3970.

MELTON, D. A., P. A. KRIEG, M. R. REBAGLIATI, T. MANIATIS, K. ZINN and M. R. GREEN, 1984 Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:** 7035–7056.

MILLER, D. M., F. E. STOCKDALE and J. KARN, 1986 Immunological identification of the genes encoding the four myosin heavy chain isoforms of *C. elegans*. Proc. Natl. Acad. Sci. USA **83:** 2305–2309.

MILLER D. M., I. ORTIZ, G. C. BERLINER and H. F. EPSTEIN, 1983 Differential localization of two myosins within nematode thick filaments. Cell **34:** 477–490.

MITCHELL, E. J., J. KARN, D. M. BROWN, A. NEWMAN, R. JAKES and J. KENDRICK-JONES, 1989 Regulatory and essential light-chain-binding sites in myosin heavy chain subfragment-1 mapped by site-directed mutagenesis. J. Mol. Biol. **208:** 199–205.

NEUBERGER, M., and G. T. WILLIAMS, 1988 The intron requirement for immunoglobin gene expression is dependent on the promoter. Nucleic Acids Res. **16:** 3969–3977.

PRIESS, J. R., H. SCHNABEL and R. SCHNABEL, 1987 The *glp-1* locus and cellular interactions in early *C. elegans* embryos. Cell **51:** 601–611.

PTASHNE, M., 1986 *A Genetic Switch: Gene Control and Phage Lambda*. Cell Press, Cambridge, Mass.

PULAK R. A., and P. ANDERSON, 1988 Structures of spontaneous deletions in *Caenorhabditis elegans*. Mol. Cell. Biol. **8:** 3748–3754.

ROSENQUIST, T. A., and J. KIMBLE, 1988 Molecular cloning and transcript analysis of *fem-3*: a sex-determination gene in *Caenorhabditis elegans*. Genes Dev. **2:** 606–616.

SCHULTZ, J. R., T. TANSEY, L. GREMKE and R. V. STORTI, 1991 A muscle-specific intron enhancer required for rescue of indirect flight muscle and jump muscle function regulates *Drosophila* tropomyosin I gene expression. Mol. Cell. Biol. **11:** 1901–1911

SPIETH, J., K. DENISON, S. KIRTLAND, J. CANE and T. BLUMENTHAL, 1985 The *C. elegans* vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes. Nucleic Acids Res. **13:** 5283–5295.

SULSTON, J., and J. HODGKIN, 1988 Methods, pp. 587–606 in *The nematode Caenorhabditis elegans*, W. B. Wood, ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SULSTON, J., Z. DU, K. THOMAS, R. WILSON, L. HILLIER, R. STADEN, N. HALLORAN, P. GREEN, J. THIERRY-MIEG, L. QIU, S. DEAR, A. COULSON, M. CRAXTON, R. DURBIN, M. BERKS, M. METZSTEIN, T. HAWKINS, R. AINSCOUGH and R. WATERSTON, 1992 The *C. elegans* genome sequencing project: a beginning. Nature **356:** 37–41

TRENT, C., N. TSUNG and H. R. HORVITZ, 1983 Egg-laying defective mutants of the nematode *C. elegans*. Genetics **104:** 619–647.

WATERSTON R. H., 1988 Muscle, pp. 281–335 in *The Nematode*

Caenorhabditis elegans, edited by W. B. WOOD. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

WAY, J. C., and M. CHALFIE, 1988 mec-3, a homeobox-containing gene that specifies differentiation in the touch receptor neurons in C. elegans. Cell **54:** 5–16.

WAY, J. C., WANG, L., J.-Q. RUN and A. WANG, 1991 The *mec-3* gene contains *cis*-acting elements mediating positive and negative regulation in cells produced by asymmetric cell division in *Caenorhabditis elegans*. Genes Dev. **5:** 2199–2211.

WEINTRAUB, H., 1988 Formation of stable transcription complexes as assayed by analysis of individual templates. Proc. Natl. Acad. Sci. USA **85:** 5819–5823.

WHITE J., 1988 The anatomy, pp. 81–122 in *The Nematode Caenorhabditis elegans*, edited by W. B. WOOD. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

WICKENS, M., 1990 How the messenger got its tail: addition of poly(A) in the nucleus. Trends Biochem. Sci. **15:** 277–281.

YAOITA, Y., Y. SHI and D. D. BROWN, 1990 *Xenopus Laevis* α and β thyroid hormone receptors Proc. Natl. Acad. Sci. USA **87:** 7090–7094.

YOCHEM, J., K. WESTON and I. GREENWALD, 1988 The *Caenorhabditis elegans lin-12* gene encodes a transmembrane protein with overall similarity to Drosophila Notch. Nature **335:** 547–550.

Communicating editor: R. K. HERMAN