Mol Gen Genet (1994) 242:346–357

MGG

The *Caenorhabditis elegans unc-60* gene encodes proteins homologous to a family of actin-binding proteins

Kim S. McKim*, Camela Matheson, Marco A. Marra, Marcia F. Wakarchuk, David L. Baillie

Institute of Molecular Biology and Biochemistry, Department of Biological Sciences, Simon Fraser University, Burnaby, B.C. Canada V5A 1S6

Received: 10 May 1993 / Accepted: 28 July 1993

Abstract. Mutations in the *unc-60* gene of the nematode *Caenorhabditis elegans* result in paralysis. The thin filaments of the muscle cells are severely disorganized and not bundled with myosin into functional contractile units. Here we report the cloning and sequence of *unc-60*. Two *unc-60* transcripts, 1.3 and 0.7 kb in size, were detected. The transcripts share a single exon encoding only the initial methionine, yet encode proteins with homologous sequences. The predicted protein products are 165 and 152 amino acids in length and their sequences are 38% identical. Both proteins are homologous to a family of actin depolymerizing proteins identified in vertebrate, plant and protozoan systems. We propose that the *unc-60* locus encodes proteins that depolymerize growing actin filaments in muscle cells, and that these proteins are required for the assembly of actin filaments into the contractile myofilament lattice of *C. elegans* muscle. *unc-60* has an essential function in development, since one *unc-60* allele, *s1586*, has a recessive lethal phenotype. Our characterization of *s1586* has shown that it is a small deletion which disrupts both coding regions.

Key words: *Caenorhabditis elegans* – *unc-60* – Muscle – Actin-binding protein

Introduction

Muscle structure in the nematode *Caenorhabditis elegans* has striking similarities to vertebrate muscle (reviewed in Waterston 1988). The force of contraction is generated by the interaction between the myosin-containing thick filaments and the actin-containing thin filaments within a unit analogous to the vertebrate sarcomere. The thin filaments are attached to dense bodies, analogous to vertebrate Z lines, and interdigitate with the thick filaments which are centrally stacked in the sarcomere. An electron-dense material analogous to the vertebrate M line is also present. While there are also significant differences, the similarities to vertebrate muscle make *C. elegans* an attractive model system for the study of muscle structure and function.

Genetic analysis has identified more than 30 genes involved in muscle development and function (Waterston 1988). Most of these genes were identified through the isolation of mutants exhibiting impaired movement due to the dysfunction of muscle contraction (Brenner 1974). These genes encode components required for muscle structure and function. Many of these genes have known homologs in vertebrate muscle. *unc-54*, for example, encodes one of the four myosin isoforms. *unc-22* (Benian et al. 1989) is similar to the proteins titin and projectin, and *unc-52* encodes a protein that resembles perlecan (Rogalski et al. 1993). The genetic and molecular analyses of *unc-54*, *unc-22* and *unc-52* have provided insights into the possible functions of the related proteins in vertebrate and insect systems.

Structural analysis with polarized light and electron microscopy (Waterston et al. 1980) has revealed the defects in the body wall muscle resulting from mutations in many *C. elegans* muscle genes. These data allow the genes to be classified according to their mutant phenotypes (Waterston 1988). Mutations in *unc-54* affect thick filament organization, consistent with the fact that *unc-54* encodes the major body wall myosin. Three genes, other than the actin genes, are known to be involved in the formation of functional thin filaments. These genes are *unc-78*, *unc-94* and *unc-60* (Waterston 1988). In *unc-60* mutants, large aggregates of thin filaments are found at the ends of the body wall muscle cells (Waterston et al. 1980), but the thick filaments are almost normal. Previous genetic characterization of this gene resulted in the identification of eleven alleles (Brenner 1974; McKim et al. 1988; Johnsen and Baillie 1991; M.F. Wakarchuk, unpublished results) and these have been used to generate a recombination-based fine structure map

Communicated by B. Judd

* *Present address:* Department of Genetics, University of California, Davis, CA 95616, USA

Correspondence to: D. Baillie

(McKim et al. 1988). Here we report the cloning and sequence of *unc-60*. The predicted protein products encoded by the *unc-60* gene are homologous to a family of actin-binding proteins. These proteins function to polymerize and depolymerize actin, a role consistent with the *unc-60* mutant phenotype.

Materials and methods

Genetics. The *C. elegans* strains used in this study were derived from the wild-type N2 strain (var. Bristol). The *C. briggsae* strain G16 was obtained from the Caenorhabditis Genetics Center (Columbia, MO.). All the deficiencies used in this study were described previously (Fig. 1) (Rosenbluth et al. 1988; Johnsen and Baillie 1991; Stewart et al. 1991). Deficiency homozygotes were generated for polymerase chain reaction (PCR) by crossing *unc-46(e177)/++* males to *sDfx unc-46; dpy-18/eT1(III;V)* hermaphrodites. The Unc-46 progeny were set on plates for 12-24 h and then removed. Two to three days later the arrested progeny (either embryos or larvae) were collected for PCR analysis.

Polymerase chain reaction procedures. PCR amplifications were carried out in 25 µl volumes. For amplification from the deficiency homozygotes, we used a procedure developed by Barstead and Waterston (1991). The arrested larval homozygotes were placed on the lid of a 0.5 ml Eppendorf tube containing 2.5 µl of lysis buffer (50 mM KCl, 10 mM TRIS pH 8.2, 2.5 mM $MgCl_2$, 0.45% NP-40, 0.45% Tween 20, 0.01% gelatin, 10 µg/ml proteinase K). The deficiency homozygotes that arrested as embryos were first taken up in chitinase solution [20 mg/ml chitinase (Sigma), 50 mM NaCl, 70 mM KCl, 2.5 mM $MgCl_2$, 2.5 mM $CaCl_2$] with a drawn-out capillary tube, and then placed in the lysis buffer. The 0.5 ml Eppendorf tubes were then spun briefly, frozen at −70° C for 10 min and then incubated at 65° C for 60 min. After incubation for 15 min at 95° C to inactivate the proteinase K, the rest of the PCR components were added [final reaction concentrations were: 25–50 pmol of each primer, 10 mM TRIS pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 0.125 mM of each NTP, 0.25 units *Taq* polymerase (Cetus)]. Typically PCR was initiated by denaturing at 95° C for 10 min, annealing for 1 min at 58° C and polymerizing at 72° C for 2 min. For the succeeding 25 cycles, denaturation time was reduced to 1 min, and in the final cycle, polymerization was carried out for 10 min.

Fig. 1. Genetic map of the left half of chromosome V, modified from Johnsen and Baillie (1991) and Stewart et al. (1991). Relevant deficiencies are shown below the genetic map. Deficiency endpoints are indicated by *vertical bars*. The scale bar corresponds to one map unit

Generation of primers. Primers for PCR were generated from the sequence of cosmid subclones. The cosmids were subcloned into pBluescript (Stratagene) using either *Hin*dIII- or *Eco*RI+*Hin*dIII-cut mini-prep DNA. The primers were designed from the sequence obtained using T3 and T7 primers, with the aid of the OLIGO program (Rychlik and Rhoads 1989).

General molecular biology methods. Genomic DNA was prepared by a method modified by J. Curran and D.L. Baillie from that of Emmons et al. (1979). The details are described in Starr et al. (1989). Restriction digestions of 2–5 µg of genomic DNA were carried out under conditions recommended by the enzyme manufacturers (BRL or Pharmacia). The digested DNA was electrophoresed in 0.7% agarose gels and then transferred to Genebind (Pharmacia) or Genescreen (DuPont) membranes with 10×SSC. Probes were ^{32}P-labelled using random priming (Feinberg and Vogelstein 1984). Alternatively, probes were labelled and hybridization detected using the ECL Random Prime labelling and detection systems (Amersham). The nylon membranes were hybridized with the probe overnight at 65° C in 5×SSPE, 0.3% SDS, 2.5×Denhardt's. For high stringency, the filters were washed in 0.2×SSPE, 0.2% SDS at 68° C for 1 h. For low stringency, which was used for cross-species hybridization, the filters were washed in 2× or 5×SSPE, 0.2% SDS at 65° C for 1 h.

The restriction map of F53E2 was generated in two steps. First, all the *Sal*I, *Pst*I and *Hin*dIII fragments from F53E2 were subcloned into pBluescript (Stratagene). Second, we used Southern hybridization to determine the fragment overlaps.

DNA sequencing was by the dideoxy chain terminator method of Sanger (1977) using either Sequenase 2.0 (US-Biochemicals) or automated sequencing on an Applied Biosystems (ABI) Model 373A automated sequencing machine using protocols supplied by ABI. DNA was prepared using the alkaline lysis method. DNA to be manually sequenced was purified with Pharmacia Miniprep Spun Columns. Cycle sequencing reactions (ABI) were used for the ABI automated sequencer, and purified DNA was not required. Cosmid subclones in pBluescript were used for the analysis of genomic DNA. The cDNA clones were isolated from a Lambda ZAP library (Stratagene) constructed by Barstead and Waterston (1989). Exonuclease III-generated deletions (Henikoff 1987) of genomic or cDNA clones for sequencing were prepared using the Erase-a-Base system (Promega).

For Northern blots, 30 µg of total RNA isolated from N2 animals were separated on a 1.2% agarose gel and blotted to Genescreen (DuPont). Filters were prehybridized and then hybridized with the probe overnight at

68° C in 5 × SSPE, 0.3% SDS, and 5 × Denhardt's. Filters were washed twice with agitation at 68° C in 0.4% SDS and 1 × SSPE for 5 min and then 30 min.

Germline transformation. Experiments were performed using an inverted Zeiss microscope with Nomarski optics. Two methods were used to perform germline transformations. In the first method (Fire 1986), nematodes were immobilized on agarose pads and individual mature oocytes were injected. In the second method (Mello et al. 1991), DNA was injected into the gonadal syncytium. Most DNA preparations for germline transformation experiments were mixed with the plasmid pRf4 containing the dominant Roller allele *rol-6(su1006)* (Kramer et al. 1990). The Roller gene acted as a marker for transformation. Cosmid and plasmid DNAs used for germline transformation were prepared by the alkaline lysis procedure (Sambrook et al. 1989). DNA was purified by either repeated ethanol precipitation, CsCl density centrifugation, or Miniprep Spun Columns (Pharmacia). Purified DNA for germline transformation was mixed with Fire's (1986) injection solution or TE buffer (10 mM TRIS-HCl, 1 mM EDTA pH 8) or double-distilled sterile H_2O at 100 ng/µl.

DNA sequence analysis. Nucleotide sequences were aligned, formatted and translated using the Eyeball Sequence Editor (ESEE) program (Cabot and Beckenbach 1989). Amino acid sequences predicted using ESEE were used to search entries in the SwissProt, PIR, GenPept, and GPUpdate peptide sequence databases, using the BLAST algorithm (Altschul et al. 1990). Computations were performed at the National Center for Biotechnology Information (NCBI) using the BLAST network service.

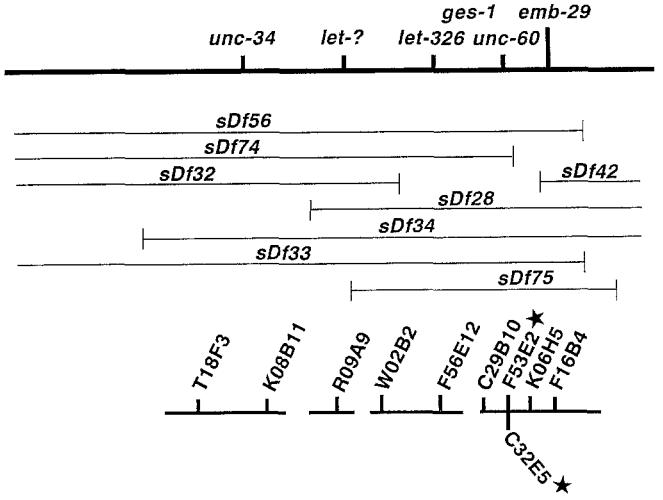

Fig. 2. Physical mapping in the *unc-60* region. The genetic map is shown on top, with the deficiencies in the middle and the cosmid contigs on the bottom. Only the cosmids from which DNA sequences were obtained in order to design primers for PCR amplifications are indicated on the contigs by *vertical lines*. Other cosmids have been left out. The gaps between the contigs are joined by YAC clones (Coulson et al. 1988). The *starred* cosmids contain *unc-60(+)* activity, as determined by transformation rescue experiments. F53E2 is deleted with respect to C32E5

Results

Alignment of the physical and genetic maps in the unc-60 *region*

On the genetic map, *unc-60* is located near the left end of chromosome V. Deficiency mapping has placed it between *let-326* and *emb-29* (Fig. 1). *ges-1* is within 1 cM of *unc-60* (McGhee and Cotrell 1986) and is deleted by *sDf74* (McGhee et al. 1990). An array of overlapping cosmids and yeast artificial chromosomes from the region ("contigs"; Fig. 2) has been constructed (Coulson et al. 1986; 1988). *ges-1* has been cloned, and is located on cosmid C29B10 (Kennedy et al. 1993).

In order to align the genetic and physical maps of the *unc-60* region, we examined genetically identified deficiencies of the region (Fig. 1) using PCR and primers derived from cosmid sequences (see Materials and methods). Deficiency homozygotes, arrested in development as embryos or early stage larvae, were used in the PCR amplifications (see Materials and methods). If the deficiency deleted at least one of a primer pair's annealing sites, an amplification product would not be observed. If the deficiency deleted neither primer annealing site, an amplification product would be observed. A primer pair from another region of chromosome V or another chromosome was included in each reaction as a positive control. The results are shown in Fig. 2.

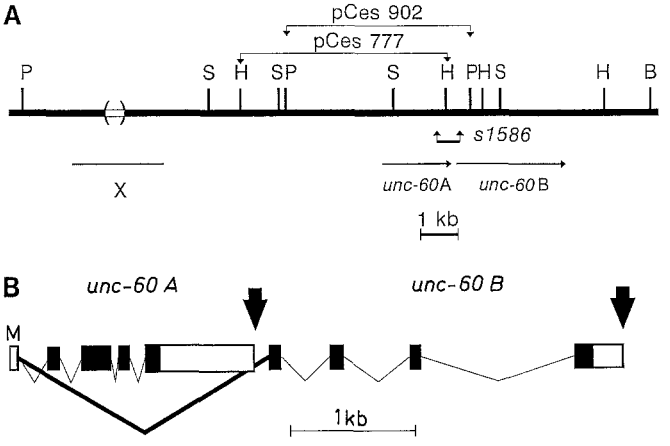

Fig. 3A, B. Molecular organization of *unc-60*. **A** Restriction map of F53E2 showing the positions of known coding regions, pCes 902, pCes 777, and the approximate region deleted in *unc-60(s1586)*. *unc-60A* and *unc-60B* are indicated. The X denotes the cDNA that resembles UBF1 and nucleolin; the exact position of the cDNA within the restriction fragment is not known. Only a portion of the entire 6.0 kb *Sal*I fragment used to detect this cDNA is shown. The *stippled box bordered by parentheses* indicates that a portion of F53E2 is deleted with respect to the other cosmid that rescues *unc-60*, C32E5. The deletion *s1586* was detected using a combination of PCR and Southern blot analysis. Three primer pairs were used to amplify all of transcript A. One pair, flanking the 3'UTR (see Fig. 6), did not amplify DNA from *s1586* homozygotes. Restriction enzyme sites are: S, *Sal*I; H, *Hin*dIII; P, *Pst*I; B, *Bam*HI. **B** Splicing pattern of *unc-60* based on genomic and cDNA sequences (Fig. 6). The *open boxes* indicate untranslated regions. The *inverted arrowheads* indicate the positions at which polyadenylation occurs. The *heavy line* shows the alternative splicing of the first exon to transcript B. M indicates the methionine that initiates both the Unc-60A and Unc-60B proteins

A primer pair from the cosmid K06H5 maps between *sDf74* and *sDf42,* in the same region as *emb-29*. A primer pair from F16B4 maps inside *sDf42*. These data orient the rightmost contig with respect to the genetic map as shown in Fig. 2. To determine which cosmid clone contained the *unc-60* gene, we transformed *unc-60(m35)* hermaphrodites with cosmid clones from the *ges-1* region. Two overlapping cosmids, F53E2 and C32E5, were found to rescue *unc-60(m35)* mutants (Fig. 2). Since these two cosmids lie to the right of the cosmid carrying *ges-1* (cosmid C29B10; Kennedy et al. 1993), we have positioned *unc-60* to the right of *ges-1* on the genetic map.

Characterization of F53E2 and C32E5 by Southern blotting revealed that F53E2 is deleted with respect to C32E5. We have concluded that the deletion is in the 6.0 kb *Sal*I fragment, a portion of which is indicated in the restriction map of F53E2 (Fig. 3), because this fragment hybridized to two large *Sal*I fragments on Southern blots containing C32E5 and *C. elegans* genomic DNA. Because F53E2 is deleted with respect to C32E5, yet still rescues *unc-60(m35)* mutant animals, we chose to use it as our *unc-60* cloning source.

Identification of conserved sequences and coding regions near unc-60

In an effort to identify coding regions in the vicinity of *unc-60,* we employed the technique of inter-species hybridization. It has been shown that coding regions can be detected using hybridization to closely related *Caenorhabditis* species (Snutch 1984; Heine and Blumenthal 1986; Prasad and Baillie 1989). Restriction fragments from F53E2 were used to probe Southern blots containing DNA from *C. elegans* and the closely related *Caenorhabditis briggsae*. To our surprise, most of the probes hybridized at low stringency to a family of *C. elegans* and *C. briggsae* fragments. Based on hybridization patterns, there are at least three different families of repetitive DNA in this cosmid (data not shown), and their sequences are conserved across species boundaries. Only one region of F53E2, represented by a 4.8 kb *Pst*I fragment (pCes 902) and a 5.0 kb *Hin*dIII fragment (pCes 777; Fig. 3), contained no repetitive DNA. Both of these fragments hybridized to single-copy sequences on Southern blots containing *C. elegans* and *C. briggsae* genomic DNA (see Fig. 4 for the pCes 902 result). As described below, these fragments contain the *unc-60* coding region.

Using an alternative search strategy for coding regions, we screened a cDNA library [obtained from Barstead and Waterston (1989)] using restriction fragments from F53E2 as probes. The positive cDNA clones obtained were analyzed by Southern blotting and DNA sequencing. Based on hybridization results, we identified three classes of cDNA clones. The positions of the cDNA clones on the F53E2 restriction map are shown in Fig. 3. The analysis described below identified the transcripts labelled *unc-60* A and *unc-60* B in Fig. 3 as products of the *unc-60* gene.

One positive clone was identified in a screen of 50 000 phage using the 6.0 kb *Sal*I fragment as a probe (Fig. 3). The protein predicted from the sequence of this cDNA is acidic and resembles certain DNA-binding proteins such as the transcription factor UBF1 and nucleolin (data not shown).

The cDNA clones corresponding to transcript A (Fig. 3) were abundant in the library. Twelve cDNA clones were isolated in a screen of 30 000 plaques using pCes 902 as a probe. Six of these clones were approximately 1.3 kb in size. The rest were smaller and, based on sequence analysis, represented truncations of the larger clones. The cDNA clones corresponding to transcript B

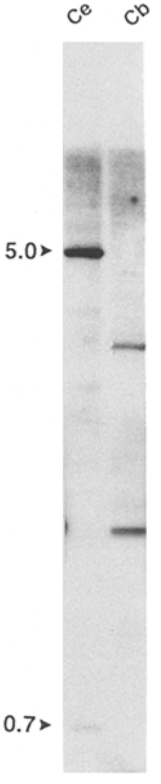

Fig. 4. Sequences conserved between *C. elegans* and *C. briggsae*. Plasmid pCes 902 (Fig. 3) was ^{32}P-labelled and hybridized to Southern blots containing *Hin*dIII-digested genomic DNA from *C. elegans* (Ce) and *C. briggsae* (Cb). The results indicate that pCes 902 contains a single copy element also found in *C. briggsae* DNA. This element is the *unc-60* gene (see Results). Size markers (in kb) are indicated by *arrowheads*

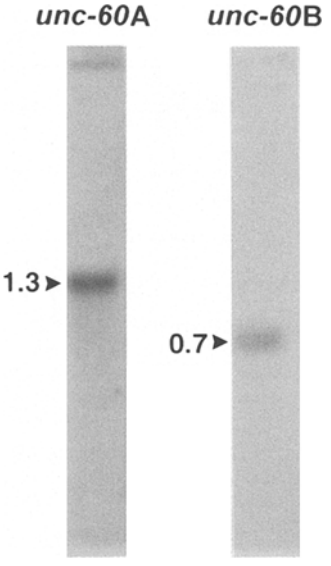

Fig. 5. Northern blot analysis of *unc-60*. Full-length *unc-60* cDNA clones were ^{32}P-labelled and hybridized to total *C. elegans* RNA (20 µg per lane). The results are shown for *unc-60A* and *unc-60B*. The sizes shown (in kb) were calculated using RNA molecular size standards (BRL)

```
aaaagctggtacaggcaagcactctgactggcggttgtgttgttggggctgggggttaggccgaagggacggccccgtctcttttttttttgcgtctttctcatcctgcttctgtgccaat  120
cacaatggactttcttggctcttgttctctctctctctgtttccatcactttaccaatttacctcacacttatgacccttatgactcttatctgggttacactttttctttgtgttga  240

                                                                                    AGTTTGAG*                                        8
gcttcttcactatcgtttttcccttctcgtgattcgattattgattgattaaaatcatttttcttgcagGAAACTCAACTTGATTCTATTTCCCCACCACAAACCAACAACAACATCTC  360

unc-60A
    M                                                                                                                        1
AAAATGgtgagtttgagattttaattcgcttcaatattttaaattgaaagaaatttggaatgttatgtttggcgtgttttttttcattgaatttccaaaatttcacactaaaatggaa  480
    M                                                                                                                        1
unc-60B

                                                                                S  S  G  V  M  V  D  P  D  V  Q  T  S  F  Q  K  L  S  E  G  R    22
tggttttctcttctttttgcctaacctaacctaacctatgtgtgcctgttttctagAGTTCCGGTGTCATGGTCGACCCAGATGTGCAGACATCTTTCCAAAAGCTCTCCGAGGGACGCA  600

K  E  Y  R  Y  I  I  F  K  I  D                                                                                            33
AGGAGTACCGCTACATCATTTTCAAGATCGACgtgagtttttaaaaaataaaatctgaatcagatcaatttaaaaaaatttcgtgccactttttgtttttttttgttgaaaaatttgaaaa  720

                                                                                    E  N  K  V  I  V  E  A  A  V  T  Q  D  Q  L  G  I  T    51
atcctcaaattaattgttttagatgaacattattgatttccctattaaattgcaacattttccagGAGAACAAGGTGATCGTGGAAGCCGCGGTGACTCAGGATCAGCTCGGCATCACTG  840

G  D  D  Y  D  D  S  S  K  A  A  F  D  K  F  V  E  D  V  K  S  R  T  D  N  L  T  D  C  R  Y  A  V  F  D  F  K  F  T  C    91
GAGACGACTATGATGACTCTTCCAAGGCCGCTTTCGACAAATTCGTCGAGGACGTGAAGTCTCGAACCGATAATCTGACCGATTGCCGCTACGCCGTTTTCGACTTCAAGTTCACGTGCA  960

S  R  V  G  A  G  T  S  K  M  D  K  I  I  F  L  Q  I                                                      C  P  D  G  A  114
GTCGTGTTGGAGCCGGCACGAGCAAGATGGACAAGATCATCTTCCTCCAGATgtaagcgcttgatcctattggtgaattattgtaccatctaatttttttttccagCTGCCCAGATGGTGC 1080

 S  I  K  K  K  M  V  Y  A  S  S  A  A  A  I  K  T  S  L  G  T  G  K  I  L  Q  F  Q                                     142
TTCTATCAAGAAAAAGATGGTGTACGCTTCGTCCGCCGCCGCCATCAAGACTTCTCTCGGAACCGGCAAAATCCTTCAGTTCCAGgtgagaaatctcgataatttttacaattagaaaaa 1200

                            V  S  D  E  S  E  M  S  H  K  E  L  L  N  K  L  G  E  K  Y  G  D  H  *                         165
aaaaatcaaattattaaaatttcagGTGTCTGACGAGAGCGAGATGAGCCACAAGGAACTCCTCAACAAGTTGGGCGAGAAATACGGAGATCACTAGATACACGGGTCTCTTCCGAAGTC 1320

GTCGTCGCCGTAACATCCTCGATTATTCTCGTCTCTGTCTGTCGCATTTTCTCCATTGCTTGATGATTTCTGATCGTTTTCACGTGTGCTCTTTGAATGTGCTAGTGCGTTGTTGTGGTG 1440
CTTTTTTCTTACCAACCTCACTACTTGTTCATTTTTTTTGGTGTCCTCAAAGAATCATCCGTTGAAAACATGACTGCGGCGATGTGTTTCGATGATGTTTCTTAAGCGTCTTCTTCACAT 1560
TTTCACCTCTTAACATTGCTAATTTATCCTCTATTTTTATCTGTTTTTTTCCTGATTGACGATTGTTTTTTTCCTGGGGGCAAGTGATTGTGTGTGCCCACTAATTTAATACAATGTGCC 1680
GGGTAACCTTCCGTACTATGAAGAAGTTGAGACATCCCCTCAGATTCTCAGTATCTTTATTGTTTAGTGAAATGCGTGGTAAATCCCTTCTCATTTACGATTATATTCTCTCTCTCTTGT 1800
AAAGCTTTCTATATTTATATAGTGTACTTTCCCCCACTGGAACTCTCAATCTCCGATCTCCTCACCATTCTTTTTTTTTCAAATCTTCTATTGCGTATTCATTTCCCTTCTTACTTTTTT 1920
CCCACCCTGCCCGATCCCCTTGAAGAATCACCCACCAAACCGTTCATGTGCTGTGCTTTTTAAGATTATCACAATGAAATCAAGTTTGGAGTTTCTAAcaattgttttaacacgcgtgga 2040

aaatttttgaaaggaaacaaaaattgacataaaaaccagaaaaatattacaaacacctttgagtacagGCTTCCGGAGTCAAAGTTGATCCATCCTGCAAGAACGCCTACGATCTTCTCCA 2160
                                                                                    A  S  G  V  K  V  D  P  S  C  K  N  A  Y  D  L  L  H    19

CAACAAGCATCAGCACAGCTACATCATCTTCAAGgtgagttattctgcgaaaaaaaaaatgcgtactgcacaacacatttgacgcgcaaaatatctcgtagcgaaaactacagtaattct 2280
 N  K  H  Q  H  S  Y  I  I  F  K                                                                                           30

ttaaatgagtactgtagctcttgtgtcgatttacgggctcaattattgcaattttctttataaatggtttccgtgaatcgacactagcgctacagtagtcatttaaaggattactgcagt 2400

tttcgctacgagatatttgatgtattttgcgcgtcaaatatgttgtgcagtacgcattttcagagttttctgttcccgtaataattaaataattttctaaatttgatttttttaaaaattt 2520

aaaaaataaataaatcgaaaaaaaataaatcaaaaaataatcagATTGACAAAAATGACACCGCCATCGTCGTCGAGAAGGTCGGAGAGAAGAATGCTCCGTACGCCGAGTTCGTTGAGG 2640
                                              I  D  K  N  D  T  A  I  V  V  E  K  V  G  E  K  N  A  P  Y  A  E  F  V  E    55

AGATGAAGAAGCTTGTTGAGGACGGAAAGGAGTGCAGgtaagcctaagcctgagactaggcctgagcctaaacctgagcataagcctaagcctaagaataagcctaagcctaacactaag 2760
E  M  K  K  L  V  E  D  G  K  E  C  R                                                                                       68

tcttagtctgagcctaaatgtcgcgtgtattctacgcgaaaagttgtgtttttgcgccaaaatacggtgctaggtctcgacacgactaatttttactagttaaatgaatgcagagaaaag 2880

aaggtgtgcgtctttaaatagtactgtaatttcaaaattattgtttccatggaatgttttattatcgaaaaaaattcacataaatattaagaaactgtgaaacattaatgaaaatttggc 3000

agcaatggaattttgaaactacattactctttaaaggcacatatgattttgaatttaacaaaaaagtttcgtgtcgagaccagttggtgcaaaaaaaccgcaaaatttcttgtctggata 3120

attatcaaacaacaaaactcgtcggaaatctcgattcaattcatccttttttgcagATATGCTGCCGTCGACGTCGAGGTTACTGTCCAAAGACAGGGAGCCGAGGGAACCAGCACACTC 3240
                                                        Y  A  A  V  D  V  E  V  T  V  Q  R  Q  G  A  E  G  T  S  T  L    89

AATAAGGTCATCTTCGTTCAATAgtgagtttcaacattgaaatattacacatacacgtggtgccaggctgtcccattgcggtttgatcttgaaaaaatgcgcgatttttttcccaaagaa 3360
 N  K  V  I  F  V  Q  Y                                                                                                    97

aatttgacgtcagcacgttcttaaccatgagaaatcagttgagaactctgcgtctctactccggctttttttgtagatctacgtagatcaagccggaatgctcgaaattgtccgcatttt 3480

tgcgttaaaaaaacggtaacactataatttcgcagcaacgaatgttttaaattactgtactctttaaaggcgcatacctgtttgaatttaacaaaaatttgtcgtgtcgagaccagatag 3600

cgtaaaatttcgcgtctggaaaatctagaaaatcgtccaaacaaaaaaccaagaaagttgataaaaatgaacgaatttactgaaattctcgttggaaaagtcggccatcaactaggaagc 3720

ctaggccaccaaaaagtcggccataaactagaaaaattcggccaccaaaaaaccggccatcaactaggatacctcggccatcaactagaaaactcggccatcaactagaaaacctgggcc 3840

atcaaaaagtcggccatcaacaaggaaaagtcggtcataaactagagaaactcggccatcaactaggaaaactcggccatcaactagaaatactcgaccagcaactagaaaaactcggct 3960

atcaactaggaaaactcggccaccaaaaaatcggccatcaacaagaaaaagtgggccatcaactagaaatcctcggccaccaaaaagtcgcccatcaactaggaaaacacggccatcaac 4080

tagaaaactcgaccatcaactagaaaacctcggccatcaaaaagtcggccattaactagaaaaagtcggccaccaaaaagtcggccaccaactaaaaaacctcggctatcaactaggaaa 4200

actcggccaccaaaaagtcggccatcaactagaaaaactcgaccatcaaaatgtcggccatcaactagaaaacctcggtcatcaactagaaaacctcggccatcaactaggaaacctcgg 4320
```

```
ccatcaactagaaaaactcggccaccaaaatgtcggccattaacaagaatctctttaaaaaaatttccatgtatttccacattccacttcccaaactcctatttttcagCTGCCCAGACAA 4440
                                                                                                                C  P  D  N  101

CGCCCCAGTCCGCAGACGTATGCTCTACGCCTCCTCGGTCCGTGCTCTCAAGGCGTCCCTCGGACTCGAGTCCCTCTTCCAAGTACAGGCCTCCGAGATGTCGGACCTCGACGAGAAGAG 4560
  A  P  V  R  R  R  M  L  Y  A  S  S  V  R  A  L  K  A  S  L  G  L  E  S  L  F  Q  V  Q  A  S  E  M  S  D  L  D  E  K  S  141

CGTGAAGAGCGACCTGATGTCCAACCAAAGAATCTAAGCCTGTCCAATTGAATCCCAATCACAAATTACAATTTCTTTCTGGTACTACTAGTCTCTCGGACCATCCAAGCCCGTCAAAAA 4680
  V  K  S  D  L  M  S  N  Q  R  I  *                                                                                          152

TGTGCCTTTAACTTTTTTTTTCTTGTTAAATTATGACTTTTCTTGTCTGTAATTCTGATTTCCCATTTTCTCTGGAATTTTTAAAAAAATATTTTTCTTTTAATAAATAAATCCGTTCCT 4800
TCAgaacatgtcagttgttgtttttccatgacgtccagacggcaaaaaacaataaaatcaattacccatttttttcagaggccccaaggggccctcagatgttatag                 4906
```

Fig. 6. Nucleotide and predicted amino acid sequences of *unc-60*. Sequence corresponding to cDNA is shown in upper case, intron and spacer DNA is *lower case*. The portion of the trans-spliced SL1 sequence attached to the cDNA clones is shown above the genomic sequence. The amino acid sequence derived from each transcript is shown above *(unc-60A)* or below *(unc-60B)* the DNA sequence. The inverted repeat sequence in the second *unc-60B* intron is *underlined,* as is the putative hairpin in the fourth *unc-60B* intron. The direct repeats in the fourth intron of *unc-60B* are indicated by a *double underline*. The primers that failed to amplify DNA from *s1586* homozygotes are indicated by *arrows* at positions 1317 and 2175 of the genomic sequence. A primer pair that was used to amplify DNA from *s1586* homozygotes is indicated at positions 832 and 1341 of the genomic sequence. The *unc-60* gene sequence is available under the Genbank accession number L18963

(Fig. 3) were less abundant in the library. Two cDNA clones were isolated in a screen of 50 000 plaques using the 3.5 kb *Sal*I-*Bam*HI fragment as a probe. Both clones were 700 bp in size. Probes which are specific for each transcript hybridize to a single band on Northern blots, and this band is the same size as the cDNA clones (Fig. 5).

Identification of unc-60 *transcripts*

Sequencing of the cDNA clones showed that transcripts A and B are generated through alternative splicing (Fig. 3 and 6). The longest cDNA clones from transcript A and transcript B share a small 58 nucleotide exon at their 5′ ends. These two transcripts are products of the *unc-60* gene, based on the following experiments, which showed that one *unc-60* mutation (*s1586;* Fig. 3) is a small deletion disrupting only these two transcripts. Originally isolated by Johnsen and Baillie (1991), *s1586* fails to complement *unc-60(m35)*and it is the only known *unc-60* mutation with a lethal phenotype; the *s1586* homozygotes do not develop past the mid-larval stage. The other 10 alleles of *unc-60* form an allelic series with varying phenotypic severity, but worms homozygous for any of these alleles are viable (McKim et al. 1988).

The deletion was detected in *unc-60(s1586)* DNA by PCR analysis. Primer pairs from transcript A were used in PCR amplifications, using DNA from *s1586* homozygotes as the amplification template (see Materials and methods). Evidence for the deletion was our inability to amplify by PCR the 3′ portion of transcript A in *unc-60(s1586)* homozygotes. Primer pairs specific to the 5′ portion of transcript A were used successfully to obtain amplification products. These results revealed a deletion in the 3′ end of transcript A in *unc-60(s1586)* mutants (Fig. 3) which is outside of the translated portion of the gene. We determined that the deletion does not extend beyond the neighboring cosmid (K06H5, Fig. 2) since primers from this cosmid amplified DNA from *s1586* homozygotes.

In order to determine the extent of the *s1586* deletion, genomic DNA from *s1586* heterozygotes and a *C. elegans* wild-type strain (N2) was blotted and probed with the

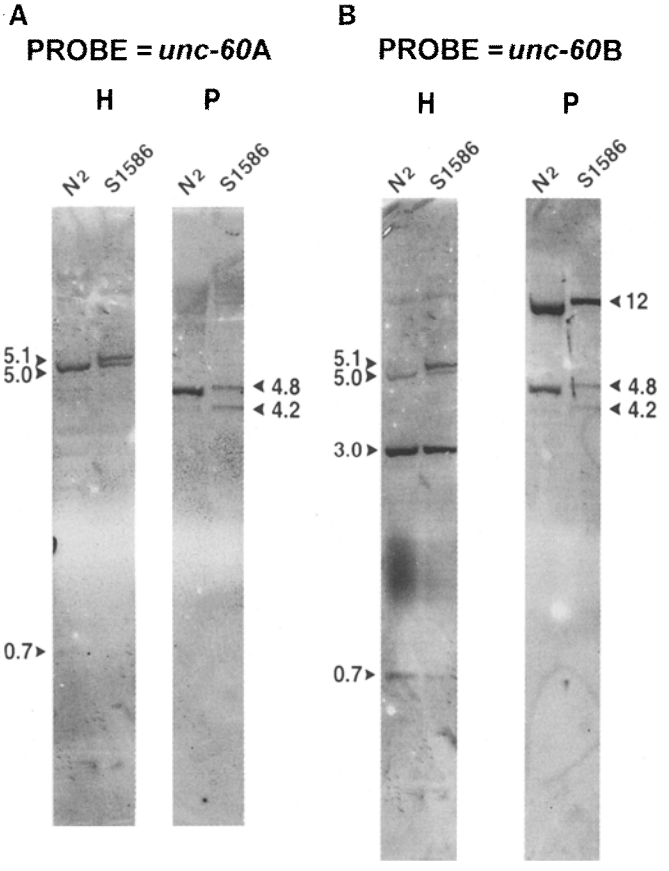

Fig. 7A, B. Identification of a DNA alteration associated with the *s1586* allele. Genomic DNA from wild-type (N2) and *s1586* heterozygotes was digested with *Hin*dIII (H) or *Pst*I (P), blotted, and hybridized with an *unc-60B* cDNA probe (**B**). The same filter was stripped and hybridized with an *unc-60A* cDNA probe (**A**). Probe preparation, hybridization, and hybridization detection were performed using the ECL Random Prime labelling and detection systems, using protocols and reagents supplied by the manufacturer (Amersham). **A** *unc-60A* hybridization showing the deletion of the *Hin*dIII site that separates *unc-60A* from *unc-60B* (see Fig. 3) in genomic DNA prepared from *s1586* heterozygotes. Molecular sizes are shown on the left in kb. **B** *unc-60B* hybridization showing the same polymorphisms detected by the *unc-60A* cDNA probe. For an interpretation of these hybridization patterns, see Results

cDNA clones for transcripts A and B. The two cDNA clones detected the same polymorphisms in genomic DNA digested with *Pst*I, *Sal*I, *Xba*I and *Hin*dIII. Examples of the *Hin*dIII and *Pst*I polymorphisms are shown in Fig. 7. Since the cDNAs do not overlap, and because PCR analysis had revealed that the left breakpoint of the deletion is in the 3′ UTR of transcript A, we conclude that the right breakpoint of the deletion must be within transcript B. The presence of a polymorphic 5.1 kb *Hin*dIII fragment detected by the two cDNAs suggests that the deletion removes the *Hin*dIII site separating transcript A from transcript B. Furthermore, the fact that the polymorphic *Pst*I fragment is smaller than the wild-type fragment suggests that the rightmost deficiency breakpoint does not delete the *Pst*I site immediately to the right of the deleted *Hin*dIII site (Fig. 3 and 7). Comparisons of the sizes of the various polymorphic fragments to the wild-type band sizes reveals the *sl586* mutation is a deletion of approximately 600 bp.

In an attempt to show which transcript encodes the *unc-60* muscle-specific function, we injected *unc-60(m35)* mutants with F53E2 restriction fragments. Plasmid pCes 902, which contains all of transcript A and 2 kb of 5′ sequence, did not rescue *unc-60(m35)* homozygotes. These experiments were complicated by the fact that pCes 902 was poisonous to the worms; no transformed worms could be isolated unless the concentration of pCes 902 was kept low (1 ng/μl). Since this fragment alone does not rescue *unc-60(m35)*, it is possible that transcript B is required for *unc-60(+)* function. Alternatively, correct expression of transcript A could require the presence of transcript B, or vice versa. It is also possible that both transcripts are required for *unc-60* function. We do not think the latter possibility is likely since *unc-60* mutants are isolated at a frequency consistent with an average size gene requiring a single mutation for an observable phenotype (McKim et al. 1988). We have not injected a fragment containing only transcript B since all of transcript A is contained within the first intron of transcript B.

The unc-60 *sequence*

Sequence data were generated from several of the cDNA clones and genomic subclones obtained from the cosmid F53E2 (Fig. 6). The 1.3 kb cDNA clones from transcript A and the two 700 bp cDNA clones from transcript B are probably full length for three reasons. Firstly, analysis of the genomic and cDNA sequence indicates the mRNAs are trans-spliced at their 5′ ends to the SL-1 trans-spliced leader sequence (Krause and Hirsh 1987). The first 6 basepairs of the cDNA sequence are not found in the genomic sequence (Fig. 6). This divergent sequence is identical to the 3′ terminal nucleotides of the SL-1 sequence. At the point where the genomic and cDNA sequences diverge, the genomic sequence has a 3′ splice acceptor site (CTTGCAG; Fig. 6). Secondly, both cDNA clones end with a poly(A) sequence at their 3′ ends. Thirdly, cDNA probes which are specific for each transcript hybridize to mRNAs of a single size on Northern blots, and these mRNAs are the sizes predicted from the sequence analysis (Fig. 5). We have also identified a putative initiator ATG codon in the first exon of each transcript contained within the context of a *C. elegans* consensus sequence for translation initiation (M.D. Perry, G.L. Hertz and W.B. Wood, personal communication). The open reading frame of transcript A is 495 nucleotides long, encoding a 165 amino acid protein. The remainder of transcript A comprises 719 nucleotides of 3′ untranslated region (Fig. 6). The open reading frame of transcript B is 456 nucleotides long and encodes a 152 amino acid protein. Transcript B has 205 nucleotides of 3′ untranslated sequence.

From comparison of the *unc-60* cDNA and genomic sequences, we have deduced the splicing patterns shown in Fig. 3. The only exon shared by the two transcripts is the first. Transcript A is composed of the first five exons. All four introns are in the coding portion of the transcript and are 109, 153, 53 and 60 nucleotides in length (Fig. 3 and 6). Transcript B is composed of the first exon and the last four exons, and has relatively large introns, the first of which includes all of transcript A. The last three introns in transcript B are 370, 499 and 1165 nucleotides in length. The majority of the donor and acceptor site sequences are similar to the consensus sequence for *C. elegans* introns (A/GAGGTAAGTT-----TTTCAG-G/A; Emmons 1988). Two notable exceptions, which match the consensus sequence poorly, are the acceptor sites for the splice event between the first and second exons of each transcript (TTCTAG and GTACAG; Fig. 6). Perhaps this is a point at which the alternative splicing of this transcription unit is controlled.

The introns of transcript B contain repetitive elements

Within the 370 nucleotide second intron of transcript B is a sequence with a high degree of identity to sequences found in the *C. elegans* sex determination gene *fem-1* (Spence et al. 1990), the putative myogenic regulatory gene *hlh-1* (Krause et al. 1990) and two genes controlling cell death, *ced-3* and *ced-9* (S. Shaham and R. Horvitz, personal communication) (Fig. 6). As Spence et al. (1990) observed, this sequence is an inverted repeat and could form a stem-loop structure. The first 20 nucleotides at the 5′ end of the conserved sequence, however, are not part of the inverted repeat. At the beginning of the next intron, the sequence TAAGCCT is repeated six times, and the sequence GAG(A/C)(A/C)T is repeated four times. The fourth intron of transcript B contains a sequence that could form a stem-loop structure, beginning 60 nucleotides into the intron. In addition, the fourth intron contains direct repeats commencing 430 nucleotides into the intron, and extending for 675 nucleotides. These repeats are oriented head to tail with no spacer DNA between the repeat units. The repeat units are of two types; either long (21 nucleotides) or short (15 nucleotides). The consensus sequence for the repeating unit is (A/G)AAAACTCGGCCATC{AACTAG}, where the nucleotides in brackets are absent in the short form of the repeat. BLAST searches of the NIH nonredundant

```
Unc60A      MS---SGVmVDPDVqtSFQkLs--------EgRKeYRYIIFKIDENKviv
Unc60B      MA---SGVkVDPscKNAYd-Ll--------hnKhQHsYIIFKIDkN----
lily        MAnssSGMaVDdEcKlkFmeLka---------KRNFRFIVFKIEEkv---
protozoan   M----SGIaVsdDcvQkFNeLkl---------ghQHRYVtFKMnasn---
human       MA---SGVaVsdgVikvFNdMkvrksstpeEvKKrkKaVLFcLsEdK---
chick       MA---SGVtVndEVikvFNdMkvrksstpeEiKKrkKaVLFcLsddK---

Unc60A      EAAVTqDqLGItgdDydDSskAAFDkFVEDVKsrtDnltDCRYAVFDFkF
Unc60B      DTAIVVEKVG---------ekNAPYAEFVEEMKklvEdgKECRYAavDvEv
lily        -qqVTVERLG--------qp-NeSYDDFtEcLpp------nECRYAVFDFDF
protozoan   tevVVehvgGp---------NATYEDFksqLpe------RDCRYAIFDYEF
human       kniILeEgkeIlvgDvgqTvddPYAtFVkmLpd------KDCRYALYDatY
chick       kqiIVeEatrIlvgD-gDTvedPYTaFVklLpl------nDCRYALYDatY

Unc60A      TcsRvGA-GTSKMDKIIFLQiCPDgAsIKKKMVYASSaaAIKKSLGTGkI
Unc60B      TvqRqGAEGTStLnKVIFVQYCPDnAPVRRRMLYASSvrALKaSLGlesL
lily        v------TDencqksKIFFIsWsPDtSrVRsKMLYASTkdrFKreLD-Giq
protozoan   q------vDGg-qrnKItFIlWaPDsAPIKsKMMYTSTkdSIKKkL-vGiq
human       e------TkeS-KkEdLVFIfWaPEsAPLKsKMIYASSkdAIKKkL-TGik
chick       e------TkeS-KkEdLVFIfWaPEsAPLKsKMIYASSkdAIKKkf-TGik

Unc60A      --LQfQvSDE-SEMShkELLnKLGeKygDh------
Unc60B      --FQVQa----SEMSd--LdeKSV-KS-DlMsnqrI
lily        veLQatdp---SEMSm-DII-KAraf----------
protozoan   veVQatdAaEiSEdAvsErakKdVk----------
human       heLQaNcyEEvkDrctlae--KLGgsAvisLegkpL
chick       heWQVNglDDikDrStlge--KLGgnvvvsLegkpL
```

Fig. 8. Comparison of the Unc-60 proteins to actin depolymerizing factors. The BLAST algorithm was used to search a number of databanks for similarities to either Unc-60A or Unc-60B (see Materials and methods). Significant similarities to actin-depolymerizing factors (ADFs) were discovered. The Unc-60A and Unc-60B proteins are indicated. Comparisons are to *Lilium longiflorum* (lily; S. Kim, Y. Kim, G. An, personal communication), actophorin from *Acanthamoeba castellanii* (protozoan; S. Quirk, S.K. Maciver, J. VanDamme, personal communication), human cofilin (human; Ogawa et al. 1990) and chick cofilin isolated from skeletal muscle (chick; Abe et al. 1990). Other known members of this family not shown in this Figure include *Saccharomyces cerevisiae* COF1 (Moon et al. 1993) and ABP1 (Drubin et al. 1990), chick brain destrin (Adams et al. 1990), destrin from chick skeletal muscle (Abe et al. 1990), porcine brain cofilin (Matsuzaki et al. 1988), porcine brain destrin (Moriyama et al. 1990b), and mouse cofilin (Moriyama et al. 1990a). ADF residues identical to residues in either the Unc-60A or Unc-60B proteins are in *bold capitals*. Conservative replacements are indicated by *plain capitals*. Non-conservative replacements are indicated by *lightface lower case lettering*. Gaps have been introduced to maximize the similarity to either the Unc-60A or Unc-60B proteins

database revealed that sequences in the first third of this intron have high degrees of nucleotide identity to other *C. elegans* sequences, including sequences reported from two cDNAs. Given the repetitive hybridization patterns seen with the Southern blot results described above, the presence of repetitive DNA in our genomic sequence was not surprising. Furthermore, based on the repetitive cross-species hybridization results, these sequences could be conserved in *C. briggsae*. The function, if any, of these non-coding sequences is not known.

The unc-60 *gene encodes two putative actin-binding proteins*

Transcripts A and B share only one exon, but their respective proteins (designated Unc-60A and Unc-60B) are homologous. There is 38% identity between the 152 amino acids of Unc-60B and the Unc-60A protein (Fig. 8). It is possible that the two transcripts originated from a duplication of the final four exons of one of the transcripts. Allowing for some drift (a small number of nucleotides) in the location of introns, three of the introns are in the same location in both transcripts A and B. In addition, each transcript has an intron in a unique location. This could be accounted for if the original gene had five introns and one was subsequently lost from each transcript after the duplication event.

Both transcripts encode proteins that are homologous to the cofilin and destrin family of actin-binding proteins first identified in vertebrate systems (Fig. 8). More recently, cofilin or destrin homologs have been identified in plants (*Brassica napa* and *Lilium longiflorum*) and the amoeba *Acanthamoeba castellanii*. There is also weak homology to the *Saccharomyces cerevisiae* actin binding protein ABP1 (data not shown; Drubin et al. 1990). In vertebrates, the difference between the cofilin and destrin proteins was first defined by their in vitro biochemical activities (see Discussion), but they can also be distinguished by amino acid sequence (see references in Fig. 8). The plant proteins and amoeba sequences are known only as actin depolymerizing factors. The degree of identity between the *C. elegans* Unc-60A amino acid sequence and the cofilins is 28%, and between Unc-60A and the plant sequences is 40%. Identity between the Unc-60B amino acid sequence and the plant and amoeba sequences is approximately 32%.

Discussion

The *unc-60* gene is essential for the correct positioning of the thin filaments in *C. elegans* body wall muscle (Waterston et al. 1980). In Unc-60 mutants, most of the actin-containing thin filaments aggregate at the ends of the body wall muscle cell. The myosin-containing thick fila-

ments are left relatively intact. In this paper, we report the cloning and sequencing of the *unc-60* gene and, consistent with the mutant phenotype, we show the predicted protein is homologous to a family of actin binding proteins found in plants and animals.

We have shown by germline transformation rescue experiments that the cosmid F53E2 carries the *unc-60* gene. Molecular analysis of the *unc-60* allele *s1586* has defined the coding region corresponding to the *unc-60* locus. *s1586* is a small deletion involving a single gene. This gene produces at least two transcripts, which we have named *unc-60A* and *unc-60B*. Although the two transcripts share only a single exon encoding the initiator methionine, they encode homologous proteins. While we do not know which transcript is required for muscle function, it is probably not *unc-60A* for two reasons. Firstly, clones containing *unc-60A* did not rescue the Unc-60 phenotype in transformation experiments. Secondly, sequencing of four *unc-60* mutant alleles has failed to detect changes in *unc-60A* (K. McKim and C. Matheson, unpublished results). Therefore the second protein, encoded by *unc-60B,* may be required for the muscle function of this locus.

A putative actin-depolymerizing factor is required for the development of thin filaments in C. elegans *muscle*

Analysis of the DNA sequence indicates that both proteins encoded by the *unc-60* locus are *C. elegans* homologs of the cofilin and destrin family (also known as Actin Depolymerizing Factor) of actin-binding proteins (Fig. 8). This is a phylogenetically diverse family of proteins, with members identified in animals, plants and protozoans. These proteins have been most extensively studied in vertebrates, where it has been shown that cofilin and destrin bind to actin filaments and monomers and depolymerize actin filaments. The activity of cofilin is pH dependent. At high pH, cofilin depolymerizes actin filaments, but at low pH, it promotes polymerization (Yonezawa et al. 1985). In addition, cofilin is associated with the actin rods induced by treatment of cultured cells with heat shock or dimethylsulphoxide (Nishida et al. 1987). In contrast, destrin depolymerizes actin filaments independently of pH.

The level of amino acid sequence identity within this group of proteins ranges from 28% to 40%. For example, the *C. elegans* Unc-60A sequence is about 28% identical to the vertebrate cofilin proteins and 40% identical to the plant actin depolymerizing factors. This low level of sequence identity is also seen between other members of this family; sequence comparisons between plant and amoeba or vertebrate show 35–45% identity.

Several lines of evidence suggest that, despite the extensive sequence divergence, these proteins are homologs. First, many of the amino acid differences between the proteins involve conservative changes (Fig. 8). Second, the amino acid identities are not scattered throughout the proteins, but are concentrated at four islands of homology. With respect to the Unc-60A sequence, conserved domains can be found commencing at amino acid positions 1 (MSSGVMVDP), 27 (YIIFKID), 80 (CRYAVFDF) and 117 (KKKMVYASS). The last of these domains has been shown to be the actin-binding site of the vertebrate cofilin sequence. The two lysine residues of the human sequence have been shown to cross-link to actin (Yonezawa et al. 1991a, b).

Within each phylogenetic group, sequence identity is high. For example, there is greater than 80% amino acid identity within the vertebrate cofilin or destrin groups, greater than 70% identity between destrin and cofilin (see references in Fig. 8), and the two plant proteins are 77% identical. In contrast to the high degree of similarity between the sequences within the plant or vertebrate groups, the two *unc-60* proteins have only 38% sequence identity. This finding suggests that the Unc-60 protein sequences have diverged to accommodate different functions.

In addition to the sequence, three structural patterns of the vertebrate transcripts and *unc-60A* are also conserved, further supporting the hypothesis that we have identified the nematode homologs of cofilin or destrin. First, the proteins are small, in the range of 138 to 166 amino acids long. All destrins and Unc-60A contain 165 amino acids, while all cofilins contain 166 amino acids. Second, the four conserved domains have the same relative location in each protein (Fig. 8). Third, the transcripts of all the vertebrate cofilin and destrin genes and *unc-60A* have 3′ untranslated regions which, at approximately 700 nucleotides in length, are longer than the coding regions (see Fig. 8 for references). The large untranslated region could be involved in the regulation of translation, as described for the *fem-3* (Ahringer and Kimble 1991) and *tra-2* genes of *C. elegans* (Kuwabara et al. 1992). Alternatively, the large 3′ UTR could be involved in the cellular localization of the transcript, as with beta actin (Sundell and Singer 1991) and *bicoid* mRNAs (MacDonald and Struhl 1988).

As noted above, the differences between *unc-60A* and *unc-60B* suggest that each protein has evolved for a unique function. These differences include the following. First, Unc-60B retains the conserved sequence motifs found in Unc-60A, but with some interesting differences. The YASS sequence of the actin-binding domain is conserved in both *C. elegans* proteins, but the upstream lysines are found only in cofilin and Unc-60A. Furthermore, the sequence CRYA is conserved in all the proteins, but the aspartic acid residue immediately upstream is present in Unc-60A, vertebrate and amoeba proteins, but absent in Unc-60B. The changes, however, are usually conservative. For example, the three lysines in Unc-60A are replaced by three arginines in Unc-60B. Second, *unc-60B* has a shorter open reading frame and 3′ untranslated region than *unc-60A*. Finally, *unc-60B* is expressed at a lower level than *unc-60A* as judged from Northern analysis and the representation of cDNA clones in the library we screened. Correlating with the lower level of expression is the larger size of the *unc-60B* introns. *unc-60A* contains characteristically small introns for *C. elegans,* in the range of 53 to 153 nucleotides. *unc-60B,* on the other hand, has introns characteristic of the largest in *C. elegans,* ranging in size from 370 to 1165 nucleo-

tides. This does not include the largest intron, the 1680 nucleotide first intron which contains most of the *unc-60A* transcript. The significance of the correlation between increased expression and small introns is not known.

The vertebrate cofilins and destrins have been isolated from non-muscle cells and from skeletal muscle (Abe et al. 1989, 1990). Perhaps the more divergent Unc-60B protein has evolved for a specialized function in nematode muscle cells (see below). This is not the case in the chick; identical proteins were isolated from brain and skeletal muscle (Abe et al. 1990; Adams et al. 1990). It is also probable that Unc-60 is not the only *C. elegans* actin-binding protein involved in actin filament polymerization. As many as thirteen different actin-binding proteins have been identified (Vandekerckhove and Vancompernalle 1992). Candidates for additional nematode actin-binding proteins are the uncharactized products of the *unc-78* and *unc-94* genes, as mutations in these genes disrupt I-band organization (Waterston et al. 1980; Zengel and Epstein 1980). Another potential nematode actin-binding protein is the product of *sup-12,* mutations in which act as recessive suppressors of *unc-60* mutations (R. Francis and R. Waterston, personal communication; McKim et al. 1988). There is little sequence identity between the different families, but most bind to the amino terminus of the actin peptide. The formation of functional *C. elegans* thin filaments probably results from the combined activities of several actin-binding proteins.

A model for unc-60 *function*

The observation that *unc-60* encodes a cofilin or destrin-like actin-binding protein is entirely consistent with the phenotype of *unc-60* mutants. Since thin filaments form in Unc-60 mutant muscle, the *unc-60* protein is not required for polymerization. More probably, the improper aggregation of thin filaments in the muscle cells of *unc-60* mutants (Waterston et al. 1980) can be explained if the *unc-60* gene product is required to regulate the polymerization of actin. We propose that if the rate of actin polymerization is not controlled by a depolymerizing factor, the thin filaments will improperly aggregate. It is also possible that Unc-60 does not depolymerize actin filaments, but binds and sequesters monomeric actin, thereby eliminating it from the actin pool available for polymerization.

This model is supported by the observation that the phenotype characteristic of *unc-60* mutants, the aggregation of thin filaments at the ends of the body wall muscle cells, is observed in another mutant defective in actin polymerization. Dominant antimorphic mutations in the *act-1* or *act-3 C. elegans* actin genes have this phenotype (Landel et al. 1984; Waterston et al. 1984). In this case, the mutant peptides probably interfere with the polymerization of peptides derived from the other three actin-encoding loci. Thus, there is precedence for the aggregation of thin filaments at the ends of the body wall muscle cells being caused by abnormal actin polymerization. Furthermore, Shimizu and Obinata (1986) proposed that actin polymerization in developing skeletal muscle is controlled or suppressed by regulatory proteins. Thus, in Unc-60 muscle cells, the absence of depolymerizing activity from the *unc-60* locus could result in unregulated actin polymerization, which could then result in the improper aggregation of thin filaments. That Unc-60 might regulate actin polymerization through a depolymerizing activity is similar to the proposal for the function of the Unc-60 homolog actophorin in *Acanthamoeba castellani.* Maciver et al. (1991a, b) showed that actophorin depolymerizing activity is required in conjunction with alpha-actinin to stimulate bundling of actin filaments in vitro. Without the depolymerizing activity of actophorin, the actin filaments grow to an abnormally long length and are not bundled by alpha-actinin.

In addition to the muscle-specific activity, the *unc-60* locus encodes an essential function required for the viability of the organism. Individuals homozygous for *unc-60(s1586)* are arrested in development at a mid-larval stage. This phenotype is distinct from the Pat phenotype described for other mutations affecting muscle, including alleles of *unc-52* (Gilchrist and Moerman 1992), *deb-1* (Barstead and Waterston 1991), *unc-45* (Venolia and Waterston 1990) and *myo-3* (Waterston 1989). The Pat phenotype is characterized by a failure of the animal to elongate past the two-fold stage, and arrest at the L1 larval stage (B. Williams, personal communication). Our molecular analysis of the lethal *unc-60* allele, *s1586,* shows that the essential function of this locus could be provided by either one or both of the *unc-60* transcripts, since *s1586* deletes portions of *unc-60A* and *unc-60B.* Unc-60 is not the only cofilin homolog demonstrated to have an essential in vivo function. An essential in vivo role for the *Saccharomyces cerevisiae* homolog of mammalian cofilin has been reported; complete disruptions of the *COF1* gene in yeast result in lethality (Moon et al. 1993). As cofilin homologs from two systems have been shown to encode essential functions, it is likely that other actin-binding proteins could also be identified by lethal mutation in these systems. Unc-60 is the first example of a member of the cofilin/destrin family of actin-depolymerizing proteins described in a metazoan amenable to classical genetic analysis. Utilizing mutational analysis, possible with the powerful genetics of the *C. elegans* system, further studies of the *unc-60* gene products will be valuable in elucidating the role of cofilin and destrin in actin filament assembly.

Acknowledgments. We are grateful to Ann Rose, Raja Rosenbluth, Sharon Gorski and Erin Gilchrist for helpful discussions and comments on the manuscript, to Sarah Baldry for technical assistance with the ABI sequencer and oligonucleotide primer synthesis, and to D. Collins for her assistance with a Figure. We thank Shiv Prasad for his gifts of total RNA and Northern blots. We also wish to thank Jim McGhee for providing us with unpublished information on the location of *ges-1,* which greatly assisted us in the cloning of *unc-60.* Some nematode strains used in this work were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR). This work was supported by a Natural Sciences and Engineering Research Council (NSERC) postdoctoral award to K. McK. and a Muscular Dystrophy Association of Canada grant to D.L.B. Funds for the maintenance of the stocks used in this study were provided by a NSERC grant to D.L.B.

References

Abe H, Obinata S, Obinata T (1989) A cofilin-like protein is involved in the regulation of actin assembly in developing skeletal muscle. J Biol Chem 106:696–702

Abe H, Endo T, Yamamoto K, Obinata T (1990) Sequence of cDNAs encoding actin depolymerizing factor and cofilin embryonic chicken skeletal muscle: two functionally distinct actin-regulatory proteins exhibit high structural homology. Biochemistry 29:7420–7425

Adams ME, Minamide LS, Duester G, Bamburg JR (1990) Nucleotide sequence and expression of a cDNA encoding chick brain actin depolymerizing factor. Biochemistry 29:7414–7420

Ahringer J, Kimble J (1991) Control of the sperm-oocyte switch in *Caenorhabditis elegans* hermaphrodites by the *fem-3* 3′ untranslated region. Nature 349:346–348

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. J Mol Biol 215:403–410

Barstead RJ, Waterston RH (1989) The basal component of the nematode dense-body is vinculin. J Biol Chem 264:10177–10185

Barstead RJ, Waterston RH (1991) Vinculin is essential for muscle function in the nematode. J Cell Biol 114:715–724

Benian GM, Kiff JE, Neckelmann N, Moerman DG, Waterston RH (1989) Sequence of an unusually large protein implicated in regulation of myosin activity in *C. elegans*. Nature 342:45–50

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

Cabot EL, Beckenbach AT (1989) Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comp Appl Biol Sci 5:233–234

Coulson A, Sulston J, Brenner S, Karn J (1986) Towards a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc Natl Acad Sci 83:7821–7825

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

Drubin DG, Mulholland J, Zhu Z, Botstein D (1990) Homology of a yeast actin-binding protein to signal transduction proteins and myosin-I. Nature 343:288–290

Emmons SE, Klass MR, Hirsh D (1979) Analysis of the constancy of DNA sequences during development and evolution of the nematode *C. elegans*. Proc Natl Acad Sci USA 76:1333–1337

Emmons SW (1988) The genome. In: Wood WB (ed) The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp 47–79

Feinberg AP, Vogelstein B (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267

Fire A (1986) Integrative transformation of *Caenorhabditis elegans*. EMBO J 5:2673–2680

Gilchrist EJ, Moerman DG (1992) Mutations in the *sup-38* gene of *Caenorhabditis elegans* suppress muscle-attachment defects in *unc-52* mutants. Genetics 132:431–442

Heine U, Blumenthal T (1986) Characterization of regions of the *Caenorhabditis elegans* X chromosome containing vitellogenin genes. J Mol Biol 188:310–312

Henikoff S (1987) Unidirectional digestion with Exonuclease III in DNA sequence analysis. Methods Enzymol 155:156–165

Johnsen RC, Baillie DL (1991) Genetic analysis of a major segment [LGV(left)] of the genome of *Caenorhabditis elegans*. Genetics 129:735–752

Kennedy BP, Aamodt EJ, Allen FL, Chung MA, Heschl MFP, McGhee JD (1993) The gut esterase gene *(ges-1)* from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. J Mol Biol 229:890–908

Kramer JM, French RP, Park EC, Johnson JJ (1990) The *Caenorhabditis elegans rol-6* gene, which interacts with the *sqt-1* collagen gene to determine organismal morphology, encodes a collagen. Mol Cel Biol 10:2081–2089

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

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

Kuwabara PE, Okkema PG, Kimble J (1992) *Tra-2* encodes a membrane protein and may mediate cell communication in the *Caenorhabditis elegans* sex determination pathway. Mol Biol Cell 3:461–473

Landel CP, Krause M, Waterston RH, Hirsh D (1984) DNA rearrangements of the actin gene cluster in *Caenorhabditis elegans* accompany reversion of 3 muscle mutants. J Mol Biol 180:497–513

MacDonald PM, Struhl G (1988) Cis-acting sequences responsible for anterior localization of *bicoid* mRNA in *Drosophila* embryos. Nature 366:595–598

Maciver SK, Zot HG, Pollard TD (1991a) Characterization of actin filament severing by actophorin from *Acanthamoeba castellanii*. J Cell Biol 115:1611–1620

Maciver SK, Wachsstock DH, Schwarz WH, Pollard TD (1991b) The actin filament-severing protein actophorin promotes the formation of rigid bundles of actin filaments crosslinked with alpha-actinin. J Cell Biol 115:1621–1628

Matsuzaki F, Matsumoto S, Yahara I, Yonezawa N, Nishida E, Sakai H (1988) Cloning and characterization of porcine brain cofilin cDNA. J Biol Chem 263:11564–11568

McGhee JD, Cottrell DA (1986) The major gut esterase locus in the nematode *Caenorhabditis elegans*. Mol Gen Genet 202:30–34

McGhee JD, Birchall JC, Chung MA, Cottrell DA, Edgar LG, Svendsen PC, Ferrari DC (1990) Production of null mutants in the major intestinal esterase gene *(ges-1)* of the nematode *Caenorhabditis elegans*. Genetics 125:505–514

McKim KS, Heschl MFP, Rosenbluth RE, Baillie DL (1988) Genetic organization of the *unc-60* region in *Caenorhabditis elegans*. Genetics 118:49–59

Mello CC, Kramer JM, Stinchcomb D, Ambros V (1991) Efficient gene transfer in *C. elegans:* extrachromosomal maintenance and integration of transforming sequences. EMBO J 10:3959–3970

Moon AL, Janmey PA, Louie KA, Drubin DG (1993) Cofilin is an essential component of the yeast cortical cytoskeleton. J Cell Biol 120:421–435

Moriyama K, Matsumoto S, Nishida E, Sakai H, Yahara I (1990a) Nucleotide sequence of mouse cofilin cDNA. Nucleic Acids Res 18:3053

Moriyama K, Matsumoto S, Nishida E, Sakai H, Yahara I (1990b) Destrin, a mammalian actin-depolymerizing protein, is closely related to cofilin. Cloning and expression of porcine brain destrin cDNA. J Biol Chem 265:5768–5773

Nishida E, Iida K, Yonezawa N, Koyasu S, Yahara I, Sakai H (1987) Cofilin is a component of intranuclear and cytoplasmic actin rods induced in cultured cells. Proc Natl Acad Sci USA 84:5262–5266

Ogawa K, Tashima M, Yumoto Y, Okuda T, Sawada H, Okuma M, Maruyama Y (1990) Coding sequence of human placental cofilin cDNA. Nucleic Acid Res 18:7169

Prasad SS, Baillie DL (1989) Evolutionarily conserved coding sequences in the *dpy-20-unc-22* region of *Caenorhabditis elegans*. Genomics 5:185–198

Rogalski TM, Williams BD, Mullen GP, Moerman DG (1993) The products of the *unc-52* gene in *Caenorhabditis elegans* are homologous to the core protein of the mammalian basement membrane heparin sulfate proteoglycan. Genes Dev 7:1471–1484

Rosenbluth RE, Rogalski TM, Johnsen RC, Addison LM, Baillie DL (1988) Genomic organization in *Caenorhabditis elegans*: deficiency mapping on linkage group V(left). Genet Res 52:105–118

Rychlik W, Rhoads RE (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. Nucleic Acids Res 17:8543–51

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. 2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Sanger F, Coulson AR, Barrell BG, Smith AJH, Roe BA (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Snutch TP (1984) A molecular and genetic analysis of the heat shock response of *Caenorhabditis elegans*. PhD thesis, Simon Fraser University, Burnaby, B.C., Canada.

Shimizu N, Obinata T (1986) Actin concentration and monomer-polymer ratio in developing chicken skeletal muscle. J Biochem (Tokyo) 99:751–759

Spence AM, Coulson A, Hodgkin J (1990) The product of *fem-1*, a nematode sex-determining gene, contains a motif found in the cell cycle control proteins and receptors for cell-cell interactions. Cell 60:981–990

Starr T, Howell AM, McDowall J, Peters K, Rose AM (1989) Isolation and mapping of DNA probes within the linkage group I gene cluster of *Caenorhabditis elegans*. Genome 32:365–372

Stewart HI, Rosenbluth RE, Baillie DL (1991) Most ultraviolet irradiation-induced mutations in the nematode *Caenorhabditis elegans* are chromosomal rearrangements. Mutat Res 249:37–54

Sundell CL, Singer RH (1991) Requirement of microfilaments in sorting of actin messenger RNA. Science 253:1275–1277

Vandekerckhove L, Vancompernalle K (1992). Structural relationships of actin-binding proteins. Curr Opin Cell Biol 4:36–42

Venolia L, Waterston RH (1990) The *unc-45* gene of *Caenorhabditis elegans* is an essential muscle-affecting gene with maternal expression. Genetics 126:345–353

Waterston RH, Thomson JN, Brenner S (1980) Mutants with altered muscle structure in *Caenorhabditis elegans*. Dev Biol 77:271–302

Waterston RH, Hirsh D, Lane TR (1984) Dominant mutations affecting muscle structure in *Caenorhabditis elegans* that map near the actin gene cluster. J Mol Biol 180:473–496

Waterston RH (1988) Muscle. In: Wood WB (ed) The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York pp 281–335

Waterston RH (1989) The minor myosin heavy chain, mhcA, of *Caenorhabditis elegans* is necessary for the initiation of thick filament assembly. EMBO J 8:3429–3436

Yonezawa N, Nishida E, Sakai H (1985) pH control of actin polymerization by cofilin. J Biol Chem 260:14410–14412

Yonezawa N, Nishida E, Iida K, Kumagai H, Yahara I, Sakai H (1991a) Inhibition of actin polymerization by a synthetic dodecapeptide patterned on the sequence around the actin-binding site of cofilin. J Biol Chem 266:10485–10489

Yonezawa N, Homma Y, Yahara I, Sakai H, Nishida E (1991b) A short sequence responsible for both phosphoinositide binding and actin binding activities of cofilin. J Biol Chem 266:17218–17221

Zengel JM, Epstein HF (1980) Identification of genetic elements associated with muscle structure in the nematode *Caenorhabditis elegans*. Cell Motil 1:73–97