

The *Caenorhabditis elegans* Gene *unc-89*, Required for Muscle M-line Assembly, Encodes a Giant Modular Protein Composed of Ig and Signal Transduction Domains

Guy M. Benian,* Tina L. Tinley,* Xuexin Tang,* and Mark Borodovsky‡

*Department of Pathology, Emory University, Atlanta, Georgia 30322; and ‡School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332

Abstract. Mutations in the *Caenorhabditis elegans* gene *unc-89* result in nematodes having disorganized muscle structure in which thick filaments are not organized into A-bands, and there are no M-lines (Waterston, R.H., J.N. Thomson, and S. Brenner. 1980. *Dev. Biol.* 77:271–302). Beginning with a partial cDNA from the *C. elegans* sequencing project, we have cloned and sequenced the *unc-89* gene. An *unc-89* allele, *st515*, was found to contain an 84-bp deletion and a 10-bp duplication, resulting in an in-frame stop codon within predicted *unc-89* coding sequence. Analysis of the complete coding sequence for *unc-89* predicts a novel 6,632–amino acid polypeptide consisting of sequence motifs which have been implicated in protein–protein interactions. UNC-89 begins with 67 residues of unique sequence, SH3, dbl/CDC24, and PH domains, 7 immunoglobulin (Ig) domains, a putative KSP-containing

multiphosphorylation domain, and ends with 46 Ig domains. A polyclonal antiserum raised to a portion of *unc-89* encoded sequence reacts to a twitchin-sized polypeptide from wild type, but truncated polypeptides from *st515* and from the amber allele *e2338*. By immunofluorescent microscopy, this antiserum localizes to the middle of A-bands, consistent with UNC-89 being a structural component of the M-line. Previous studies indicate that myofilament lattice assembly begins with positional cues laid down in the basement membrane and muscle cell membrane (Williams, B.D., and R.H. Waterston. 1994. *J. Cell Biol.* 124:475–490; Hresko, M.C., B.D. Williams, and R.H. Waterston. 1994. *J. Cell Biol.* 124:491–506). We propose that the intracellular protein UNC-89 responds to these signals, localizes, and then participates in assembling an M-line.

IN striated muscle, the bare zone is the central part of the A-band in which bipolar thick filaments are free of myosin heads. In the center of the bare zone lies the electron dense M-line, also known as the M-band or M-disc, which is 75–100 nm wide and is believed to anchor thick filaments and help maintain them in proper register. From electron microscopic examinations, models for the complex lattice-work that constitutes the M-line in vertebrates, especially the frog, have been developed (Knappes and Carlsen, 1968; Luther and Squire, 1978), but the assignment of every known component to structures is not well understood.

Five non-myosin proteins are known to reside in the vertebrate M-line: (a) MM-creatine kinase (43 kD; Strehler et al., 1983), (b) the COOH-terminal portion of the giant 3×10^6 -D polypeptide called titin, also known as connectin (Gautel et al., 1993; Maruyama, 1994; Labeit and Kolmerer, 1995b), (c) M-protein (165 kD; Noguchi et al., 1992), (d) myomesin (190 kD; Vinkemeier et al., 1993), and (e) skelemin (195 kD; Price and Gomer, 1993). Titin/

connectin copurifies with M-protein and myomesin (Nave et al., 1989). The last four of these proteins are members of the intracellular muscle branch of the immunoglobulin (Ig) superfamily and consist primarily of multiple copies of Ig and fibronectin type III (FnIII)¹ domains. The founding member of this branch of the Ig superfamily is *Caenorhabditis elegans* twitchin, a 753,494-D polypeptide (Benian et al., 1989, 1993) encoded by the gene *unc-22*, and localized to muscle A-bands (Moerman et al., 1988). Twitchin consists of a single protein kinase domain, 31 FnIII domains, and 30 Ig domains. Other members of this intracellular branch of the Ig superfamily include insect projectin (Ayme-Southgate et al., 1991, 1995; Fyrberg et al., 1992), and the vertebrate proteins smooth muscle (Olson et al., 1990) and non-muscle (Shoemaker et al., 1990) myosin light chain kinases, telokin (Gallagher and Herring, 1991), C-protein (Einheber and Fischman, 1990), 86-kD protein (Vaughan et al., 1993), kettin (Lakey et al., 1993), and the above-mentioned M-line proteins. The Ig and FnIII domains are likely to mediate interaction with myosin and other proteins, as has been shown for titin/con-

Address correspondence to Guy M. Benian, Dept. of Pathology, Woodruff Memorial Building, Emory University, Atlanta, Georgia 30322. Tel.: (404) 727-5953. Fax: (404) 727-8540. E-mail: pathgb@bimcore.emory.edu

1. *Abbreviations used in this paper:* FnIII, fibronectin type III; KSP, lysine serine proline.

nectin (A-band portion; Labeit et al., 1992), telokin (Shirinsky et al., 1993), C-protein (Okagaki et al., 1993), and ketin (Lakey et al., 1993).

The nematode *C. elegans* is an excellent organism in which to study muscle because of its simple anatomy, simple genome and sophisticated genetics. The obliquely-striated body wall muscle, which permits the animal to move, as first described by Waterston et al. (1980), has M-lines that appear on electron microscopy as lightly stained amorphous material up to 100 nm in width and $\sim 1 \mu\text{m}$ in depth. Similar to the dense body (the nematode's Z line analogue), the M-line is anchored at the muscle cell plasma membrane at which are small projections extending into the basement membrane. The basement membrane separates the muscle cell from the overlying hypodermis and cuticle. The molecular composition of the nematode M-line is not known. Of all the ~ 50 genes that are important for the assembly and function of muscle in *C. elegans* (Waterston, 1988; Williams and Waterston, 1994), the only one which has so far been described to affect the M-line is *unc-89*. One allele of *unc-89*, *e1460*, had been described by Waterston et al. (1980): these animals move as fast as wild type, but are thinner and more transparent. In body wall muscle, polarized light microscopy revealed much disorganization with only suggestions of A-bands and dense bodies. Electron microscopy showed a normal number of thick filaments, but these were not organized into A-bands and most interestingly, there were no M-lines. In pharyngeal muscle, polarized light revealed decreased birefringence in some muscle groups, and misplacement of other groups of fibers.

In this report, we show that *unc-89* encodes a 732-kD polypeptide consisting of sequence motifs that have been implicated in protein-protein interactions. The UNC-89 protein begins at the NH₂ terminus with 67 residues of unique sequence, followed by an SH3 domain, a dbp/CDC24 domain, a PH domain, 7 Ig domains, then 645 residues containing 44 copies of the amino acid sequence lysine serine proline (KSP), followed by 46 Ig domains. An antiserum raised to expressed *unc-89* sequences reacts on Western blots to a twitchin-sized polypeptide from wild type, but truncated polypeptides from mutator-induced and amber alleles. By immunofluorescent microscopy this antiserum stains both body wall and pharyngeal muscles in wild type, but in *unc-89* mutants staining is absent or of reduced intensity. UNC-89 is localized to the middle of A-bands, consistent with UNC-89 being an M-line component. UNC-89 is the first muscle architectural protein that is likely to be interacting with a G-protein-mediated signal transduction pathway. Taking into account that myofilament lattice assembly is likely to begin at the extracellular matrix and cell membrane, and then proceed inwards (Hresko et al., 1994; Williams and Waterston, 1994), we propose a mechanism by which the intracellular protein UNC-89 responds to these signals, and then participates in the assembly of the M-line.

Materials and Methods

Nematode Strains

C. elegans strains were obtained from the *C. elegans* stock center, R.H.

Waterston, H.F. Epstein, L. Avery, and D.L. Riddle. Wild-type worms were *C. elegans* variety Bristol, strain N2. The following *unc-89* alleles were tested for polymorphisms with cosmids C34E1 and K08D2: *ad539*, *e1460*, *e2338*, *st79*, *st85*, *st86*, *st151*, *su75*, *su78*, *su227*, and *su240*. *unc-89(st151)* was originally isolated from a mutator strain (and thus likely to be transposon-induced) and then out-crossed 8 \times to N2 by S. Rioux in R.H. Waterston's laboratory (Washington University). *st85* and *e2338* are known to be amber (UAG) chain termination mutations because their phenotypes are suppressible by the amber suppressor mutation *sup-7*. *eDp23*; *unc-54(e190)* is null for myoB (encoded by the *unc-54* gene) but can produce thick filaments because it has a duplication of the *myo-3* gene and overproduces myoA (Riddle and Brenner, 1978; Maruyama et al., 1989).

Northern and Southern Blots

Total RNA was isolated from wild type and *unc-22(ct37)* by a slight modification of published procedures (Chirgwin et al., 1979). Northern blots were prepared and hybridized as described previously (Moerman et al., 1988). Genomic DNA was prepared from wild type and each of the above-mentioned *unc-89* strains. Separate aliquots were digested with eight restriction enzymes, and probed with cosmids C34E1 and K08D2 on Southern blots as described in Moerman et al. (1988).

DNA Sequencing

The entire sequence of the cosmid C34E1 was determined by sequencing clones from a shotgun library in mp10M13. Cosmid mini-prep DNA was sheared by sonication, separated on an agarose gel, and fragments larger than 1.1 kb were purified by GeneClean (BIO 101 Inc., La Jolla, CA). This DNA was end-repaired with sequential treatment with T4 DNA polymerase (+dNTPs) and Klenow fragment, phosphorylated with T4 polynucleotide kinase, ligated into SmaI-cut and phosphatased mp10M13 (Amersham Corp., Arlington Heights, IL) and transformed into competent JM101 cells. Single-stranded DNA from over 800 independent clones was prepared and sequenced (~ 400 bp each) with the M13 universal primer using Taq DyeDeoxy Terminator Cycle Sequencing Kits (Applied Biosystems, Inc., Foster City, CA) on an Applied Biosystems 373A automated fluorescent DNA sequencer. Each ABI raw data file was transferred to a SUN computer, bases were called and M13 vector sequences were clipped-off with AUTOTED (Wilson et al., 1994), and assembled into contigs and edited with XBAP (Dear and Staden, 1991). To obtain one contig for the entire cosmid sequence on both strands, 31 oligonucleotides were used as sequencing primers on selected shotgun clones.

Exon Prediction

The GeneMark program (Borodovsky and McIninch, 1993) was used for predicting exons in the C34E1 sequence. The key element of the GeneMark method is an inhomogeneous (three-periodical) Markov model that provides an accurate statistical description of a protein-coding sequence (Borodovsky et al., 1986; Tavares and Song, 1989; Kleffe and Borodovsky, 1992). A given DNA fragment is identified as residing in a region that is either (a) protein-coding, (b) non-coding but complementary to coding sequence (gene shadow), or (c) totally non-coding. Parameters of the Markov model for DNA sequence of a particular type are determined from training sets of experimentally identified sequences of the same type. A fragment S, of newly sequenced DNA, is associated with one of several possible sequence types (models, $k = 1, \dots, K$) using a probabilistic measure defined by Bayes theorem:

$$P(\text{model } k \mid \text{sequence } S) = \frac{P(\text{sequence } S \mid \text{model } k) P(\text{model } k)}{\sum_{k=1}^K p(\text{sequence } S \mid \text{model } k) P(\text{model } k)}$$

The above expression is used to identify possible reading frames in coding sequence ($k = 7$). Then, seven probability values P_i , $i = 1, \dots, 7$ describe the likelihoods that the sequence S belongs to one of seven mutually exclusive states: six states of protein-coding in six different frames and one non-coding state. If one of P_i , $i = 1, \dots, 6$ is greater than threshold 0.5, then the DNA fragment S (or its complement) is identified as protein-coding in a proper reading frame. If P_7 is larger than 0.5 or if none of P_i , $i = 1, \dots, 6$ is larger than 0.5, then S (and its complement) is identified as non-coding. Long sequences are analyzed using a sliding window technique and parsed into coding and non-coding regions. The GeneMark method

has proven to be accurate enough to detect previously unrecognized genes in GenBank sequences (Borodovsky et al., 1994a,b).

The application of GeneMark to *C. elegans* required some modifications. Since the non-homogeneity of the *C. elegans* genome hampers the accuracy of gene identification, a large set of known *C. elegans* DNA sequences was divided into two sets with "high" and "low" GC content using a cutoff value of 46%, the average GC content for the *C. elegans* genome. There were 153 spliced genes (227,010 bp) and 155 non-coding regions (130,950 bp) that fell into the high GC range as well as 197 spliced genes (294,615 bp) and 338 non-coding regions (363,337 bp) that fell into the low GC range. These data were sufficient to derive parameters of Markov models up to fourth order for each GC content range. When these models were employed in the program, the correct prediction for true 96-bp coding fragments was observed in 86% of cases for low GC range and in 89% of cases for high GC range. For 96 bp non-coding fragments the correct identification was done in 91% of cases for the low GC range and 94% of cases for the high GC range. Since the *unc-89* sequence falls into the low GC content range, the low GC models were used in the GeneMark program. Predicted exons were translated and compared to protein sequences in the databases by the BLASTP method (Altschul et al., 1990).

Partial cDNA Clones, Reverse Transcriptase PCR, and 5' RACE

The original cDNA cm20d2 was used as a probe to screen an oligo-dT-primed cDNA library (kindly provided by R. Barstead, Oklahoma Medical Research Foundation) and two additional overlapping partial cDNAs were recovered, 1aA2 and 17b-2a. These clones were sequenced manually (Sanger et al., 1977) using Sequenase (Stratagene, Inc., La Jolla, CA). All three clones lack in-frame stop codons and have artifactual polyA tails probably due to annealing of oligo-dT to A-rich coding sequences often found in the AT-rich *C. elegans* genome. Comparison of these cDNA sequences to genomic sequence defined 8 exons. These encode sequence from the last 15 amino acids of Ig no. 2 until the middle of the KSP domain. cm20d2 and 17b-2a have one 66-bp exon, not found in 1aA2 (see Fig. 3 a). The remaining exons were first predicted by GeneMark and then confirmed by sequencing ~25 200-900-bp reverse transcriptase PCR products. PCR primers were chosen such that products would include the 3' end of one exon, jump across a predicted intron, and include the 5' end of the next exon. First-strand cDNA was synthesized from total nematode RNA using an antisense oligonucleotide primer, Superscript II reverse transcriptase, and reagents supplied in a Superscript Pre-amplification System kit, according to the manufacturer (Life Technologies, Inc., Grand Island, NY). One-tenth of this material was used in a PCR that included some of the same primer used to generate the first-strand cDNA together with a sense oligo-primer and Taq DNA polymerase (Promega Corp., Madison, WI). Primers were 20-23 mers with approximate T_m of 60-62°C. The PCR involved denaturation at 94°C for 1 min, annealing at 57°C for 2 min, and extension at 72°C for 1 min for a total of 40 cycles. After agarose gel purification and recovery by GeneClean, fragments were ligated into the pT7Blue(R) T-vector and transformed into competent NovaBlue *Escherichia coli* (Novagen, Inc., Madison, WI). Plasmid DNA was prepared using a Wizard kit (Promega Corp.) and sequenced either manually or by ABI machine. These sequences were compared to the genomic sequence. In some cases very small (less than 150 bp) exons were defined which had not been predicted by GeneMark. The most 5' exon predicted by GeneMark with high probability begins at position 5870 in the genomic sequence. From this point until the 3' end of coding sequence, as compared to the size of mRNA on Northern blot, and polypeptide on Western blot, we expected to have defined ~99% of the *unc-89* coding sequence. Thus, we conducted 5' RACE to determine the 5' end of the *unc-89* mRNA. This was performed using a 5' RACE System kit (Life Technologies, Inc.), making first-strand cDNA from nematode total RNA with an antisense primer beginning at position 14,574, oligo-dC tailing with terminal transferase, and then PCR using an antisense primer beginning at position 5870 and the kit's anchor primer. Whereas no product was obtained from a sample that was not tailed, the tailed sample yielded a single 450-bp fragment, which was cloned into pT7Blue vector and sequenced. An in-frame initiator methionine (position 4920) which matches a consensus for translation initiation in *C. elegans* (Benian et al., 1993), and a 5' untranslated sequence of 145 bp were inferred. To recover a cDNA for the 3' end of the *unc-89* mRNA, we screened the above-mentioned cDNA library by PCR. Approximately 10^7 phage particles were added to a PCR reaction containing a sense primer corresponding to sequence 45 bp upstream of

the 3' end of the insert in C34E1 and a primer lying 109 bp downstream of the EcoRI cloning site in pBluescript SK (Stratagene, Inc.). The resulting 1.4-kb product was gel purified and ligated into pT7Blue(R) and sequenced with the ABI machine on both strands using primers placed every ~330 bp. Unfortunately, this sequence is all coding sequence with a false 13-nt polyA tail. Therefore, the cDNA library was rescreened by hybridization with this 1.4-kb fragment. All three clones recovered had different 5' ends but ended at this same place, with false polyA tails. Thus, a different oligo-dT-primed cDNA library was screened (λ RB1 in which inserts had been cloned into the XhoI site of the vector λ ACT; library also provided by R. Barstead). The 3'-most 520 bp of the 1.4-kb fragment was used as a hybridization probe to screen ~300,000 plaques. One positive clone was obtained and called 3'11A and found to extend further 3'. Upon sequencing, it was found to contain an in-frame stop codon, 3' untranslated region of 680 bp, a polyadenylation signal, and a polyA tail. Thus, the 3'-most 1,982 bp of the *unc-89* mRNA was determined only by sequencing cDNAs and not genomic DNA.

Antiserum Production and Western Blots

A 364 residue polypeptide, corresponding to sequence including the end of Ig no. 2 through the middle of Ig no. 6, was expressed in *E. coli* as a glutathione S-transferase fusion protein. This was done by PCR using Pfu polymerase (Stratagene) with primers corresponding to sequence within cDNA cm20d2, to which were added BamHI and EcoRI sites and cloned into pGEX-2T (Smith and Johnson, 1988). The clone used for expression was verified, by sequencing, not to contain errors introduced by PCR. The GST fusion protein was expressed and purified with a glutathione-agarose batch method essentially as described by Smith and Johnson (1988) except that IPTG induction was conducted at room temperature. Milligram quantities of GST fusion protein were supplied to Spring Valley Laboratories (Sykesville, MD) for production of polyclonal rabbit antisera. From the preimmune sera of 10 rabbits, 2 rabbits were selected for immunization because they had the lowest level of antibodies cross-reacting to total protein extracts of nematodes on Western blots. Following the third series of immunizations, Western blots showed reaction to a polypeptide of ~750 kD. Removal of most of the anti-GST antibodies and affinity purification was performed as described in Benian et al. (1993). Immunoblot experiments were conducted as follows: wild-type and *unc-89* mutant strains were grown on single 10-cm Petri plates seeded with *E. coli* OP50 until the bacteria was nearly depleted. Worms were washed free from bacteria by several washes and pelletings through M9 buffer (86 mM NaCl, 22 mM KH_2PO_4 , 43 mM Na_2HPO_4 , 1 mM MgSO_4) in a microfuge. To the ~0.1 ml of packed worms was added 1 ml of 95°C 2 \times Laemmli buffer (Laemmli, 1970), vortexed for 20 s, heated at 95°C for 5 min, and vortexed again for 20 s. This mixture was then spun in a microfuge for 10 min, after which the supernatant was decanted to a fresh Eppendorf tube and stored at -70°C until used. Samples were thawed briefly at 37°C, and 10 μ l of each were run on lanes of a 1-mm thick 5% polyacrylamide SDS Laemmli gel (Laemmli, 1970) with a 3% stacking gel run in a Bio-Rad, Inc.'s minigel as follows: 50 volts for 40 min, 100 volts for 10 min, and 200 volts for 50 min. Contents of the gel were transferred to nitrocellulose membrane (Towbin et al., 1979) for 2 h at 100 constant volts with cooling in a Bio-Rad mini trans-blot apparatus. After staining with Ponceau S solution (Sigma Chem Co., St. Louis, MO), the membrane was blocked at 37°C for 4 h in 5% non-fat dry milk, 0.1% Tween-20, tris-buffered saline; this solution was also used for dilution of primary and secondary antibodies. Monoclonal MH42 (kindly provided by M. Coutu Hresko) was used at a 1:400 dilution, and the rabbit polyclonal anti-twitchin (developed to central Fn III Fn III Ig motifs of twitchin by M. Valenzuela and G.M. Benian) at 1:10,000 dilution. The secondary antisera were peroxidase-conjugated donkey anti-mouse or anti-rabbit immunoglobulin, and detection was by enhanced chemiluminescence (Amersham).

Indirect Immunofluorescence

Most of our observations were made on whole nematodes using a method originally designed by Ruvkun and Giusto (1989) and Finney and Ruvkun (1990) and modified for the study of muscle by D.M. Miller (1995). For localizing UNC-89 in the pharynx, we performed fluorescence microscopy on frozen sections, as described later in this section.

For the whole worm procedure, most spins were carried out at 764 g. 430-g spins were used for fragile mutant strains. Disposable plastic conical centrifuge tubes, with screw-on lids were used for most steps. Glass conical centrifuge tubes were necessary for the β -mercaptoethanol (β ME),

dithiothreitol (DTT), and H₂O₂ incubations to prevent worms from sticking to the sides of the tube. All reactions were carried out at room temperature, unless otherwise noted. One volume is equivalent to 5 ml.

Worms were washed off plates with 2 vol M9 buffer and spun for 2 min, and then washed several times, leaving behind a worm pellet of 0.1–0.5 ml. The worms were resuspended in 1 vol M9 and left on ice until ready to use. The fixative, containing 1% formaldehyde and 71% methanol was made as follows: 0.25 ml 16% paraformaldehyde (obtained from E M Sciences), 2.85 ml 100% methanol, and 0.9 ml 4× Modified Ruvkun's Witches Brew (MRWB) without methanol (320 mM KCl, 80 mM NaCl, 40 mM EGTA [pH 7.4], 20 mM spermidine, 60 mM Pipes [pH 7.4]). The worms were spun again, and excess M9 was aspirated off. The 4 ml of fixative were quickly added. The tube was inverted to mix and plunged into liquid nitrogen until frozen. The tube was then placed on ice to thaw. In some cases, the freeze-thaw cycle was repeated two more times. The worms were then incubated for 1 h on ice, and mixed by gentle inversion at 5-min intervals. The worms were again pelleted and resuspended in 1 vol 1× TTB (Tris Triton buffer: 100 mM Tris-HCl [pH 7.4], 1% Triton X-100, 1 mM EDTA). After gently inverting three times, the worms were spun and washed again twice. The worms were then resuspended in 1 vol 1× TTB and transferred to a glass tube. Then, 1% βME was added and the worms were incubated at 37°C for 2 h on a rotating wheel at low speed. The worms were then washed in 1 vol 1× BO₃ buffer (50× BO₃ buffer: 1 M H₂BO₃, 0.5 M NaOH, pH 9.2). They were then resuspended in 10% 0.1 M DTT in 1 1/2 vol 1× BO₃ and gently rotated on the wheel for 15 min. The worms were then washed in 1 1/2 vol 1× BO₃ twice. The worms were then resuspended in 1 1/2 vol 1× BO₃ and 0.075 ml 30% H₂O₂ was added. The tube was allowed to set, uncovered, for 15 min, with gentle inversion every 5 min. The worms were then washed in 1 1/2 vol of 1× BO₃. The worms were resuspended in 1 vol AbB (Antibody buffer B: 1× PBS, 0.1% BSA, 0.5% Triton X-100, 1 mM NaN₃, 1 mM EDTA), and left on the wheel for 20 min. The worms were then resuspended in 1 vol AbA (antibody buffer A: same as AbB, except 1.0% BSA). The solution was divided among the number of tubes necessary and then pelleted. Primary antibodies were diluted at 1:50 in AbA to a volume of 0.5 ml, and added to worm pellets. The lids were placed on the tubes, and wrapped in parafilm, then incubated on the wheel overnight. The worms were then washed four times in 1 vol of AbB for 30 min each on the wheel. The worms were then resuspended in 1 vol AbA and pelleted. Rhodamine- or fluorescein-conjugated secondary antibodies (Cappel, Durham, NC) were diluted 1:250 or 1:500 in AbA and added to the pellets. The tubes were capped, covered in foil, and incubated for 2 h on the wheel. Samples were handled individually as follows. Worms were pelleted and the solution was aspirated, leaving a small amount of supernatant above the worms. 30-μl samples were placed on slides and 50 μl AquaMount (Lerner Labs., Pittsburgh, PA) was then added per slide. Coverslips were applied and then the slides were viewed with a Zeiss Axioplan microscope, equipped with an HBO 100 watt mercury illuminator. Photographs were taken with an MC100 Spot camera, using Kodak Ektachrome P1600, EPH 135-36 or Fujichrome 1600, EPH 135-36 film. Fluorescein samples were photographed at 1600 ASA and rhodamine samples at 800 ASA. Film was push processed two stops.

To prepare frozen sections, ~200 μl of a 50:50 mixture of nematodes in M9 buffer were injected into a mold filled approximately halfway with OCT compound (Miles, Inc., Elkhart, IN). Additional OCT compound was added gently to encircle the area containing the worms and to fill up the mold. Freezing was accomplished by immersing the mold in 2-methylbutane at -72°C. Frozen sections of ~3 μm thickness were mounted on Superfrost/Plus (charged) slides (Fisher Scientific, Pittsburgh, PA). After staining a representative section by hematoxylin and eosin, we observed that ~10% of the sections were cross sections through the pharynx. We obtained best results, in terms of numbers of sections retained on slides and intensity of staining, when the entire procedure was carried to completion on the same day that the sections were cut. The following washes were performed in Copland jars at room temperature: PBS for 5 min, 95% ethanol for 10 min, and PBS for 5 min (×2). After blotting away excess liquid, 200 μl of 10% goat serum in PBS was applied and incubated at room temperature for 15 min to block nonspecific binding of secondary antibodies. After blotting off excess goat serum, 200 μl of primary antibody diluted 1:200 in PBS was added and the slide was placed in a humidified chamber at 37°C for 1 h. The slide was then washed at room temperature in PBS for 5 min, and then in 1% BSA, PBS for 5 min. After removing excess liquid, 200 μl of fluor-conjugated goat secondary antibodies diluted 1:200 or 1:400 in 1% BSA, PBS was added and the slide placed in a humidified chamber at 37°C for 40 min (in the dark). This was followed by two washes at room temperature in the dark: 1% BSA, PBS

for 5 min, and then PBS for 5 min. After removing most of the liquid, the sections were covered with 30 μl of AquaMount and a coverslip, and photographed as noted above.

Results

cm20d2 Encodes Six Ig Domains and Hybridizes to an ~22-kb mRNA

Recently, Waterston et al. (1992) determined ~400 bp of sequence from the 5' ends of each of 1517 clones from a sorted cDNA library, then translated and compared these sequences to the databases. One cDNA, *cm20d2*, had highest similarity to myosin light chain kinases and twitchin. We were alerted and kindly given the clone. We finished sequencing the 1978-bp insert. It has no similarity to the 55-kb genomic sequence that contains *unc-22* (Benian et al., 1989, 1993). *cm20d2* encodes part of one and five complete Ig domains (No. 2–7 in Fig. 3 a). These six domains give the best BLAST scores to Ig domains from members of the intracellular muscle branch of the Ig superfamily (data not shown).

When *cm20d2* was used as a probe against Northern blots, a very large mRNA, approximately the size of the twitchin mRNA (21.6 kb; Benian et al., 1993) was detected (Fig. 1). To eliminate the possibility that *cm20d2* merely cross-hybridized to the twitchin mRNA, we also probed RNA prepared from the *unc-22* allele *ct37*. *ct37* has a several-kilobase deletion of *unc-22* coding sequence, and lacks the twitchin polypeptide on an immunoblot (Moerman et al., 1988). As shown in Fig. 1, although no mRNA was detected from *ct37* with an *unc-22* probe, an mRNA was detected from *ct37* with *cm20d2* as probe.

cm20d2 Maps to a Genomic Cosmid Which, as a Probe, Detects a Rearrangement in the unc-89 Allele st515

Upon hybridization to a filter of 960 physically-mapped YAC clones (kindly provided by R. Waterston), the *cm20d2* cDNA detects 2 YACs, Y56D12 and Y74A11. We obtained cosmids covering the overlapping region from J. Sulston and A. Coulson. *cm20d2* hybridized to the cosmids as indicated in Fig. 2. This portion of the physical map corresponds to the central portion of chromosome I. The closest gene that had been cloned was *unc-73*. *unc-73* is required for axon guidance and segregation of developmental potential. *unc-73* has been cloned via transposon tagging (Steven, R., and J. Culotti; Mt. Sinai Hospital Research Institute, Toronto, personal communication). *cm20d2* is not encoded by *unc-73* for the following reasons: (a) just to the left and slightly overlapping cosmids that hybridize to *cm20d2* is cosmid C11B5, which as a transgene rescues *unc-73*; (b) there is no DNA sequence homology between *cm20d2* and the *unc-73* sequence; (c) *unc-73* mRNAs are 6.3 and 7.7 kb; and (d) *unc-73* encodes a protein homologous to yeast CDC24, and lacks Ig domains. Inspection of the genetic map revealed that one of the closest genes to *unc-73* is *unc-89*, only 0.21 map units away. Mutations in *unc-89*, as described above, specifically affect muscle. Eleven *unc-89* alleles were examined by Southern blots for RFLPs with cosmid clones C34E1 and K08D2 as probes. For the allele *st515*, isolated from a mutator background (Rioux, S., and R. Waterston, personal communication), and thus

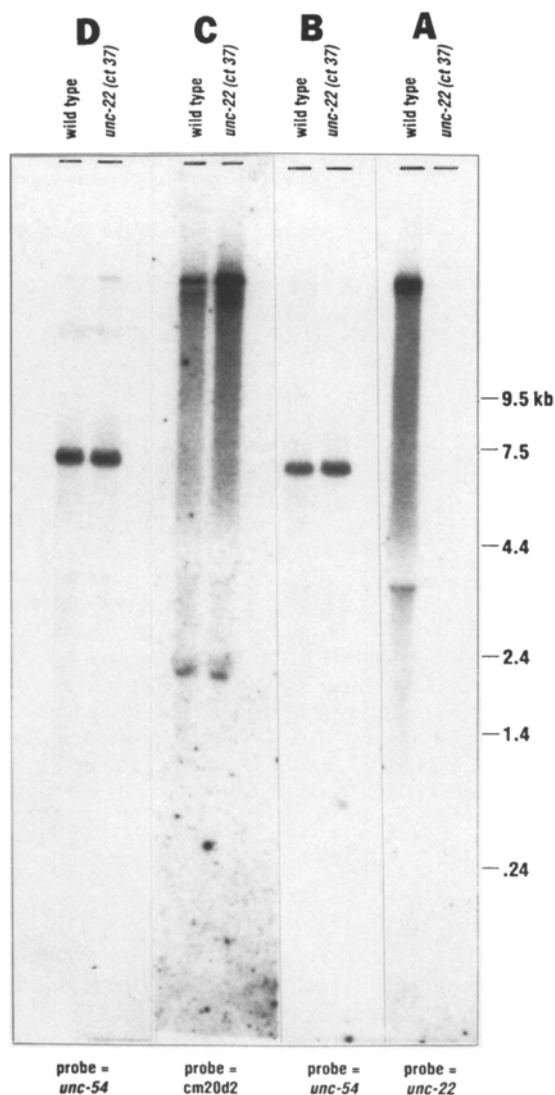


Figure 1. cDNA cm20d2 hybridizes to an ~22-kb mRNA. *C. elegans* total RNA (~20 µg per lane) was fractionated on a 0.7% agarose-formaldehyde gel, transferred to nitrocellulose membrane and hybridized with the indicated probes. *unc-22(ct37)* is a several kilobase deletion within the *unc-22* (twitchin) gene which produces no detectable twitchin (Moerman et al., 1988). As shown in *A*, a probe specific for *unc-22* detects the mRNA known from sequence analysis to be 21.6 kb from wild type, but no mRNA from *ct37*. This blot was stripped of probe and rehybridized with a probe specific for *unc-54* (myosin heavy chain B) to show that the RNA was largely undegraded and that RNA was present in the *ct37* lane (*B*). A portion of the same blot was hybridized with cDNA cm20d2. As shown in *C*, cm20d2 hybridizes to an mRNA of approximately the same size as the mRNA for *unc-22* (twitchin). Since this message can be detected in RNA from *ct37*, it rules out the possibility that cm20d2 is merely cross-hybridizing to the *unc-22* message. *D* shows the result of rehybridizing the membrane used in *C* with an *unc-54* probe.

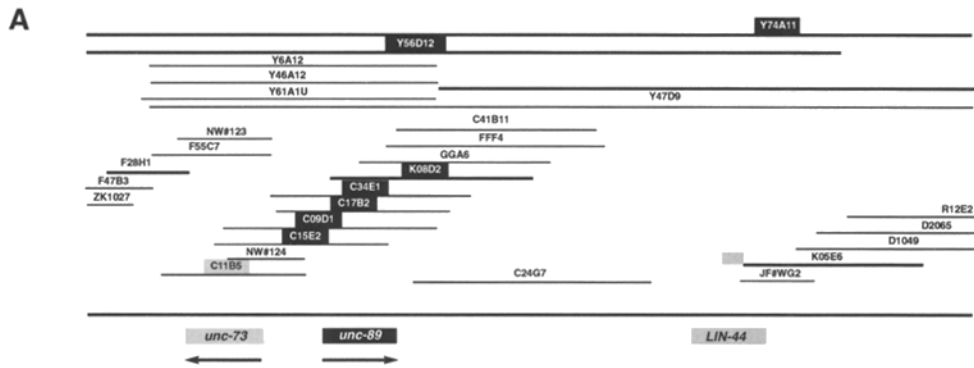
likely to be transposon-induced, a polymorphism was seen with C34E1 as probe, consistent with a deletion of ~100 bp within a 2.0-kb Hind III fragment. This 2.0-kb fragment was subcloned into Bluescript, and several hundred bp of sequence were determined from each end. Primers were designed from the sequence and used to produce, by PCR,

the corresponding segment from both wild type and *st515*. After sequencing each segment, *st515* was found to contain an 84-bp deletion and 10-bp duplication, resulting in an in-phase TGA stop codon (Fig. 2*b*).

unc-89 Encodes a 731,897-D Polypeptide Consisting of SH3, dbl/CDC24, PH, Multiphosphorylation, and Ig Domains

Because cosmid C34E1 is the most centrally placed of the five cosmids to which cm20d2 hybridizes, and C34E1 contains the site of deletion for *st515*, we determined the complete 44,848-bp sequence of the C34E1 insert. Exons were predicted by the computer program, GeneMark, and then, exon-intron boundaries were confirmed by sequencing 200–900-bp reverse transcriptase PCR products. These PCR products revealed several small (less than 100 bp) exons not predicted by the program. Alignment of cm20d2 and two other partial cDNAs (each ~2.2 kb) with this genomic sequence confirmed several other exons. The confirmed exons predicted one transcriptional unit, beginning at position 5,870 and continuing until the end of the cosmid insert. Because an in-frame stop codon was not found, screens of oligo-dT-generated cDNA libraries were performed, and yielded two cDNA clones that extended the sequence an additional 1,982 bp. The sequence of the furthest 3'-clone revealed an in-frame stop codon, a 680 bp 3'-untranslated sequence, and an AATAAA consensus polyadenylation signal 19 bp upstream of a poly(A) tail. After finding coding sequence that could account for a protein ~99% the size of twitchin, a primer designed from the furthest 5' exon was used in 5' RACE to determine the 5' end of the *unc-89* message. An in-frame initiator methionine beginning at position 4,920 and a 145-bp 5' untranslated sequence starting at position 4,774 were found. The *unc-89* gene extends over at least 42,041 bp of the genome and contains at least 30 introns. These sequence data are available from GenBank under accession number U33058. The 5' end of the *unc-73* gene lies at position 773 in our sequence and is transcribed in opposite orientation to *unc-89* (Steven, R., and J. Culotti, personal communication), giving a distance of 4,001 bp between *unc-73* and *unc-89*.

The entire UNC-89 polypeptide consists of 6,632 amino acids with a calculated molecular weight of 731,897. As shown in Fig. 3*a*, the UNC-89 sequence begins with 67 residues having no homologies to sequences in the databases, followed by an SH3 domain, a dbl/CDC24 domain, a PH domain, 7 Ig domains, 645 amino acids consisting of several repeats including 44 copies of the amino acid triplet lysine serine proline (KSP), and then 46 Ig domains. The region with homology to dbl/CDC24 was recognized by a BLASTP search which gave the highest scores to products of the *dbl* (Eva et al., 1988; Ron et al., 1991) and *Ost* oncogenes (Horii et al., 1994), which are members of the CDC24 family of guanine-nucleotide release factors that link rho and rac signaling pathways (Fig. 3*b*). Because these domains are often followed by PH domains (Gibson et al., 1994), such a sequence was searched for, by inspection. As shown in Fig. 3*c*, there is indeed a PH domain which matches a consensus for PH domains and a prediction of secondary structure indicates that the key el-



B

wild type	GGCGTCAAGA...84bp...	ctgaaaatttgatattgg
<i>unc-89 (st515)</i>	GGCGTCAAGA---84bp deletion---	aaatttgatactgaaaatttgatattgg
Ig no. 5	gly val lys	lys ile stop tyr stop

Steven's analysis of *unc-73* indicates that the transcription start sites of the two genes are separated by 4001 bp, and they are oppositely oriented (directions indicated by arrows). On the genetic map, *unc-73* and *unc-89* are separated by 0.21 map units. *lin-44* lies in the 20 kb between C24G7 and K05E6, within the 4.1 kb just left of K05E6 (Herman et al., 1995; indicated by light shading). *unc-89* and *lin-44* have not been separated by recombination. The 3' end of *unc-89* has not been placed on this physical map. (B) Sequence alteration in the *unc-89* allele *st515*. The mutator-induced *st515* mutation consists of an 84-bp deletion and 10-bp duplication, resulting in an in-frame TGA stop codon within coding sequence for Ig no. 5 of the predicted UNC-89 protein (see Fig. 3). This is predicted to result in a truncated polypeptide of 135 kD that can indeed be detected by Western blot (see Fig. 4).

elements (primarily β -strands and COOH-terminal α -helix) of the known structure are present. When the UNC-89 sequence was tested by an experimental WWW profile server (Lausanne, Switzerland) against a set of profiles, an SH3 domain was detected. It extends from position 68–122 in the UNC-89 sequence (Fig. 3 d). The 44 KSPs and related sequences are mostly arranged in a repeat of 10 residues (typically KSPTKKEKSP) which is present in 26 copies (Fig. 3 e). For the Ig domains, a consensus sequence derived by the PILEUP program (Fig. 3 f) consisting of amino acids found in 20 of the 53 Ig domains, has high similarity to the consensus for the 30 Ig domains of twitchin (Benian et al., 1989) and a consensus for six intracellular Ig superfamily members derived by Price and Gomer (1993). The 7 and 46 Ig domains are essentially in tandem with only 0–19 amino acids separating consecutive domains, except after Ig domains 17 and 18, and 26 and 27, where larger spacers of 38 and 36 residues are found. It is also interesting to note that near the COOH terminus, separating Ig domains 41–50, there are somewhat larger spacers of 9–19 residues. The stop codon in the *unc-89* allele *st515* lies in the middle of Ig domain 5.

Antisera Raised to an *unc-89* Encoded Peptide React with a Twitchin-sized Polypeptide in Wild Type and Truncated Polypeptides in Two *unc-89* Alleles

To confirm the presence of the very large polypeptide predicted from sequence analysis, we raised antisera to a portion of the polypeptide. A 364-amino acid peptide, en-

Figure 2. Molecular cloning of *unc-89*. (A) cm20d2 localizes to a region of the physical map that corresponds to the *unc-73-unc-89* region of the genetic map. A portion of the physical map of chromosome I, adapted from ACeDB, is shown. Long lines indicate YAC clones and short lines denote cosmid or lambda clones. cDNA cm20d2 hybridizes to the YACs and cosmids indicated by white letters in black boxes. Cosmid C34E1 was sequenced in this study. The mutationally-defined gene *unc-73* has been cloned, and the cosmid C11B5 (lightly shaded) can rescue *unc-73* mutants as a transgene (Stevens, R., and J. Culotti, personal communication). Comparison of our sequence data on *unc-89* and R.

coded by the majority of cDNA cm20d2 and corresponding to sequence including the end of Ig domain 2 through the middle of Ig domain 6, was expressed as a glutathione-S transferase fusion protein. An affinity-purified rabbit antiserum to this fusion protein (EU30) reacts to a polypeptide of about the same size as twitchin (~750 kD) on immunoblots containing total Laemmli-soluble nematode proteins (Fig. 4 a). This antiserum reacts with smaller, probably truncated polypeptides from *st515* (Fig. 4 a) and *e2338* (an *unc-89* amber allele; Fig. 4 b). The 135,000-D size of the truncated polypeptide from *st515* is consistent with the position of the stop codon introduced in the coding sequence.

R. Francis had previously generated ~40 monoclonal antibodies to muscle cells, and their underlying basement membranes and hypodermis (Francis and Waterston, 1985, 1991). One of these monoclonals, MH42, localized to the center of A-bands by immunofluorescence (Waterston, 1988). Previously, R. Francis failed to detect a polypeptide on a Western blot with MH42. Using the more recently developed and highly sensitive enhanced chemiluminescent method, we detect reaction of MH42 with a polypeptide of the same size as with EU30 (Fig. 4). MH42 fails to react to a polypeptide in *st515*, but detects approximately the same-sized truncated polypeptides in *e2338* as EU30. That the polyclonal antiserum EU30 and monoclonal MH42 react to the same protein, the product of the *unc-89* gene, is suggested by the results from both these Western experiments and immunofluorescent localization in both wild type and in *unc-89* mutants (see below).

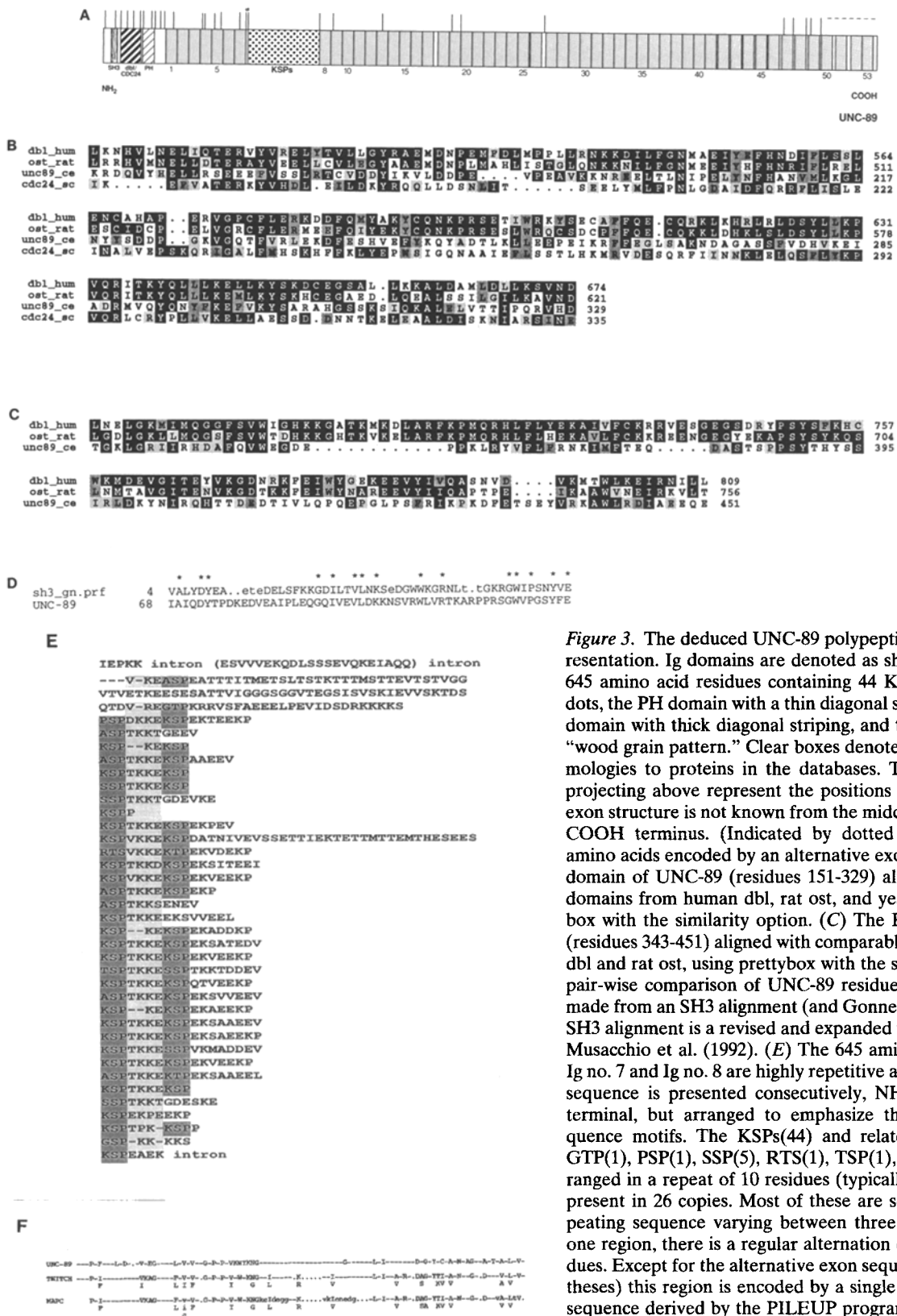


Figure 3. The deduced UNC-89 polypeptide. (A) Schematic representation. Ig domains are denoted as shaded boxes (1-53), the 645 amino acid residues containing 44 KSPs are indicated with dots, the PH domain with a thin diagonal striping, the dbl/CDC24 domain with thick diagonal striping, and the SH3 domain with a “wood grain pattern.” Clear boxes denote sequences with no homologies to proteins in the databases. The short vertical lines projecting above represent the positions of introns. The intron-exon structure is not known from the middle of Ig no. 50 until the COOH terminus. (Indicated by dotted line) *a* represents 22 amino acids encoded by an alternative exon. (B) The dbl/CDC24 domain of UNC-89 (residues 151-329) aligned with comparable domains from human dbl, rat ost, and yeast cdc24, using prettybox with the similarity option. (C) The PH domain of UNC-89 (residues 343-451) aligned with comparable domains from human dbl and rat ost, using prettybox with the similarity option. (D) A pair-wise comparison of UNC-89 residues 68-122 with a profile made from an SH3 alignment (and Gonnet PAM160 matrix). The SH3 alignment is a revised and expanded version of one made by Musacchio et al. (1992). (E) The 645 amino acids lying between Ig no. 7 and Ig no. 8 are highly repetitive and highly ordered. This sequence is presented consecutively, NH₂-terminal to COOH-terminal, but arranged to emphasize the repeating linear sequence motifs. The KSPs(44) and related sequences ASPs(7), GTP(1), PSP(1), SSP(5), RTS(1), TSP(1), KTP(1) are mostly arranged in a repeat of 10 residues (typically KSPTKKEKSP) and present in 26 copies. Most of these are separated by a third repeating sequence varying between three and eight residues. In one region, there is a regular alternation of eight and seven residues. Except for the alternative exon sequence (limited by parentheses) this region is encoded by a single exon. (F) A consensus sequence derived by the PILEUP program for all 53 Ig domains which consists of any residue present at a given position in at least 20 of the 53 Ig domains. This UNC-89 consensus is aligned to a consensus for the 30 Ig domains of twitchin (Benian et al., 1989) and a consensus for Ig domains found in six of the myosin-associated intracellular Ig proteins (MAPC) calculated by Price and Gomer (1993). Dots are gaps introduced to align the three consensus sequences.

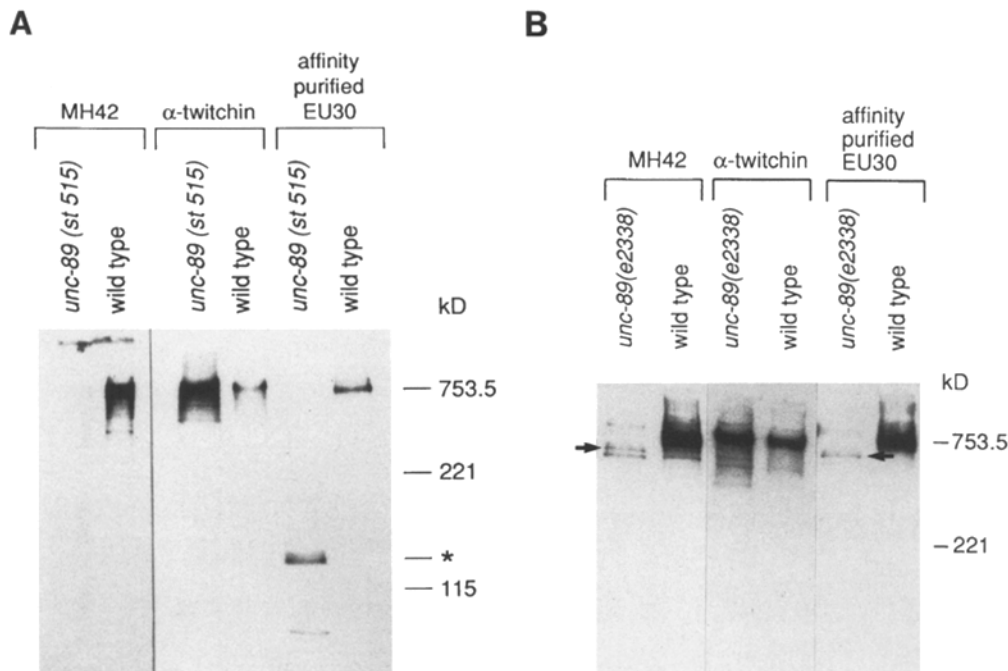


Figure 4. Antisera raised to a part of the predicted UNC-89 polypeptide react on Western blots with an ~750-kD polypeptide in wild type and truncated polypeptides in two *unc-89* alleles. Total SDS-soluble nematode proteins were separated on 5% SDS-PAGE, transferred to nitrocellulose, reacted with affinity-purified EU30, α -twitchin, or monoclonal MH42, followed by ECL detection. EU30 is a rabbit polyclonal antiserum generated to a 364 residue segment of UNC-89 including the end of Ig no. 2 through the middle of Ig no. 6 expressed as an *E. coli* fusion protein. Monoclonal MH42 is one of ~40 monoclonal antibodies generated to preparations enriched for muscle (Francis and Waterston, 1985; 1991).

MH42 was previously known to localize to the center of muscle A-bands (Waterston, 1988). (A) EU30 reacts with a polypeptide of about the same size as twitchin (known from complete sequence analysis to be 753.5 kD) from wild-type animals. MH42 detects a similarly sized polypeptide (and a small smear which probably represents degradation) from wild type. In contrast, EU30 detects an ~135-kD polypeptide from *unc-89(st515)*, the size of a polypeptide predicted from the position of the TGA stop codon created by the deletion/duplication in the *st515* sequence. Inadvertently, the *st515* lane was loaded with more protein than the wild-type lane, as shown by the α -twitchin antibody reaction. This overloading allowed easier detection of the truncated polypeptide with EU30, and increased the certainty that MH42 does not react with the truncated polypeptide. MH42 fails to detect a polypeptide from *unc-89(st515)* probably because its epitope lies COOH-terminal to the stop codon in *unc-89(st515)*. (B) Greatly reduced quantities of a slightly smaller polypeptide or several polypeptides (arrows) are detectable with both EU30 and MH42 from the *unc-89* amber terminator mutation *e2338*. Comparable amounts of total protein were loaded from both wild type and *e2338*, as shown by the reaction with α -twitchin.

Antisera Localize the UNC-89 Protein to the Center of Muscle A-Bands

Indirect immunofluorescence was used to determine the location of this protein in the sarcomere. Wild-type animals were simultaneously reacted with an antiserum to twitchin and to MH42 (Fig. 5, A and B). This reaction was visualized by the use of secondary antibodies conjugated to rhodamine and fluorescein. Moerman et al. (1988) observed that twitchin is localized to the A-bands with a central gap of staining. In Fig. 5, A and B, the arrows point to the middle of the same A-band, and indicate that this gap is filled with the protein recognized by MH42. In Fig. 5, C and D are shown results of staining the same animals with EU30 and a monoclonal to myosin heavy chain A (myoA). The thick filaments of nematode body wall muscle contain two myosin heavy chain isoforms, myoB distributed along most of the length of the thick filament, and myoA located in the center (Miller et al., 1983). When viewed by immunofluorescence with monoclonals specific for each isoform, myoB localizes to the A-band with a gap in the middle, and this gap is filled with myoA. We find that EU30 reacts with the center of A-bands, with the same width of distribution as myoA (Fig. 5, C and D). This colocalization is further demonstrated when the same portions of muscle cells are viewed with a filter that allows detection of both fluorophores (data not shown). Both EU30 and MH42 also react to pharyngeal muscles which is consistent with

the abnormal pharyngeal muscle seen in *unc-89* mutants by polarized light microscopy. As shown in the cross-sections through the pharynx, EU30 and MH42 staining occurs in the middle (Fig. 5, F and G) of the radial muscle fibers of the pharynx, as compared with the broad A-band distribution (Fig. 5 E) of the pharyngeal muscle-specific myosin heavy chain, myoC (Miller et al., 1986).

In *unc-89* Mutants, UNC-89 Protein Is Missing, or Is Present in Reduced Amounts, but Is Still Located in the Middle of A-Bands

To obtain further evidence that the antibody reagents are directed against the *unc-89* product, we examined *unc-89* mutants by immunofluorescence microscopy. One group of nematodes was stained at the same time with anti-twitchin and MH42 and viewed with different filters. A separate set of animals was reacted with EU30. Reaction to twitchin antibodies was used to check the success of the immunofluorescence procedure and as an independent measure of A-band organization. Two nonsense (amber) alleles of *unc-89*, *st85*, and *e2338*, (Waterston, R.H., personal communication) show very much reduced staining as compared to wild type. *st85* still shows some localization to the center of A-bands (data not shown), but the staining in *e2338* is confined to dots or clumps (Fig. 6, B and C). *unc-89(e1460)* yields an abundant normal-sized polypeptide on

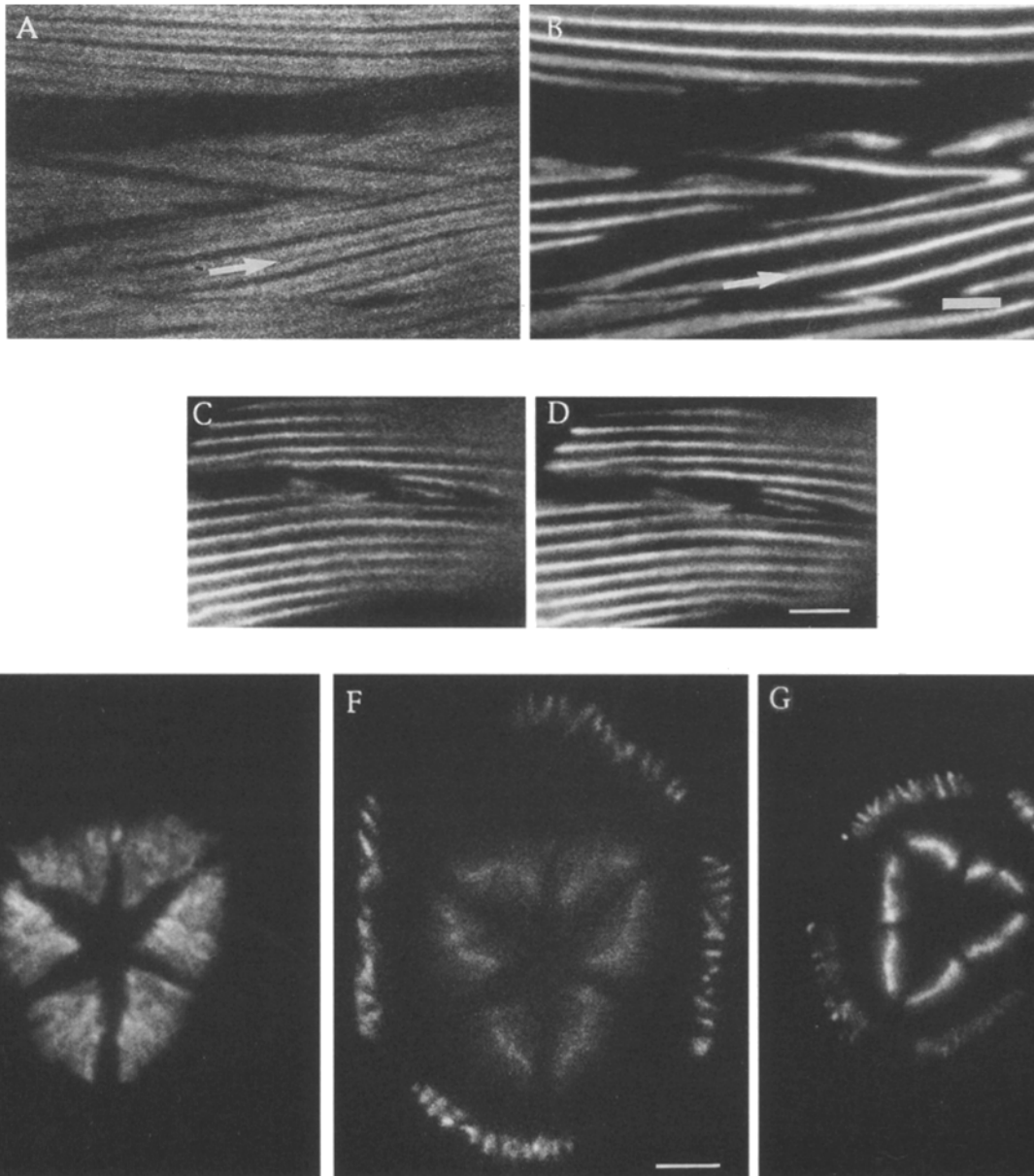


Figure 5. EU30 and MH42 localize to the middle of A-bands in both body wall and pharyngeal muscle from wild-type animals. (*A* and *B*) Animals were simultaneously reacted with a polyclonal α -twitchin and monoclonal MH42, and secondary antibodies of goat α -rabbit-fluorescein and goat α -mouse rhodamine as described in Materials and Methods. A portion of body wall muscle is shown, as viewed with the fluorescein channel in *A* (to show α -twitchin reaction), and the rhodamine channel in *B* (to show MH42 reaction). The arrows point to the middle of the same muscle A-band. (*C* and *D*) Animals were simultaneously reacted with a monoclonal to myosin heavy chain A (α -myoA) and polyclonal EU30, and appropriate fluor-labeled secondary antibodies. The same portion of body wall muscle is shown, as seen with the rhodamine channel in *C* (to reveal α -myoA reaction) and as seen with the fluorescein channel in *D* (to reveal EU30 reaction). (*E* and *F*) Frozen cross sections were simultaneously reacted with a monoclonal to pharyngeal muscle-specific myosin heavy chain C (α -myoC; shown in *E*), polyclonal EU30 (shown in *F*), and appropriate fluor-labeled secondary antibodies. *E* and *F* are from the same cross section of a pharynx as viewed with the rhodamine and fluorescein channels, respectively. In the cross-sections in *F* and *G* (below), EU30 and MH42 also react with the four body wall muscle quadrants. (*G*) Frozen cross section reacted with monoclonal MH42 and goat α -mouse-rhodamine. Bars, 5 μ m.

immunoblots with either EU30 or MH42 (data not shown). Consistent with this, immunofluorescence microscopy shows heavy but discontinuous reaction to the center of A-bands (Fig. 6, *D* and *E*). Both MH42 and EU30 show the same intensity of staining and localization for wild type and these three *unc-89* mutants (*e1460*, *st85*, and *e2338*).

Fig. 6, *F* and *G* display the results of staining one *st515*

animal with MH42, and another *st515* animal with EU30. Consistent with the Western data, no staining is seen with MH42, but weak staining, surprisingly much of it localized to the center of A-bands, is seen with EU30. This suggests that most of the determinants for localization of the UNC-89 protein to the middle of A-bands reside in the first 135 kD of the UNC-89 polypeptide.

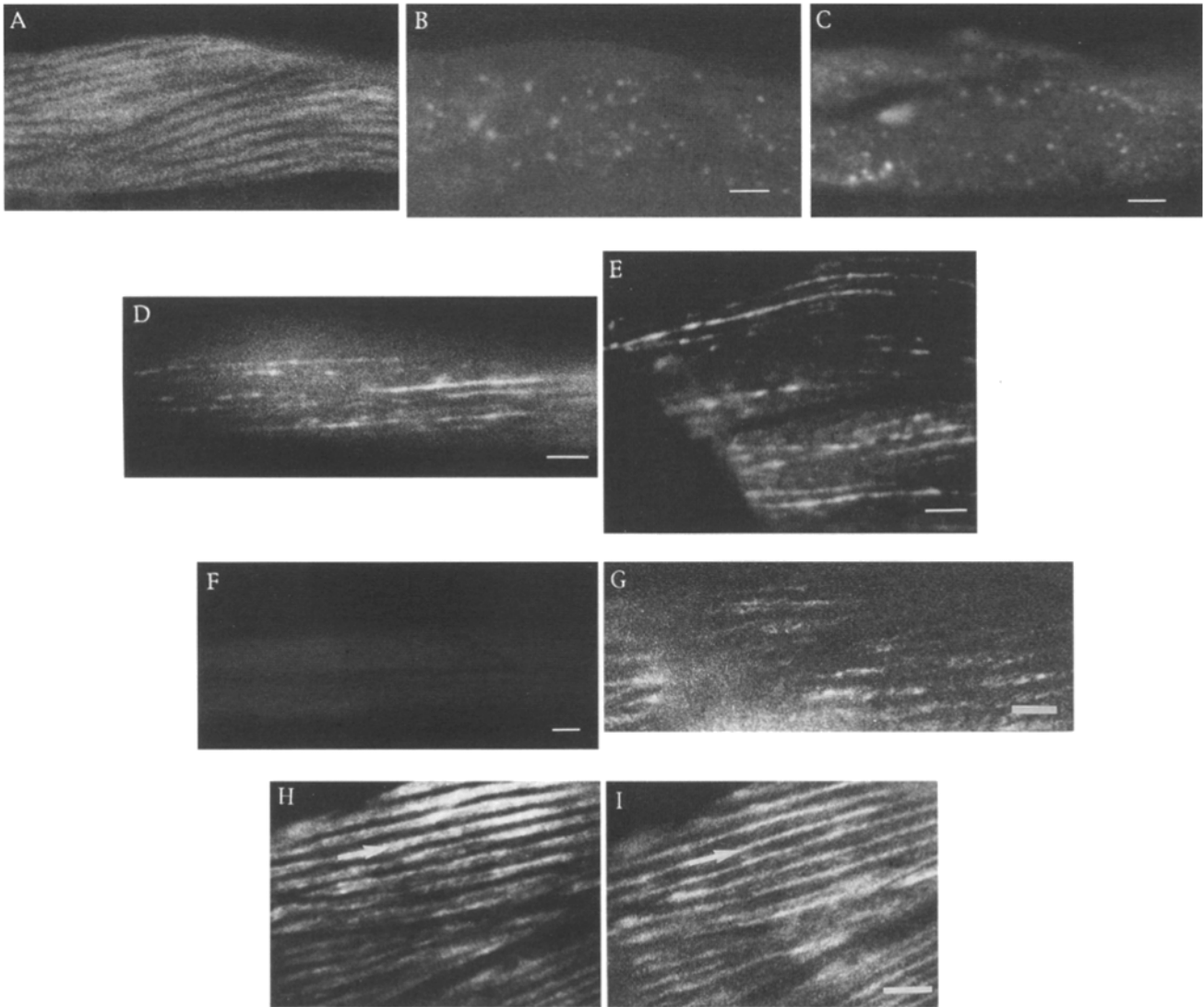


Figure 6. EU30 and MH42 staining of body wall muscle in mutant animals. (A and B) Photographs of the same portion of muscle from an *e2338* animal simultaneously reacted with α -twitchin (in A) and MH42 (in B). (C) Photograph of a separate *e2338* animal reacted with EU30. (D) Photograph of body wall muscle of *e1460* stained with MH42 (simultaneous staining with α -twitchin showed a nearly normal staining pattern). (E) A separate *e1460* animal stained with EU30. (F) A *st515* animal stained with MH42. (G) A *st515* animal stained with EU30. This mutant that produces a fairly abundant truncated 135-kD polypeptide shows reduced staining but much of it is still restricted to the middle of A-bands. (H and I) Localization of UNC-89 protein in body wall muscle with thick filaments having myosin heavy chains of only the myoA isoform. Nematodes of genotype *eDp23;unc-54(e190)* were stained with both α -myoA and EU30. The arrows point to the center of the same A-band. In H it can be seen that MyoA is distributed throughout the A-band, whereas in I it can be seen that UNC-89 (EU30 staining) is confined to the center of A-bands, as it is in wild type. Twitchin is also distributed as it is in wild type, with a lack of staining in the middle (data not shown). Bars, 5 μ m.

MyoA Is Not the Determinant for UNC-89 Localization

Given that twitchin colocalizes with myoB (Moerman et al., 1988), and UNC-89 seems to colocalize with myoA, it is tempting to hypothesize that in wild-type, twitchin associates with myoB and UNC-89 protein associates with myoA. This was tested by performing immunofluorescence microscopy on the strain *eDp23;unc-54(e190)* which produces no myoB (encoded by the *unc-54* gene) but forms thick filaments because of overexpression of myoA due to a duplication of the myoA gene. As shown in Fig. 6 H, anti-myoA stains the entire, broad A-band. Nevertheless, UNC-89 (EU30) is still narrowly restricted to the cen-

ter of A-bands (Fig. 6 I). Arrows point to the center of the same A band in both Fig. 6, H and I. Similar results have been obtained with MH42 staining of *unc-54(e190)* animals transformed with extrachromosomal arrays of the myoA gene (Coutu Hresko, M., P. Hoppe, and R. Waterston, personal communication). Thus, the most important determinants for UNC-89 localization appear to reside in proteins/structures different from myoA. In addition, twitchin is seen to have the same distribution as in wild type, with the same gap in the middle of A-bands (data not shown). Thus, in this strain, it would seem that twitchin, in the absence of myoB can instead associate with myoA and/or other proteins.

Discussion

We have characterized a >42-kb gene that encodes a 20.7-kb mRNA and 732-kD polypeptide consisting primarily of 53 Ig domains, a possible multiphosphorylation sequence, and regions with homology to SH3, CDC24, and PH domains. That this gene corresponds to *unc-89* is based on the following evidence: (a) genomic and cDNA clones localize to a region of the physical map that corresponds to a region of the genetic map which contains *unc-89*. (b) An antiserum raised to a peptide encoded by a portion of the sequence reacts specifically to muscle, localizing to the middle of the A-band, a region that contains the M-line. The M-line is missing in one *unc-89* allele examined by EM. On Western blots, this antiserum reacts with an ~750-kD polypeptide from wild type, and with smaller, truncated polypeptides from an *unc-89* deletion (*st515*) and an *unc-89* amber nonsense mutation (*e2338*). Indirect immunofluorescence microscopy results are consistent with the immunoblot data in that staining is reduced or greatly reduced in four *unc-89* alleles. (c) *unc-89(st515)* was shown to contain an 84-bp deletion and 10-bp duplication resulting in an in-frame stop codon in the coding sequence of the gene, consistent with the 135-kD polypeptide detected by Western blot.

In nematode obliquely-striated body wall muscle, we have localized the UNC-89 protein to the middle of the A-band, where the M-line resides, but it may extend into a broader central region of the A-band. Although at the resolution of immunofluorescence microscopy, the UNC-89 protein is located in the same central region of the A-band as the minor myosin heavy chain *myoA*, we have shown that the localization of UNC-89 is independent of *myoA*. This conclusion is based on finding UNC-89 restricted to the center of A-bands even when the entire A-band contains *myoA*, as in the mutant *eDp23;unc-54(e190)*. Moreover, this is consistent with UNC-89 being a component of the M-line itself. The pharynx of the nematode is a neuromuscular pump used for feeding. The pharynx has radially oriented myofilaments (Albertson and Thomson, 1976). We found that UNC-89 is also localized to the center of the A-bands in pharyngeal muscle, which suggests that this muscle also has an M-line. Thus, in two different muscle types, UNC-89 has a similar location, and quite probably a similar function. Possibly different isoforms of UNC-89, generated by alternative splicing, are expressed in pharyngeal vs. body wall muscle. At least one example of alternative splicing in *unc-89* is known, and is indicated in Fig. 3 a. Also, there is at least one allele of *unc-89*, *ad539*, which has abnormal pharyngeal, but normal body wall muscle structure (Avery, 1993; T.L. Tinley and G.M. Benian, unpublished data).

The UNC-89 polypeptide is unusual in being composed primarily of Ig domains, and having no FnIII domains. Except for telokin, which is essentially one Ig domain, all other members of the intracellular branch of the Ig superfamily contain both Ig and FnIII domains. Although having no linear sequence homology, Ig and FnIII domains form similar structures (Leahy et al., 1992). Whereas both Ig and FnIII domains have been implicated in protein-protein interactions, functional differences, such as in binding affinities, preferences for target domains etc., are

not known. The observation of 46 Ig domains in tandem in UNC-89 is surpassed only by the 90 tandem Ig domains in the I-band region of some skeletal muscle isoforms of vertebrate titin/connectin (Labeit and Kolmerer, 1995). Given the immunolocalization of UNC-89 and the phenotype of *unc-89* mutants, it is likely that the function of these tandem Ig domains in UNC-89 is to interact with other proteins in or near the M-line, especially myosin in the shaft of the thick filament. The postulated function of the very large arrays of Ig domains in I-band titin/connectin is different, probably providing length to the titin/connectin filament and resistance to stretching, rather than interacting with other proteins (Labeit and Kolmerer, 1995).

Large numbers of KSPs have previously been found in two of three subunits of human neurofilament proteins (NF-M has 12 copies [Myers et al., 1987]; and NF-H has 41 copies [Lees et al., 1988]). Neurofilaments comprise the major cytoskeleton in axons and consist of parallel arrays of 10-nm filaments linked to each other by cross-bridges. Antibody decoration experiments have shown that the COOH-terminal tail of NF-H forms cross-bridges (Hirokawa et al., 1984). In addition, by transfection experiments with NF-L and various deletion mutants of NF-M, it appears that the COOH-terminal tail of NF-M also forms cross-bridges (Nakagawa et al., 1995). The KSPs reside in the COOH-terminal tails of NF-M and NF-H and become phosphorylated at the serines by a neuronal *cdc2*-like kinase (Lew and Wang, 1995). It is hypothesized that phosphorylation of the KSPs causes the COOH-terminal tails of NF-M and NF-H to project out perpendicular to the filament core, thus forming cross-bridges (Nixon and Sihag, 1991). Interestingly, 1 KSP has been found in the first FnIII domain of myomesin (Vinkemeier et al., 1993), and 4 KSPs have been found in the COOH-terminal, M-line portion of human cardiac titin/connectin (Gautel et al., 1993). The KSPs in titin/connectin are phosphorylated by a *cdc2*-like kinase activity in developing, but not differentiated muscle (Gautel et al., 1993). These authors suggest that early in development, the COOH-terminal portion of titin/connectin, phosphorylated at the KSPs might be inhibited from attaching to M-line proteins. By analogy, the KSP-containing region of the UNC-89 protein might be a major binding site of the UNC-89 protein to itself or to other M-line proteins. The fact that the truncated polypeptide produced in *unc-89(st515)*, which lacks all of the KSPs, can still localize to the center of A-bands (perhaps to the M-line) suggests that, for UNC-89, the KSPs are not necessary for this localization. In addition, whether the serines of the KSPs in UNC-89 are phosphorylated, and whether this phosphorylation is developmentally regulated is not yet known. It is also interesting to note that the organization of the KSPs is roughly similar in four proteins: the separation of KSPs alternates between four and primarily seven or eight residues in UNC-89, between three and five in NF-H, between two and five in NF-M, and in the pattern of four, three, and three residues separating the four KSPs in titin/connectin. These intervening sequences are very similar within one protein (e.g., in UNC-89, TKKE is the predominant sequence that separates two closely-spaced KSPs, whereas for NF-H it is predominantly EKA and for NF-M it is most frequently, VP). Presumably, this reflects how the region evolved by multi-

ple rounds of duplication. Moreover, these “intervening sequences” are very similar between UNC-89 and NF-H; in fact, there is a strong bias for K occurring as the second residue past each KSP (i.e., KSP_K) in both proteins. It is interesting to note that the consensus phosphorylation site motif for p34cdc2-cyclin is (S/T)P_(K/R) (Moreno and Nurse, 1990). Thus, it is likely that in UNC-89, the KSP_K sequences serve as multiple sites for serine phosphorylation by a nematode muscle cdc2-like kinase.

The most surprising domains encountered in the UNC-89 sequence are the SH3, dbl/CDC24 and PH domains, which are well-known in signal transduction molecules (Cohen et al., 1995). The SH3 domain is an ~60-amino acid residue domain first identified as a conserved sequence in the NH₂ termini of src tyrosine protein kinases and subsequently found in a large number of other proteins. The functions of SH3 domains are not certain, but because many SH3-containing proteins are localized to the plasma membrane or the cytoskeleton, it suggests that SH3 domains mediate localization to these regions (Cohen et al., 1995). A number of ligands for SH3 have been identified and the binding sites all contain proline-rich sequences, especially a P__P motif (Musacchio et al., 1994). Precedence for an SH3 domain in a myofilament lattice protein is provided by the giant I-band protein called nebulin, where an SH3 domain at its COOH terminus is probably partly responsible for its anchorage at the Z-disc (Wright et al., 1993; Labeit and Kolmerer, 1995a).

The human oncogene *dbl*, the yeast cell division cycle protein CDC24, and an expanding family of growth regulatory proteins share a homologous 238-amino acid sequence, generally termed a CDC24 domain. These domains from CDC24 (Zheng et al., 1994) and *dbl* (Hart et al., 1991) have been shown to stimulate the exchange of GDP for GTP on Rho-like GTPases, thereby activating Rho-like GTPase activity. Activation of Rho has been shown to cause a reorganization of actin filaments via an unknown mechanism. This triggers a diverse set of cell processes including bud formation in yeast, cytokinesis in zygotes, maintenance of cell shape, formation of stress fibers and focal adhesions, cell aggregation, smooth muscle contraction, etc. (Takai et al., 1995). Similar to UNC-89, some known guanine nucleotide exchange factors (e.g., Vav and CDC25) also contain an SH3 domain (Mayer and Baltimore, 1993). The presence of a possible Rho-like stimulator as a domain of the giant UNC-89 polypeptide, suggests that Rho-like molecules might also trigger the assembly of M-lines in striated muscle.

The PH domain, first defined in the platelet protein called plekstrin, is a 100-residue sequence that has been found in many different proteins involved in intracellular signalling or the cytoskeleton (Gibson et al., 1994). PH domains from several proteins have been shown to bind to inositol phosphates (Harlan et al., 1994; Hyvonen et al., 1995). Hyvonen et al. (1995) have proposed that this binding has two functions: (a) The anchoring of some proteins to membranes would occur through the interaction of a PH domain with membrane phosphoinositides. (b) In the case of CDC24 and its homologs, PH domains reside just COOH-terminal to the *dbl*/CDC24 domains, as is true for UNC-89. The binding of an inositol phosphate compound, such as Ins(1,4,5)P₃, to the PH domain, might regulate the

nucleotide exchange activity of the neighboring CDC24 domain.

In the obliquely-striated muscle of *C. elegans*, M-lines and dense bodies are attached to the muscle cell membrane. At these positions are the locations of β -integrin (Gettner et al., 1995) in the plasma membrane, and the UNC-52 protein, which is homologous to vertebrate perlecan (Rogalski et al., 1993), in the overlying basement membrane (Francis and Waterston, 1985, 1991). Certain mutations in *unc-52* and *pat-3* (which encodes β -integrin) result in a “pat” embryonic lethal phenotype which is characteristic of mutations in at least 13 genes essential for muscle development in *C. elegans* (Williams and Waterston, 1994). Failure of these *unc-52* and *pat-3* mutants to assemble actin and myosin (although both are synthesized) into thin and thick filaments suggests that myofilament lattice assembly begins at the extracellular matrix and the cell membrane. Further support for this model comes from antibody staining of embryos at different stages of development (Hresko et al., 1994). The data on UNC-89 can be incorporated into this model. Where an M-line is formed is dictated by the accumulation of UNC-52, β -integrin, and M-line-specific basement membrane and muscle cell membrane molecules, which in some way send a signal. This signal might be an inositol phosphate compound that binds to the PH domain of UNC-89 and increases the nucleotide exchange activity of the adjacent CDC24 domain. Regardless of the nature of the signal and how it is received, the signal causes the CDC24 domain of UNC-89 to activate a Rho-like GTPase. This results in a reorganization of the actin cytoskeleton near the muscle cell membrane. Anchorage of UNC-89, and hence the M-line, at the cell membrane might be provided by the SH3 domain interacting with this reorganized actin cytoskeleton. Alternatively, this anchorage might occur through a direct interaction of the PH domain of UNC-89 with the muscle cell membrane. The M-line structure would be built by interactions of UNC-89 with itself and other, as yet uncharacterized proteins. Some examples include the 200-kD H-zone polypeptide detected with monoclonal MH9 (Waterston, 1988), and nematode homologues of vertebrate M-line proteins (creatinine kinase, M-protein and myomesin). These interactions could occur through some of the Ig domains of UNC-89, and the KSP region (by analogy to the role of the four KSPs in vertebrate titin/connectin). The long tandem array of 46 Ig domains might be designed to interact with myosin rods in the shaft of the thick filament, as is postulated for the many copies of Ig and FnIII domains in the A-band portion of vertebrate titin/connectin (Labeit et al., 1992; Labeit and Kolmerer, 1995b). This idea, together with the role of the KSP-containing tails of NF-M and NF-H in neurofilament cross-bridge formation, suggests an attractive topology for UNC-89. From the NH₂ terminus to some position within the KSP domain, UNC-89 is part of the M-line per se, at which point UNC-89 makes a bend which permits the 46 tandem Ig domains to bind to a thick filament.

The authors wish to thank the following colleagues: R. Waterston for suggesting the project to us and generously providing the clone *cm20d2*; T. Gibson for recognizing the SH3 domain, and his alignments of regions of UNC-89 to *dbl*/CDC24 and PH domains; R. Steven for pointing us towards the PH domain; S. Sammons for installing and trouble-shooting the

Staden programs on a Sun computer and generous help with sequence analysis using the GCG programs; R. Santoianni for devising and helping us with the frozen section immunofluorescence method; S. Warren and members of his laboratory (especially H. Liener and J. Iber) for instruction and generous access to their ABI sequencing machine; J. McIninch for valuable programming assistance in obtaining the GeneMark program parameters for *C. elegans*; H. Joshi, D. Moerman, and J. Heierhorst for critically reading the manuscript; R. Francis and M. Coutu Hresko for generating and then suggesting and providing the monoclonal MH42; D. Miller for providing the monoclonal antibodies to myoA and myoC. G. White, J. Waddle, and M. Gilbert for suggestions on fluorescence microscopy; D. DeLa Garza for technical assistance; N. Boguslavsky for oligonucleotide synthesis; D. Maurer and N. Larsen for running some sequencing reactions on their ABI machine; and R. Waterston, B. Williams, L. Avery, H. Epstein, and D. Riddle for providing nematode strains.

This work was supported in part by the National Institutes of Health (NIH) (to G.M. Benian and M. Borodovsky). Some nematode strains were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources. G.M. Benian is an Established Investigator of the American Heart Association.

Received for publication 11 October 1995 and in revised form 8 January 1996.

References

- Albertson, D.G., and J.N. Thomson. 1976. The pharynx of *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 275:299–325.
- Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Avery, L. 1993. The genetics of feeding in *C. elegans*. *Genetics*. 133:897–917.
- Ayme-Southgate, A., J. Vigoreaux, G. Benian, and M.L. Pardue. 1991. *Drosophila* has a twitchin/titin gene that appears to encode projectin. *Proc. Natl. Acad. Sci. USA*. 88:7973–7977.
- Ayme-Southgate, A., R. Southgate, J.D. Saide, G.M. Benian, and M.L. Pardue. 1995. Both synchronous and asynchronous muscle isoforms of projectin, the *Drosophila* bent locus product, contain functional kinase domains. *J. Cell Biol.* 128:393–403.
- Benian, G.M., J.E. Kiff, N. Neckelmann, D.G. Moerman, and R.H. Waterston. 1989. Sequence of an unusually large protein implicated in regulation of myosin activity in *C. elegans*. *Nature (Lond.)*. 342:45–50.
- Benian, G.M., S.W. L'Hernault, and M.E. Morris. 1993. Additional sequence complexity in the muscle gene *unc-22* and its encoded protein, twitchin, of *C. elegans*. *Genetics*. 134:1097–1104.
- Billingsley, P. 1961. Statistical methods in Markov chains. *Annals of Mathematical Statistics*. 82:12–40.
- Borodovsky, M., and J. McIninch. 1993. GeneMark: gene prediction of both DNA strands. *Comput & Chem.* 17:123–133.
- Borodovsky, M., E. Koonin, and K. Rudd. 1994a. New genes in old sequence: a strategy for finding genes in the bacterial genome. *Trends Biochem. Sci.* 19:309–313.
- Borodovsky, M., K. Rudd, and E. Koonin. 1994b. Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. *Nucleic Acids Res.* 22:4756–4767.
- Borodovsky, M., Y.A. Sprizhitsky, E.I. Golovanov, and A.A. Alexandrov. 1986. Statistical features in the primary structure of *E. coli* genomic DNA. II: non-homogeneous Markov chains. *Mol. Biol.* 20:833–840.
- Cohen, G.B., R. Ren, and D. Baltimore. 1995. Modular binding domains in signal transduction proteins. *Cell*. 80:237–248.
- Chirgwin, J.M., A.E. Przybyla, R.J. MacDonald, and W.J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry*. 18:5294–5299.
- Dear, S., and R. Staden. 1991. A sequence assembly and editing program for efficient management of large projects. *Nucleic Acids Res.* 19:3907–3911.
- Einheber, S., and D.A. Fischman. 1990. Isolation and characterization of a cDNA clone encoding avian skeletal muscle C-protein: an intracellular member of the immunoglobulin superfamily. *Proc. Natl. Acad. Sci. USA*. 87:2157–2161.
- Eva, A., G. Vecchio, C.D. Rao, S.R. Tronick, and S.A. Aaronson. 1988. The predicted DBL oncogene product defines a distinct class of transforming proteins. *Proc. Natl. Acad. Sci. USA*. 85:2061–2065.
- Finney, M., and G. Ruvkun. 1990. The *unc-86* gene product couples cell lineage and cell identity in *C. elegans*. *Cell*. 63:895–905.
- Francis, G.R., and R.H. Waterston. 1985. Muscle organization in *Caenorhabditis elegans*: localization of proteins implicated in thin filament attachment and I-band organization. *J. Cell Biol.* 101:1532–1549.
- Francis, G.R., and R.H. Waterston. 1991. Muscle cell attachment in *C. elegans*. *J. Cell Biol.* 114:465–479.
- Fyrberg, C., S. Labeit, B. Bullard, K. Leonard, and E. Fyrberg. 1992. *Drosophila*

- projectin: relatedness to titin and twitchin and correlation with lethal (4)102cda and bent-Dominant mutants. *Proc. R. Soc. Lond. B*. 249:33–40.
- Gallagher, P.J., and B.P. Herring. 1991. The carboxyl terminus of the smooth muscle myosin light chain kinase is expressed as an independent protein, telokin. *J. Biol. Chem.* 266:23945–23952.
- Gautel, M., K. Leonard, and S. Labeit. 1993. Phosphorylation of KSP motifs in the C-terminal region of titin in differentiating myoblasts. *EMBO (Eur. Mol. Biol. Organ.) J.* 12:3827–3834.
- Gettner, S.N., C. Kenyon, and L.F. Reichardt. 1995. Characterization of β pat-3 heterodimers, a family of essential integrin receptors in *C. elegans*. *J. Cell Biol.* 129:1127–1141.
- Gibson, T.J., M. Hyvonen, A. Musacchio, M. Saraste, and E. Birney. 1994. PH domain: the first anniversary. *Trends Biochem. Sci.* 19:349–353.
- Harlan, J.E., P.J. Hajduk, H.S. Yoon, and S.W. Fesik. 1994. Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate. *Nature (Lond.)*. 371:168–170.
- Hart, M.J., A. Eva, T. Evans, S.A. Aaronson, and R.A. Cerione. 1991. Catalysis of guanine nucleotide exchange on the CDC42Hs protein by the dbl oncogene product. *Nature (Lond.)*. 354:311–314.
- Herman, M.A., L.L. Vassilieva, H.R. Horvitz, J.E. Shaw, and R.K. Herman. 1995. The *C. elegans lin-44* gene, which controls the polarity of certain asymmetric cell divisions encodes a Wnt protein and acts cell nonautonomously. *Cell*. 83:101–110.
- Hirokawa, N., M.A. Glicksman, and M.B. Willard. 1984. Organization of mammalian neurofilament polypeptides within the neuronal cytoskeleton. *J. Cell Biol.* 98:1523–1536.
- Hori, Y., J.F. Beeler, K. Sakaguchi, M. Tachibana, and T. Miki. 1994. A novel oncogene, ost, encodes a guanine nucleotide exchange factor that potentially links Rho and Rac signaling pathways. *EMBO (Eur. Mol. Biol. Organ.) J.* 13:4776–4786.
- Hresko, M.C., B.D. Williams, and R.H. Waterston. 1994. Assembly of body wall muscle and muscle cell attachment structures in *Caenorhabditis elegans*. *J. Cell Biol.* 124:491–506.
- Hyvonen, M., M.J. Macias, M. Nilges, H. Oschkinat, M. Saraste, and M. Wilmanns. 1995. Structure of the binding site for inositol phosphates in a PH domain. *EMBO (Eur. Mol. Biol. Organ.) J.* 14:4676–4685.
- Kleffe, J., and M. Borodovsky. 1992. First and second moment of counts of words in random texts generated by Markov chains. *Comput. Appl. Biosci.* 8:433–441.
- Knappeis, G.G., and F. Carlsen. 1968. The ultrastructure of the M line in skeletal muscle. *J. Cell Biol.* 38:202–211.
- Labeit, S., and B. Kolmerer. 1995a. The complete primary structure of human nebulin and its correlation to muscle structure. *J. Mol. Biol.* 248:308–315.
- Labeit, S., and B. Kolmerer. 1995b. Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science (Wash. DC)*. 270:293–296.
- Labeit, S., M. Gautel, A. Lakey, and J. Trinick. 1992. Towards a molecular understanding of titin. *EMBO (Eur. Mol. Biol. Organ.) J.* 11:1711–1716.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.)*. 227:680–685.
- Lakey, A., S. Labeit, M. Gautel, C. Ferguson, D.P. Barlow, K. Leonard, and B. Bullard. 1993. Kettin, a large modular protein in the Z-disc of insect muscles. *EMBO (Eur. Mol. Biol. Organ.) J.* 12:2863–2871.
- Leahy, D.J., W.A. Hendrickson, I. Aukhil, and H.P. Erickson. 1992. Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science (Wash. DC)*. 258:987–991.
- Lees, J.F., P.S. Shneidman, S.F. Skuntz, M.J. Carden, and R.A. Lazzarini. 1988. The structure and organization of the human heavy neurofilament subunit (NF-H) and the gene encoding it. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:1947–1955.
- Lew, J., and J.H. Wang. 1995. Neuronal cdc2-like kinase. *Trends Biochem. Sci.* 20:33–37.
- Luther, P., and J. Squire. 1978. Three-dimensional structure of the vertebrate muscle M-region. *J. Mol. Biol.* 125:313–324.
- Maruyama, I.N., D.M. Miller, and S. Brenner. 1989. Myosin heavy chain gene amplification as a suppressor mutation in *C. elegans*. *Mol. Gen. Genet.* 219:113–118.
- Maruyama, K. 1994. Connectin, an elastic protein of striated muscle. *Biophys. Chem.* 50:73–85.
- Mayer, B., and D. Baltimore. 1993. Signalling through SH2 and SH3 domains. *Trends Cell Biol.* 3:8–13.
- 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.
- 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.* 83:2305–2309.
- Miller, D.M., and D.C. Shakes. 1995. Immunofluorescence microscopy. In *C. elegans: Modern Biological Analysis of an Organism*. H.F. Epstein and D.C. Shakes, editors. Academic Press, San Diego, CA. 365–394.
- Moerman, D.G., G.M. Benian, R.J. Barstead, L. Schreifer, and R.H. Waterston. 1988. Identification and intracellular localization of the *unc-22* gene product of *C. elegans*. *Genes & Dev.* 2:93–105.
- Moreno, S., and P. Nurse. 1990. Substrates for p34cdc2: in vivo veritas? *Cell*. 61:549–551.
- Musacchio, A., T. Gibson, V.-P. Lehto, and M. Saraste. 1992. SH3-an abundant protein domain in search of a function. *FEBS Lett.* 307:55–61.

- Musacchio, A., M. Wilmanns, and M. Saraste. 1994. Structure and function of the SH3 domain. *Prog. Biophys. Mol. Biol.* 61:283-297.
- Myers, M.W., R.A. Lazzarini, V.M.-Y. Lee, W.W. Schlaepfer, and D.L. Nelson. 1987. The human mid-size neurofilament subunit: a repeated protein sequence and the relationship of its gene to the intermediate filament gene family. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:1617-1626.
- Nakagawa, T., J. Chen, Z. Zhang, Y. Kanai, and N. Hirokawa. 1995. Two distinct functions of the carboxyl-terminal tail domain of NF-M upon neurofilament assembly: cross-bridge formation and longitudinal elongation of filaments. *J. Cell Biol.* 129:411-429.
- Nave, R., D.O. Furst, and K. Weber. 1989. Visualization of the polarity of isolated titin molecules: a single globular head on a long thin rod as the M-band anchoring domain? *J. Cell Biol.* 109:2177-2187.
- Nixon, R.A., and R.K. Sihag. 1991. Neurofilament phosphorylation: a new look at regulation and function. *Trends Neurosci.* 14:501-506.
- Noguchi, J., M. Yanagisawa, M. Imamura, Y. Kasuya, T. Sakurai, T. Tanaka, and T. Masaki. 1992. Complete primary structure and tissue expression of chicken pectoralis M-protein. *J. Biol. Chem.* 267:20302-20310.
- Okagaki, T., F.E. Weber, D.A. Fischman, K.T. Vaughan, T. Mikawa, and F.C. Reinach. 1993. The major myosin-binding domain of skeletal muscle MyBP-C (C-protein) resides in the COOH-terminal, immunoglobulin C2 motif. *J. Cell Biol.* 123:619-626.
- Olson, N.J., R.B. Pearson, D.S. Needleman, M.Y. Hurwitz, B.E. Kemp, and A.R. Means. 1990. Regulatory and structural motifs of chicken gizzard myosin light chain kinase. *Proc. Natl. Acad. Sci. USA* 87:2284-2288.
- Price, M. G., and R.H. Gomer. 1993. Skelemin, a cytoskeletal M-disc periphery protein, contains motifs of adhesion/recognition and intermediate filament proteins. *J. Biol. Chem.* 268:21800-21810.
- Riddle, D.L., and S. Brenner. 1978. Indirect suppression in *Caenorhabditis elegans*. *Genetics.* 89:299-314.
- Rogalski, T.M., B.D. Williams, G.P. Mullen, and D.G. Moerman. 1993. Products of the *unc-52* gene in *C. elegans* are homologous to the core protein of the mammalian basement membrane heparan sulfate proteoglycan. *Genes & Dev.* 7:1471-1484.
- Ron, D., M. Zannini, M. Lewis, R.B. Wickner, L.T. Hunt, G. Graziani, S.R. Tronick, S.A. Aaronson, and A. Eva. 1991. A region of proto-dbl essential for its transforming activity shows sequence similarity to a yeast cell cycle gene, CDC24, and the human breakpoint cluster gene, bcr. *The New Biol.* 3:372-379.
- Ruvkun, G., and J. Giusto. 1989. The *C. elegans* heterochronic gene *lin-14* encodes a nuclear protein that forms a temporal developmental switch. *Nature (Lond.)* 338:313-319.
- Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463-5467.
- Shoemaker, M.O., W. Lau, A.P. Shattuck, A.P. Kwiatkowski, P.E. Matrisian, L. Guerra-Santos, E. Wilson, T.J. Lukas, L.J. Van Eldik, and D.M. Watterson. 1990. Use of DNA sequence and mutant analyses and antisense oligodeoxynucleotides to examine the molecular basis of nonmuscle myosin light chain kinase autoinhibition, calmodulin recognition, and activity. *J. Cell Biol.* 111:1107-1125.
- Shirinsky, V.P., A.V. Vorotnikov, K.G. Birukov, A.K. Nanaev, M. Collinge, T.J. Lukas, J.R. Sellers, and D.M. Watterson. 1993. A kinase-related protein stabilizes unphosphorylated smooth muscle myosin minifilaments in the presence of ATP. *J. Biol. Chem.* 268:16578-16583.
- Smith, D.B., and K.S. Johnson. 1988. Single-step purification of polypeptides expressed in *E. coli* as fusions with glutathione S-transferase. *Gene (Amst.)* 67:31-40.
- Strehler, E.E., E. Carlsson, H.M. Eppenberger, and L.-E. Thornell. 1983. Ultrastructural localization of M-band proteins in chicken breast muscle as revealed by combined immunocytochemistry and ultramicrotomy. *J. Mol. Biol.* 166:141-158.
- Takai, Y., T. Sasaki, K. Tanaka, and H. Nakanishi. 1995. Rho as a regulator of the cytoskeleton. *Trends Biochem. Sci.* 20:227-231.
- Tavare, S., and B. Song. 1989. Codon preference and primary sequence structure in protein-coding regions. *Bull. Math. Biol.* 51:95-115.
- Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA.* 76:4350-4354.
- Vaughan, K.T., F.E. Weber, S. Einheber, and D.A. Fischman. 1993. Molecular cloning of chicken myosin-binding protein H (86-kDa protein) reveals extensive homology with MYBP-C (C-protein) with conserved immunoglobulin C2 and fibronectin type III motifs. *J. Biol. Chem.* 268:3670-3676.
- Vinkemeier, U., W. Obermann, K. Weber, and D.O. Furst. 1993. The globular head domain of titin extends into the center of the sarcomeric M band: cDNA cloning, epitope mapping and immunoelectron microscopy of two titin-associated proteins. *J. Cell Sci.* 106:319-330.
- Waterston, R.H. 1988. Muscle. In *The Nematode Caenorhabditis elegans*. W.B. Wood, editor. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. 281-335.
- Waterston, R.H., J.N. Thomson, and S. Brenner. 1980. Mutants with altered muscle structure in *C. elegans*. *Dev. Biol.* 77:271-302.
- Waterston, R., C. Martin, M. Craxton, C. Huywh, A. Corlson, L. Hillier, R. Durbin, P. Green, R. Shown Keen, N. Halloran, et al. 1992. A survey of expressed genes in *Caenorhabditis elegans*. *Nat. Genet.* 1:114-123.
- Williams, B.D., and R.H. Waterston. 1994. Genes critical for muscle development and function in *Caenorhabditis elegans* identified through lethal mutations. *J. Cell Biol.* 124:475-490.
- Wilson, R., R. Ainscongh, K. Anderson, C. Baynes, M. Berks, J. Bonfield, J. Burton, M. Connell, T. Copsey, J. Cooper et al. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature (Lond.)* 368:32-38.
- Wright, J., Q.-Q. Huang, and K. Wang. 1993. Nebulin is a full-length template of actin filaments in the skeletal muscle sarcomere: an immunoelectron microscopic study of its orientation and span with site-specific monoclonal antibodies. *J. Muscle Res. Cell Motil.* 14:476-483.
- Zheng, Y., R. Cerione, and A. Bender. 1994. Control of the yeast bud-site assembly GTPase CDC42. *J. Biol. Chem.* 269:2369-2372.