Characterization of a Pair of Interacting Genes, *unc-112* and *dim-1*, that Affect Sarcomere Development and Organization

by

MARY MARGARET GILBERT

BA, Scripps College, 1984
M.Sc., The University of Oregon, 1989

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Zoology

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

June 26, 1997

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Zoology

The University of British Columbia
Vancouver, Canada

Date July 10, 1997

DE-6 (2/88)

Abstract

This dissertation focuses on two genes, *unc-112* and *dim-1*, that affect bodywall muscle differentiation in the nematode *Caenorhabditis elegans*. A characterization of these two genes and their gene products will contribute to answering a fundamental question: how does a nematode make a collection of muscle proteins organize into a structure capable of movement? The roles that the *unc-112* and *dim-1* gene products play in the assembly, organization and contraction of sarcomeres in the various subtypes of muscle were examined using genetic, cytological, and molecular techniques. Two phenotypic classes of *unc-112* alleles are known (116,123 and P. Anderson, per. comm.). The viable class allele *unc-112(r367)* demonstrates the importance of this gene in larval development and maintenance of sarcomeres in the gonad arms and muscle cells. The lethal class allele *unc-112(st562)* demonstrates that *unc-112* function in embryonic muscle cells is essential (123). In order to discover the function of UNC112 protein, I cloned and sequenced several cDNAs associated with polymorphisms in two *unc-112* intragenic revertant strains. The cDNA pUNC112-1 is derived from the gene ZC376.2 and identifies a tandem cluster of five highly similar genes covering 19 kb of cosmids T02B5 and ZC376. The predicted proteins of the five gene family share considerable similarity with the cholinesterase subfamily of serine proteases. Maximum parsimony analysis of this gene family with other members of the cholinesterase subfamily from *C. elegans* and other species indicate an ancient origin and perhaps unique function for this gene cluster.

Mutations in a new gene, *dim-1,* have been isolated as suppressors of *unc-112* paralysis. The null mutation *dim-1(ra102)* completely suppresses *unc-112(r367)* single sarcomere muscle defects and partially suppresses bodywall muscle defects. This indicates that *dim-1* interaction with *unc-112* is indirect. In addition, the *dim-1* alleles exhibit a unique muscle phenotype of nested chevrons in some muscle cells. The distribution of mutant versus wild type muscle components have been compared and show that dense body and M-line proteins assemble and remain intact in the mutant strains but are improperly distributed across the cell membrane. A model is proposed showing how *unc-112* and *dim-1* may function in a signaling pathway governing sarcomere patterning in muscle cell membranes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Dedication

To my parents, William and Dorothy Gilbert
To my parents-in-law, Bob and Betty Snutch
and to my partner and best friend, Terry Snutch.

Acknowledgments

I am grateful for the advice, encouragement, inspiration and support of my supervisor, Dr. D.G. Moerman, adjunct advisor, Dr. D.L. Baillie, and supervisory committee members, Dr. M. Adamson, Dr. J. Gosline, Dr. D. Holm, Dr. W. Jefferies, and Dr. L. Matsuuchi. I greatly appreciate the discussions and encouragement of my colleagues J. Bush, Dr. J. Couch, M. Fuse, Dr. J. Hodgson, E. Mathews, G.P. Mullen, Dr. D. Nelson, K. Norman, Dr. T. Rogalski, C. Salinas, and Dr. T-W. Soong.

Chapter 1

Muscle Structure and Development in the Nematode *Caenorhabditis elegans*

1. Muscle development as a model system for understanding determination and differentiation of cells.

The developmental path from a single fertilized cell to a multicellular organism can be divided into two parts: determination and differentiation. Determination is defined as an essentially irreversible process that increasingly specifies fates of cells being generated during embryogenesis (1). The process of cell determination restricts expression of genes not relevant to the specialized or essential metabolic functions via a variety of signaling and transcription-regulating proteins. A pluripotent cell such as a mesodermal cell may be, for example, determined to become a specialized cell such as a skeletal myocyte by limiting gene expression to those transcription factors necessary for producing the components of sarcomere assembly and contraction regulation. In contrast, differentiation occurs in response to cell type-specific transcription factors limiting gene expression to those proteins necessary for sarcomere assembly and contraction regulation (1). The determination of vertebrate presumptive myoblasts begins when the cells are in the somites and ends once the cells have migrated to their final positions (103). The process of differentiation then takes over, fusing the myoblasts into myotubes and organizing skeletal muscle proteins into sarcomeres (103). This dissertation focuses on two genes, *unc-112* and *dim-1*, that affect bodywall muscle differentiation in the nematode *Caenorhabditis elegans*. I have attempted to determine what role the products of *unc-112* and *dim-1* play in the processes of assembly, organization and contraction of differentiated muscle cells.

2. *Caenorhabditis elegans* is well suited for developmental studies.

Caenorhabditis elegans is a free-living soil nematode that is especially well suited for muscle developmental studies using genetic, molecular and cytological techniques (126). This nematode is transparent, has a simple anatomy, essentially invariant cell lineage, small size (less than 1 mm in length at maturity), short life cycle, hermaphroditic and male-mediated methods of reproduction, a small genome estimated at 100 Mb consisting of 5 autosomes and a sex chromosome (105, 126).

The transparency of this small nematode enables observation of muscle cells and all other internal structures in living and fixed animals by bright field, Nomarski, polarized light, or immunofluorescent microscopy. Moreover, the small size of *C. elegans* and the essentially invariant program of development and anatomy have enabled the entire cell lineage from the first cell division in embryogenesis to adulthood to be determined using Nomarski microscopy (62,107,108) and a comprehensive description of all cell types in both sexes to be assembled from serial electron micrographs (121). The transparency of the nematode throughout its life facilitates observation of muscle cell birth, migration and differentiation of embryos in ovo, and of muscle cell growth and contraction in larvae and adults. Using bright field, Nomarski or polarized light microscopy the dynamics of muscle determination and differentiation can be followed in living nematodes. Intact nematodes may be fixed and immunostained to examine the distribution of antigens in whole muscle cells and muscle quadrants. Likewise, Nomarski microscopy can be used to follow each muscle cell during its birth, migration, and differentiation in living animals. The small size of the nematode permits the observation of the consequences a defective protein in a muscle cell has on either the muscle quadrant or the whole nematode in which the defect is detected. How these properties can be used to great advantage to study muscle function can be illustrated using a fictitious group of mutant nematodes that are observed to vary in ability to move from moving awkwardly to paralyzed. These potential muscle gene defects can be sorted into two classes using polarized light microscopy to determine the number and morphology of muscle cells. One class of mutations that affect structural muscle proteins would produce abnormal birefringent muscle patterns. The other class of

mutations would affect movement but not structural muscle proteins and so would produce normal birefringent muscle patterns. Mutations that affect muscles can be categorized into determination or differentiation classes based on the cell lineage patterns. A mutation in a determination pathway gene member may have a non-wild type lineage that produces too few (or too many) muscle cells. Conversely, mutations in the genes that affect the differentiation or function of muscles will have a wild type lineage producing normal numbers of muscle cells that do not function properly. The details of the defective structures can be examined using either polarized light or immunofluorescence microscopy.

The 3 day life cycle of *C. elegans* facilitates faster genetic mapping and characterization relative to other multicellular organisms used in genetic studies, such as *Drosophila melanogaster* (10 days at 25° C) and mice (2 months). Moreover, the short life cycle and small size permit the screening of very large numbers of animals in a short time period with minimal effort. Consequently, the likelihood of obtaining rare genetic variants, such as intragenic recombination events used to obtain fine structure mutational maps (for example, see 83) or phenotypic revertants of a mutated gene (8), greatly increases. In addition, self-fertilizing hermaphrodites permit a variety of non-lethal, morphological and movement-defective mutations to be maintained as homozygotes since mating is not essential for reproduction (39). Males arise spontaneously at a frequency of 0.2% by nondisjunction of the X chromosome and can be used to move mutations to different genetic backgrounds (39). A variety of duplications, deficiencies, and inversions comprise a "toolkit" of aberrant chromosomes and are used for isolating, mapping, and balancing lethal and non-lethal mutations (37).

In over two decades of study, a detailed genetic linkage map has been developed for *C. elegans* consisting of approximately 1400 genes (54). About 50 of these genes are involved in the development of bodywall, pharyngeal, and sex-specific muscle cells and some of these have been cloned (4). The physical linkage map consists of cosmid (23) and yeast artificial chromosome (YAC) (24) containing overlapping genomic clones covering each *C. elegans* chromosome nearly completely (24). In addition, the nucleotide sequence of the cloned DNA is being determined and added to ACEDB (the *C. elegans* database) and other databases (105). The availability of genomic clones and sequence data should greatly simplify the molecular mapping and cloning of any gene of interest. Moreover, the *C. elegans* genome sequencing project provides insight into the possible function of any gene that has been

molecularly mapped to a sequenced chromosome by providing similarities with other gene sequences from different species and within the same species (54).

3. Comparison and contrast of nematode with vertebrate striated muscle structure.

Striated muscle derives its name from the repeated pattern of sarcomeres it contains. A single vertebrate sarcomere consists of a myofibril-wide band of thick filaments flanked by and overlapping with two bands of thin filaments (103). Many sarcomeres are laid end-to-end down the length of the myotube generating light bands (I-bands) and dark bands (A-bands) that are observed using a light microscope (103). Striated muscle is an ideal structure in which to study determination and differentiation using a variety of strategies. First, the highly ordered and repetitive nature of striated muscle enables detection of a broad range of mutant phenotypes from those which fail to produce striated muscle that is easily observed to those which produce uncoordinated movement (8,20,116,117,123). Second, the proteins of which the sarcomeres are composed are relatively abundant which facilitates immunohistochemical and biochemical studies (for example, see 40,90,96,115,122).

Many vertebrate structural components of a sarcomere have been identified and the basic architecture and contractile mechanism determined biochemically and morphologically. The I-bands consist of thin filament proteins and the Z-line membrane-attachment proteins (see figure 1a) (103). The thin filaments consist of polymerized dimers of actin, decorated with troponins and tropomyosin to regulate actin binding to myosin, and capping proteins. Thin filaments are anchored in position to the membrane at the barbed end by the Z-line complex which includes vinculin and α-actinin. Adjacent Z-lines are held in register by the intermediate filament protein, desmin. Overlying the thin filaments and inserting into the Z-line is nebulin, hypothesized to be a protein-ruler that determines the length of the thin filaments (66,68). The thick filaments extend from M-line membrane-attachment complex to overlap with the thin filaments in the A-bands (see figure 1b) (103).

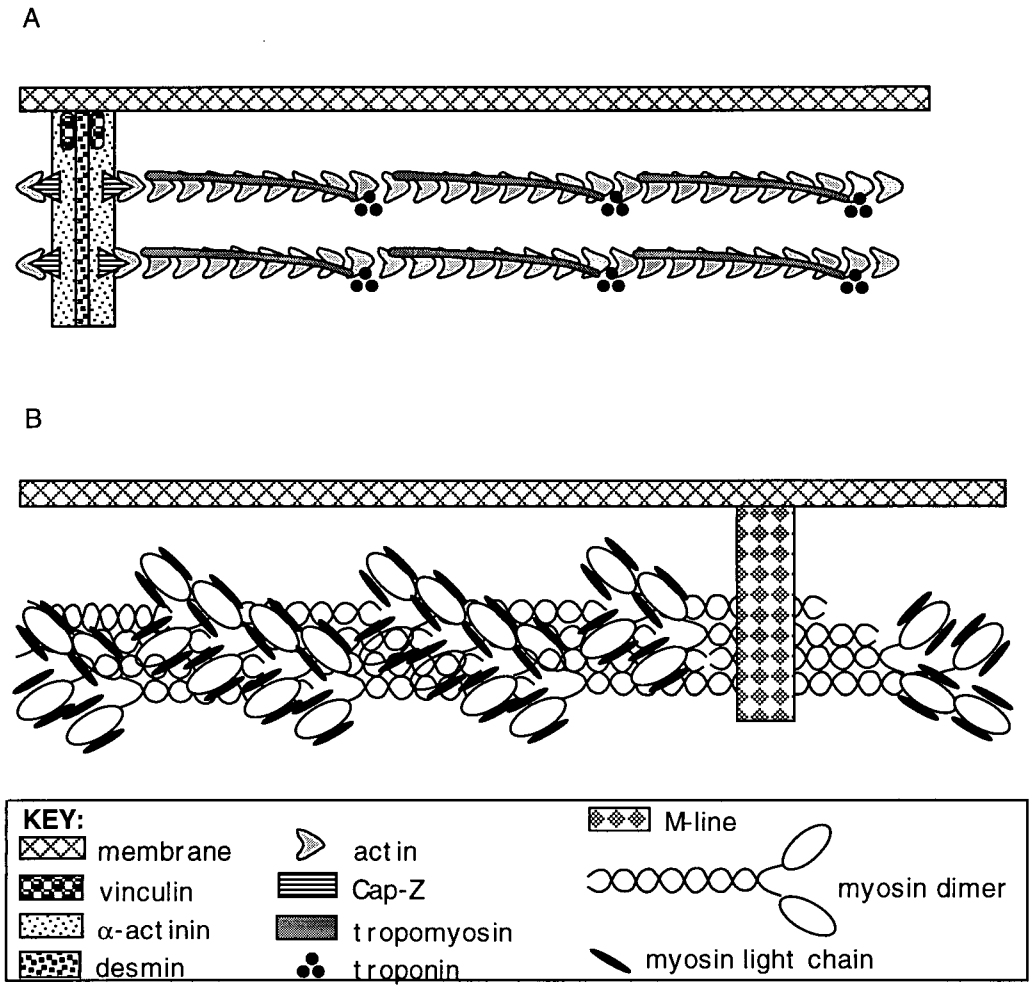


Figure 1: Location of vertebrate Z-line, M-line, thin and thick filament proteins.

In A) the Z-line attaches the thin filaments to the myotube membrane. The Z-line consists of vinculin and α-actinin. Adjacent Z-lines are held in register by desmin filaments. Actin filaments are capped at the barbed ends by CapZ and bound to both vinculin and α-actinin. The binding of actin molecules to myosin head domains is regulated in skeletal muscle by tropomyosin and the troponin T, I and C complex. In B) the M-lines attach the thick filaments to the myotube membrane. The thick filaments are composed of coiled-coil dimers of myosin tails with the myosin head domains protruding from the surface. The dimers associate in parallel except in the vicinity of the M-line where they associate in antiparallel array. This region of the thick filament is devoid of myosin heads. In smooth muscle, the binding of the myosin heads to actin filaments is regulated through the phosphorylation of myosin light chain proteins.

The thick filament primarily consists of myosin dimers associated into filaments via their coiled-coil rod domains. The rod domains of myosin are associated in antiparallel array flanking the portion of the filament proximal to the M-line and in parallel array along the distal portions. The myosin heads project from the filament and are decorated with myosin light chain proteins. The different myosin light chain proteins modulate the ATPase and actin-binding activities and are the substrates for myosin light chain kinases and phosphatases. The center of the thick filament is anchored in position to the membrane by the M-line complex. Titin extends from the M-line to the Z-line in association with the thick filaments and has been proposed to act as a protein ruler to which the sarcomere is compared and adjusted to the same length (39,67-69,114). Other proteins that associate with the M-line, titin, and myosin include myomesin (88), M-protein (88,113), and C-protein (38,113).

Muscle contraction is produced by the movement of the thin and thick filaments past each other utilizing ATP as the energy source (see figure 2) (103). Myosin hydrolyzes ATP to ADP and P_i and then binds to actin. The myosin undergoes a conformational change that pulls the thin filament in the direction of the M-line and releases ADP and P_i. A new molecule of ATP binds to myosin and the actin-myosin linkage is broken. This process of linkage, movement, and release is regulated by calcium-dependent molecules of the troponins and tropomyosin complex that sterically block actin from associating with myosin, or via myosin light chain kinases which alter the conformation of myosin via phosphorylation of the myosin light chains to prevent binding with actin. The regulation of contraction takes place on the thin filaments via the troponins and tropomyosin complex in skeletal muscle, and on the thick filaments via myosin light chain kinases in smooth muscle.

C. elegans bodywall muscle shares many biochemical and morphological properties with vertebrate striated and non-striated (smooth) muscles (103,116). *C. elegans* bodywall muscle cells, like cardiac or smooth muscle, remain mononucleate and do not fuse into myotubes. Many of the nematode muscle proteins are very similar to vertebrate muscle proteins in sequence, structure and distribution. However, *C. elegans* bodywall muscle differs from vertebrate muscle in five respects: 1) sarcomeres are localized to a shallow region beneath the muscle cell membrane in apposition to the hypodermis (the cuticle-secreting "skin" of nematodes), 2) T-tubules are absent (which carry the signal for sarcomeres to contract deep into the myotube), 3) Z-lines are replaced by discontinuous, finger-like projections called dense

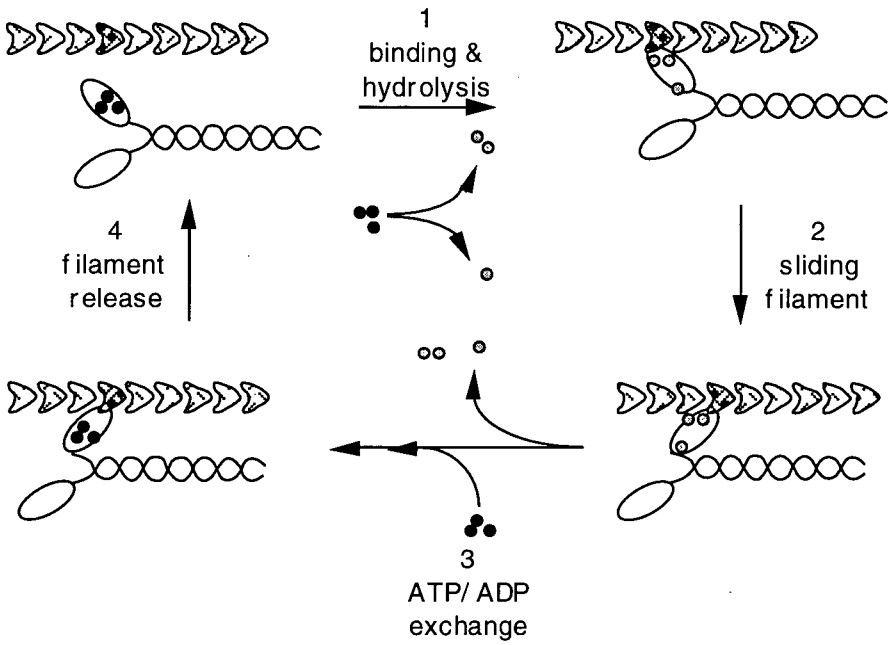


Figure 2: The force-generating crossbridge cycle of muscle contraction.

The cycle begins at 1 with the hydrolysis of ATP (3 solid circles) to ADP and P_i (gray circles), which remain bound to the myosin head (see key in Figure 1), and association with an actin molecule in a nearby thin filament. Step 2 is the conformational change in myosin that moves the thin filament towards the M-line. Step 3 is the release of ADP and P_i followed by the binding of a new molecule of ATP. Step 4 is the dissociation of the actin from the myosin head.

bodies, 4) thin and thick filaments are five-fold longer, and 5) paramyosin and twitchin are added to the thick filaments (116).

The differences in muscle structure and composition reflect the differences in lifestyle between *C. elegans* and vertebrates (103,116). Vertebrates have internal skeletons with muscles attached on both sides of a pivoting joint. The sarcomeres of vertebrate striated muscles fill the myotubes and are laid end-to-end the length of the myotube (see figure 3a); (103). The cross striations are aligned laterally between adjacent myotubes (103). The vector of contraction is directed down the length of the myotubes to shorten the muscle and alter the relative angle of the bones at the joint. In contrast, *C. elegans* has an external cuticle "skeleton" of collagen secreted by the hypodermis and no joints. The dense bodies and M-lines project only 1-2 μm into the cells from the muscle membrane adjacent to the hypodermis (see figure 3b); (116). The lateral attachment of the sarcomere array to the apical surface of the bodywall muscle cells alters the transmission of the force of contraction from parallel (seen in the vertebrate skeletal muscles) to perpendicular in relation to the muscle quadrant. Sarcomeres are offset by ~5° across the width of the cell (116) possibly to maximize the number of sarcomeres occupying the available volume. Viewed under a polarized light microscope this generates an obliquely striated pattern (see figure 5). The striations are aligned at the ends of adjacent cells maintaining the pattern the length of the quadrant.

Electron microscopy of *C. elegans* sections show that the length of thin filaments is five times longer, ~6 μm compared to ~1 μm typical of vertebrates, as is the length of thick filaments, ~10 μm compared to ~1.6 μm (116). Perhaps as a consequence of the increased filament length, paramyosin is incorporated into the thick filaments as structural reinforcement. The paramyosin protein sequence is similar to the rod portion of myosin but lacks actin-binding and ATPase activities (59). When paramyosin is absent from, or altered in muscle cells the result is flaccid paralysis and disorganized thick filaments (116).

Vertebrate muscle and nematode bodywall muscle also differ in the regulation of contraction. In *C. elegans*, contraction is regulated on both thin and thick filaments. Thin filaments are regulated by troponins and tropomyosin as they are in vertebrates (4,116). The thick filaments may be regulated by the myosin light chain kinase domain of twitchin (11,12). Twitchin resembles vertebrate titin in primary

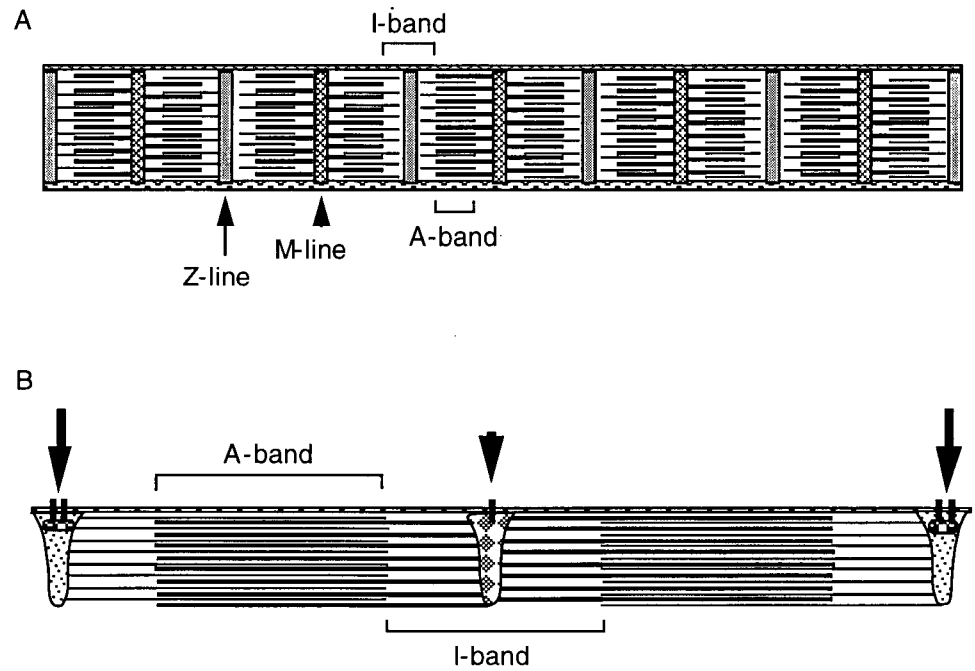


Figure 3: Comparison of vertebrate and nematode muscle structure.

A) is a representation of a vertebrate myotube in tangential section. The sarcomeres are laid end to end and the Z-lines (arrow), M-lines (arrowhead), I-bands and A-bands span the width of the myotube. **B**) is a representation of a *C. elegans* sarcomere in radial section. Dense bodies (arrow) and M-lines (arrowhead) attach the thin and thick filaments respectively to the membrane. At the base of the dense bodies vinculin (beaded ellipses) lies adjacent to the intracellular face of the membrane. Alpha-actinin (small squares) is distributed throughout the dense body projection away from the membrane. The region of thin and thick filament overlap constitutes the A-band.

structure: both are composed of numerous fibronectin type III and immunoglobulin domains and a myosin-specific kinase domain at the amino terminus (11,12,67,69). However, twitchin is not long enough to extend from the M-line to the dense body of the much larger invertebrate sarcomere (4,11,12). Furthermore, twitchin is the only myosin light chain kinase yet identified in *C. elegans* and so is more likely to regulate contraction rather than participate in determining the length of the sarcomere.

4. Comparison of nematode muscle development with that of vertebrates.

Determination of the vertebrate myogenic pathway is thought to be initiated by a family of transcription factors that includes MyoD1, Myf-5, MRF4 and myogenin (100). Members of this small gene family contain a region with homology to c-*myc* which forms a basic helix-loop-helix DNA binding motif (29). These transcription factors bind at specific times and sites in the 5' (7,16,94) and 3' (27) sequence flanking muscle protein-encoding genes and activate expression in a subset of cells in the somites (100). The presumptive myoblasts then migrate from the somites to their final positions by following cues in the extracellular matrix (ECM) that specify the dorsal-ventral, anterior-posterior, and left-right axes (100). For example, the left-right positioning of cardiac and visceral primordial cells in *Xenopus laevis* embryos is randomized if heparinase or arginine-glycine-aspartic acid (RGD) peptide is microinjected into the blastocoel (127). These treatments disrupt the deposition of heparin sulfate proteoglycans and fibronectin, respectively, which are components of the ECM overlying the cardiac and visceral primordia, and the migrating cells of each tissue type make the decision to turn left or right randomly (127). Other factors may specify the final destination of the myoblasts and trigger the terminal phase of differentiation.

Only one component of muscle determination has been characterized in *C. elegans*. The expression of the nematode homologue of myoD1, called CeMyoD, is first detected by immuno-fluorescence approximately 150 minutes post primary cleavage (ppc) at 20 °C in 2 cells of the D lineage, which produces only bodywall muscles (64). At this stage the presumptive D myoblasts are 3 to 4 cell divisions from cessation of mitosis and are not expressing myofilament proteins (see figure 4). CeMyoD is

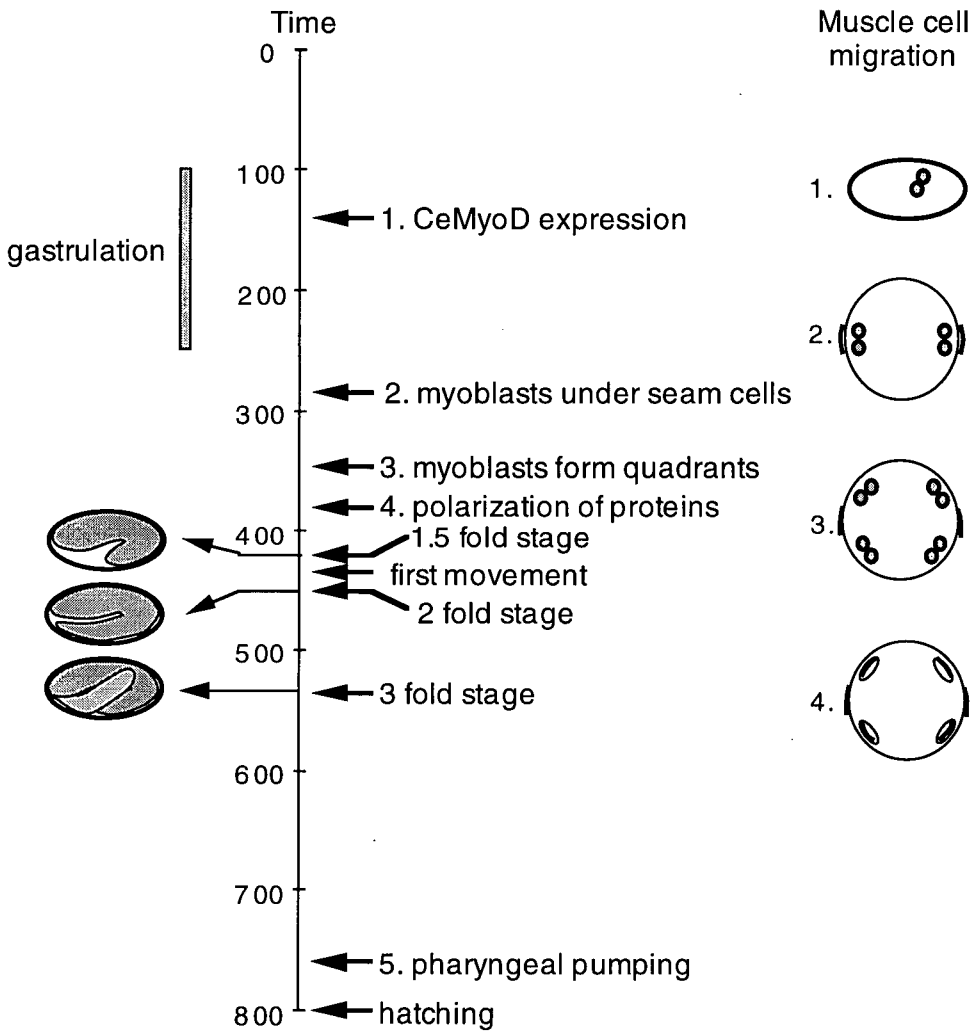


Figure 4: Time course of* C. elegans *muscle cell differentiation.

The length of time from the first cleavage to hatching is represented by a vertical bar marked every one hundred minutes. Near the midpoint of gastrulation (~150 min.) the expression of CeMyoD is detected in the two D lineage cells which will generate body wall muscle cells exclusively. By ~290 min. 32 myoblasts from the C, D and MS lineages have migrated into rows underlying each row of hypodermal seam cells and are beginning to express muscle-specific proteins. Half of the cells in each row either migrate dorsally or ventrally into their final positions by 340 min. In the meantime, 33 more myoblasts are born and added to the muscle quadrants. By 400 min., the final group of 16 embryonic myoblasts are added to the quadrants for a total of 81. The muscle cells assume a flattened spindle shape and the muscle-specific proteins accumulate in the region of the muscle cytoplasm adjacent to the hypodermis by 420 min. By ~440 min. the body wall muscle sarcomeres are functional and the embryo begins to twitch. In contrast, the pharynx does not function until 760 min. By hatching at 800 min. 2 A-bands per body wall muscle cell can be detected.

not the only myogenic transcription factor in *C. elegans*. Additional factors, which await identification, appear to be involved since a null allele of the gene encoding CeMyoD protein, *hlh-1*, does not prevent differentiation of bodywall muscle (19,20). However, the animals are inviable due to poor muscle function (19,20). From the time that CeMyoD is expressed through 420 minutes ppc, three groups of myoblasts are born and migrate to form the muscle quadrants (108). The first group of 32 myoblasts arising from the D, C (which produces hypodermal, neuronal and bodywall muscle cells), and MS (which produces anterior muscle, neuronal, gonadal and intestinal cells) lineages form a row on either lateral surface of the embryo just beneath the hypodermal seam cells and begin to express muscle proteins such as vinculin, β-integrin, and myosins A and B by approximately 290 minutes (43,56,108). The second group of 33 myoblasts are born between 290 to 340 minutes (108). Following the birth of the second group of myoblasts, one half of the myoblasts lying under each row of hypodermal seam cells migrates dorsally and the other half ventrally between 290 minutes to 350 minutes (56,108). The final group of 16 embryonic myoblasts are born and added to the quadrants between 340 and 420 minutes (108).

C. elegans bodywall muscle differentiation is known in greater detail than vertebrate muscle development due to the availability of a panel of monoclonal antibodies that recognize many muscle cell, hypodermal and basement membrane antigens (36), many genetic mutations (see Table 1) and a lineage map of embryonic cell fates (108). A panel of monoclonal antibodies has been raised from an extract of muscle cell and basement membranes to aid in the morphological analysis of wild type and mutant *C. elegans* dense bodies, M-lines, and hypodermal attachment structures (36). The monoclonal antibodies MH4, MH5 and MH46 have been used to correlate the changing distributions of the corresponding hypodermal or basement membrane proteins in wild type and mutant developing embryos with myoblast migrations and sarcomere formation (8,56,123). MH4 recognizes two intermediate filament proteins of 68 and 70 kDa (37). MH4 stains the hypodermis overlying the bodywall muscle quadrants in bands perpendicular to the sarcomeres, the sites of contact between vulva, uterine or anal depressor muscles and the hypodermis, and non-muscle cells within the pharynx (37,56). MH5 recognizes a 300 kDa component of hypodermal hemidesmosomes (called fibrous organelles in *C. elegans*) (37). The staining pattern of MH5 resembles MH4 over the muscle quadrants, but stains only the lumenal and basal surfaces of the pharyngeal marginal cells (37,56). MH46 recognizes a 380 kDa basement membrane

TABLE 1

MUSCLE-AFFECTING GENES IDENTIFIED BY MUTATION AND/OR MOLECULAR CLONING

GENE	PROTEIN	PHENOTYPE	REFERENCE
unc-15	Paramyosin	Flaccid paralysis	(103)
unc-45	phosphatase	Pat* lethal or flaccid paralysis	(D. Pilgrim and L. Venolia, per. comm., and 103)
unc-54	MHC B	Flaccid paralysis	(103)
unc-82	unknown	Slow movement	(103)
unc-89	mini-titin	Good movement, no M-line	(13,103)
myo-3	MHC A	Pat lethal	(104)
lev-11	tropomyosin (tentative)	uncoordinated movement, mild twitch, grows well in levamisole	(107)
pat-5/ egl-19	Calcium channel	Pat lethal	(L. Avery, per. comm., and 107)
pat-10	troponin C (tentative)	Pat lethal	(107)
unc-22	twitchin	Slow movement with twitch	(10,77)
unc-90	unknown	Rigid paralysis	(103)
unc-93	unknown	Rubberband movement	(71,103)
unc-105	degenerin-like	Rigid paralysis	(82,103)
sup-10	unknown	Suppressor of *unc-93(e1500)*, otherwise wildtype in appearance	(103)
mup-2	troponin T	Larval lethal, kinked bodies	(4,5)
act-1- act-3	actin	Nearly paralyzed, slow growing	(103)
unc-60	actin-binding protein	embryonic lethal and viable Unc alleles	(80)
unc-78	unknown	Slow movement	(103)
unc-94	unknown	Slow movement	(103)
deb-1	vinculin	Pat lethal, antigen for MH24	
pat-8	unknown	Pat lethal	(107)
pat-9	unknown	Pat lethal	(107)
pat-12	unknown	Pat lethal	(107)
unc-23	unknown	Weak forward movement	(103)
unc-52	perlecan	Pat lethal or flaccid paralysis, antigen for MH2 & MH3	(39,90,107)
let-805	basement membrane factor	Pat lethal but not completely paralyzed with muscle quadrants detached	M. Hresko, per. comm.
unc-95	unknown	Nearly paralyzed	(103)
unc-97	unknown	Flaccid paralysis	(103)
unc-112	unknown	Pat lethal or flaccid paralysis	(103,107)
mua-1	transcription factor	Progressive paralysis from quadrant detachment	(5,85)
mua-2	unknown	unknown	(5)
mua-3	novel matrix receptor	Progressive paralysis from quadrant detachment	(5,85)
mup-1	unknown	unknown	(5)
mup-3	unknown	unknown	(5)
pat-2	α-integrin (tentative)	Pat lethal	(107)

pat-3	β-integrin	Pat lethal, antigen for MH25	(107)
emb-9	a2 collagen IV	Late embryonic lethal	(103,107)
let-2	collagen IV	Embryonic lethal	(103,107)
unc-96	unknown	Somewhat slow movement	(103)
unc-98	unknown	Slow movement	(103)
unc-87	calponin-like	Nearly paralyzed	(40,41,103)
unc-27	unknown	Weak backward movement, slow	(103)

* Pat lethal phenotype is described on page 25.

protein that contains many repeats of the fibronectin type III motif and identifies the antigen as a candidate signaling molecule for migrating muscle cells (M. Hresko, per. comm., (37)). MH46 generates a staining pattern over the muscle quadrants similar to MH4 and MH5, but stains only the basal surfaces of the marginal and muscle cells of the pharynx (37). The similarity of the staining patterns of MH4, MH5, and MH46 over the bodywall muscle quadrants in adults suggests that the proteins form a complex that extends from the basement membrane between the muscle cells and the hypodermis (MH46), to the hypodermal hemidesmosomes (MH5), and terminating in the hypodermal intermediate filament network (MH4).

The staining patterns of MH4, MH5 and MH46 antibodies change during the interval from 290 to 420 minutes ppc and correlate with the migrations of the myoblasts (56). The proteins recognized by these antibodies allow the migration process and initiation of the formation of sarcomeres at the muscle cell surface proximal to the hypodermis to be monitored (56). All three proteins are diffusely distributed in dorsal and ventral hypodermis, but not in the hypodermal seam cells prior to 290 minutes (56). MH4 and MH5 staining redistributes into patches proximal to the myoblasts with initiation of migration of the myoblasts into subdorsal and subventral quadrants by 310 minutes (56). By 350 minutes MH46 and GM1 (a polyclonal antibody that recognizes all isoforms of perlecan) co-localize in the basement membrane between the hypodermis and bodywall muscle quadrants (56,84). Between 350 and 420 minutes MH46 antigen is organized into oblique, punctate striations similar to sarcomere striations (56). In this same time interval, the muscle proteins are redistributed within the muscle cells just beneath the basement membrane and begin assembling into sarcomeres starting with transmembrane β-integrin (56). By approximately 430 minutes, the embryonic muscles are capable of twitching and by 450 minutes, the embryo can roll in the eggshell (56). After 520 minutes the MH46 staining pattern assumes the perpendicular conformation seen in adults (56). When the nematode hatches after 800 minutes there are 81 muscle cells each of which is two A-bands in width (125). By the end of larval stage L1 the animal has acquired 14 more cells, attaining the adult complement of 95 bodywall muscle cells. The second phase of differentiation extends from larval stage L3 to adulthood and results in the expansion in number of dense bodies and M-lines, and the number of thin and thick filaments across the depth, length, and width of each muscle cell.

The assembly and placement of the Z-line and M-line complexes is crucial for determining both the orientation and the pattern of sarcomeres in vertebrates, insects and nematodes. The attachment components are among the first proteins to be expressed by the vertebrate myocytes (32,33,79,90). Similarly in invertebrates, mutations in the gene encoding the β subunit of integrin from Drosophila, *lethal(1)myospheroid* (74,87), and from *C. elegans*, *pat-3* (41), result in absence of organized myofilaments and are lethal early in development. The gene products of *unc-52* from *C. elegans* are extracellular proteins associated with the dense bodies and M-lines. Null mutations of *unc-52* also result in absence of organized myofilaments and embryonic lethality (98,123). Both *unc-52* and *pat-3* lethal mutations belong to the earliest group of myofilament organization affecting mutations identified in *C. elegans* (56,123). Assembly and placement of vertebrate membrane-bound and -associated components then initiates the independent formation of I-band and A-band structures and their interdigitation to form mature sarcomeres. Ward and Hammer (79) examined the formation of Z-lines cytologically using rhodamine labeled α-actinin and found that Z-line formation begins in chick myotubes with the diffuse aggregation of future Z-line components. These gradually consolidate into Z-bands and then Z-lines, which nucleate the formation of thin and thick filaments into functional sarcomeres. At the same time the variability in sarcomere length decreases with time as myofibrils shift positions until all Z-lines are in lateral register. The limitation of these vertebrate studies is the inability to perturb the assembly process and then go beyond the observable changes in morphology to discover the mechanism driving the assembly process.

Several mutants affected in the assembly of muscle components into functional sarcomeres in *C. elegans* have been followed morphologically with a subset of the monoclonal antibodies (8,56,123). Null mutations in β-integrin and perlecan genes, *pat-3* and *unc-52*, show disrupted antibody staining patterns of actin, MHC B, MH5, MH46 and MH42 (an M-line-specific marker) but very little disruption of the MH4 staining pattern (56,123). Null mutations in the genes for vinculin and MHC A, *deb-1* and *myo-3*, show disruption in actin and MHC B staining patterns; however, the staining pattern for β-integrin is much less disrupted (8,56,117). Likewise, the onset of developmental arrest in embryos mutant for *pat-3* and *unc-52* occurs earlier than in *deb-1* and *myo-3* mutant embryos: approximately 420 minutes ppc for *pat-3* and *unc-52*, L1 for *deb-1*, and approximately 450 minutes ppc for *myo-3* (8,117,123). These results indicate

that mutations that disrupt sarcomere components located closest to the muscle cell membrane appear to cause more extensive disruption than mutations in components more distally located. This is consistent with the vertebrate immunofluorescent staining results that showed formation of the Z-lines occurs prior to filament assembly and interdigitation (73,79). It is interesting to note that twitchin, a titin-like protein encoded by the *unc-22* gene, does not appear to be essential to this stage of the developmental process as suggested for vertebrate titin (40,67), because mutations in this gene have no effect on filament assembly and so do not disrupt this developmental stage.

The arrest of muscle differentiation followed by arrest of elongation in developing *C. elegans* embryos was first observed in *myo-3* mutant animals (117). The phenotype was called Pat (for paralyzed and arrested at two fold) and was used to select mutations in other genes whose products are also required at this early phase of muscle differentiation (123). Five classes of mutants were isolated: class I mutants show the most severe disruption of both M-lines and dense bodies; class II mutants show disruption of dense bodies and less severe disruption of M-lines; class III mutants appear to disrupt thin filament organization preferentially; class IV mutants are severely paralyzed but do not show much disruption of the filament assemblies; and class V mutants appear to disrupt thick filament organization preferentially (123). The first proteins that appear to be temporally required in the development of dense bodies and M-lines are encoded by the genes *unc-112*, *unc-52, pat-2,* and *pat-3* (123). Mutations in these genes show the most severe disruption of the actin and myosin antibody staining patterns and the earliest arrest of muscle differentiation (123). Of these four genetic mutants, *unc-112* lethal mutant embryos arrest muscle development earliest (B. Williams, per. comm.), yet the nature of the gene has not been determined. The products encoded by the other 3 genes have been determined: *pat-2* encodes an α-integrin (B. Williams, per. comm.), *pat-3* encodes β-integrin (41,123), and *unc-52* encodes several perlecan isoforms (98). Thus, the study of muscle development in *C. elegans* has lead to the identification of new gene products involved in this process that can be integrated into a detailed functional pathway based on analysis of mutations in these genes and eventually to an accurate model of the process of muscle cell differentiation.

5. History of the study of unc-112.

The identification and classification of an *unc-112* mutation as a class I Pat was only the second instance in which a mutation has been obtained in this gene. The first mutation in *unc-112* was identified in a search for suppressors of *unc-105(n490)* animals (P. Anderson, per. comm.). Examination of *unc-112(r367)* paralyzed animals using polarized light microscopy demonstrates that this phenotype is due to the progressive collapse of organized sarcomeres in the bodywall muscles (116). It is likely that *unc-112* interacts very indirectly with *unc-105,* as the hypercontracted state of sarcomeres would be relieved by any disruption of the structure, including detachment of the sarcomeres from muscle cell membranes. However, a suppressor screen such as this is probably the only way to identify a viable mutant allele of *unc-112*, as the majority of mutations in this gene are predicted to be embryonic lethals (see Ch. 2).

unc-112 mutants exhibit similar phenotypes to *unc-52* mutants. First, both genes have viable, fully penetrant alleles with similar phenotypes. The viable class of *unc-112* consists of the *r367* allele that gradually causes paralysis as do *unc-52* viable mutations. Second, paralysis appears to be caused by detachment of the thin and thick filaments from the muscle cell membranes in both *unc-112* and *unc-52* viable mutants. Last, class I Pat mutant alleles have been identified for both genes. Both *unc-112(st562)* and *unc-112(st581)* cause very early arrest of muscle differentiation about 420 minutes post-fertilization (the 1.5-fold stage of embryogenesis) as does *unc-52(st549).* Based on this apparent similarity to *unc-52,* the product of the *unc-112* gene may also have a role in either forming the attachment structures, positioning them, or both. In addition, the three members of Pat class I whose gene products have been identified are all localized within the muscle cell membrane (the integrins from *pat-2* and *pat-3*) or in very close proximity to it (perlecan from *unc-52*). It thus appears that the *unc-112* gene product is also likely to be located within or very close to the muscle cell membrane, have an essential function very early in muscle differentiation, and affect the formation of both dense bodies and M-lines. These possibilities make the study of this poorly characterized gene very compelling. Consequently, I have undertaken the characterization of *unc-112* and its product in order to understand its function.

Chapter 2

Genetic and Cellular Characterization of *unc-112* and a Suppressor, *dim-1*

1. Introduction

A. Phenotypic analysis of mutant muscle structure.

Phenotypes of the two known classes of *unc-112* mutants provide a basis for formulating mechanisms by which the product of this gene may function in *C. elegans*. The two classes are embryonic lethal, represented by *unc-112(st562)* (123), and larval paralysis, represented by *unc-112(r367)* (10). The developmental stage at which the mutant phenotypes are manifest indicates the temporal requirement for the gene product and the consequence of inactivity or abnormal activity of the gene product. Thus, the phenotype of paralysis and developmental arrest around the 1.5 fold stage of embryogenesis (Pat phenotype) of *unc-112(st562)* animals indicates the normal gene product is required before or at this early stage of embryonic muscle development. Similarly, the progressive, flaccid paralysis phenotype beginning in larval stage L3 of *unc-112(r367)* animals indicates a further requirement for the gene product midway through larval development. The two classes of *unc-112* alleles also suggest a gradient of increasing activity or amount of UNC112 protein. If *unc-112(st562)* is presumed to be a null allele, then the absence of functional UNC112 results in death mid-embryogenesis. If *unc-112(r367)* produces less than wildtype levels of UNC112 but more than zero, then death in mid-embryogenesis is avoided but the level of UNC112 is not sufficient to avoid mid-larval paralysis.

Within muscle cells the phenotypes of both classes of *unc-112* mutants are similar: *unc-112(st562)* mutant muscles exhibit stellate clumps of immunostained actin and myosin (123) and *unc-112(r367)* mutant muscles exhibit stellate clumps of birefringent A-bands (116). The disorganization of sarcomere structure is the most likely cause of paralysis in both *unc-112(st562)* and *unc-112(r367).* However, demonstrating that the embryonic process and the larval process share common features does not imply that these processes are identical. Indeed, the two processes may be fundamentally different since the embryonic process initiates the de novo formation of the first sarcomeres whereas the larval process specifically affects the growth and pattern formation of rows of sarcomeres in bodywall muscle cells. I am interested in the development of muscle structures and how the decision of sarcomere placement is made in both phases of muscle development. Alleles of *unc-112* appear to affect both structure and placement, therefore a better understanding of this gene may address both issues of interest.

The observations of disrupted thin and thick filaments in both classes of mutants suggest that *unc-112* affects either assembly of muscle components into sarcomeres, or organization of sarcomeres into functional arrays. There are four possible mechanisms by which *unc-112* mutations could produce these effects: two with *unc-112* gene product functioning as an assembly catalyst and two with *unc-112* gene product functioning as an integral structural protein. Mechanism 1) the *unc-112* gene product may have a role in assembling a portion of the sarcomere structure. Consequently, a failure to express either the essential embryonic isoform or the adult isoform in bodywall muscle cells would result in incomplete assembly of the sarcomere components and collapse of the partially assembled structures. Mechanism 2) the *unc-112* gene product is required for organization of dense bodies and M-lines in muscle cell membranes in order to effectively manage contraction stress. As a result of functional failure, the sarcomeres are pulled out of the membranes by the force of muscle contraction and the structure collapses. Mechanism 3) the *unc-112* gene product is a crucial structural component of the sarcomere. Therefore failure of the sarcomeres to form results in dispersion of the remaining components throughout the muscle cells. Mechanism 4) *unc-112* may either encode an enzymatic or a structural protein that when defective has a deleterious secondary effect on a crucial protein with which it interacts. For example, mutant UNC112 protein may another member of the protein complex susceptible to degradation.

A chain reaction of exposure and degradation would result in the destruction of the entire complex as is seen with mutations in the dystrophin complex in vertebrates (31).

In this thesis, characterization of bodywall muscle phenotypes of *unc-112(r367)* and suppressor mutants were used to identify the changes in the organization of sarcomeres and provide evidence for the specific roles *unc-112* and its suppressor *dim-1* may play in their formation. Polarized light microscopy and indirect immunofluorescence staining using antibodies to vinculin, β-integrin and perlecan were used to determine the distribution of proteins in wild type and mutant animals. This allowed identification of sites within sarcomeres requiring UNC112 and any suppressor proteins. Polarized light is rotated by highly ordered structures such as thin and thick filaments and appears as birefringent patterns on a dark gray background. Changes in the organization of these structures leads to diminished signal intensity and an altered birefringent pattern. In *C. elegans* bodywall muscles dense bodies and A-bands are strongly birefringent and so polarized light microscopy can be used to identify changes in pattern and morphology of these structures. The distribution of specific components of dense bodies, M-lines, myofilaments and extracellular basement membrane can be determined by immunostaining with specific antibodies and the pattern compared with the polarized light pattern. For example, if *unc-112* or any suppressor gene products are integral structural components of the dense bodies and M-lines, then this may alter the association of other components in these structures. This may have the effect of preventing another component from associating with maturing dense bodies, and instead the product would remain diffuse in the cytoplasm. The diffuse distribution of the protein can be detected by immunostaining with an antibody to the displaced component and comparing the mutant pattern to the wildtype pattern.

B. Genetic analysis of mutants

In order to understand the effects that mutant *unc-112* alleles have on muscle development it is important to identify other genes with which *unc-112* interacts. Genetic suppression analysis is a powerful method for examining gene interactions in *C. elegans* (49,52). Several mechanisms of suppression have been observed in *C. elegans* including informational suppression (53,118), gene duplication (77),

intragenic rearrangement (70,120) and epistasis (85,104). The types of suppression that are of primary importance to this study are epistasis and intragenic reversion. The latter type is important for the molecular analysis of *unc-112* and will be discussed in Chapter 3. Epistasis is the ability of a mutant gene, such as *unc-54* encoding MHC B, to mask the phenotype of another mutant gene, such as *unc-22* encoding twitchin, and is used to assemble developmental and enzymatic pathways (49,58,85). *unc-22* mutations cause constant twitching of the bodywall muscles, which can be suppressed by specific missense *unc-54* alleles (85). A unique class of *unc-54* mutations, located in the globular head domain of MHC B, suppress a variety of *unc-22* alleles and result in near-normal movement and the absence of twitching. In contrast, the general class of *unc-54* mutations including *e190, e1301* and *e1152,* located in regions carboxyl to the head domain, result in paralysis and the absence of twitching. In both cases, the suppression of *unc-22* mutant twitching by *unc-54* mutants demonstrates that *unc-54* is epistatic to *unc-22*. The *unc-54* gene product MHC B acts in the terminal step of the bodywall muscle contraction/ relaxation biochemical pathway. Suppression of *unc-22* by *unc-54* indicates that *unc-22* functions upstream of *unc-54* in the contraction/relaxation pathway. This hypothesis was confirmed when the sequence of *unc-22* revealed a myosin light chain kinase domain in the amino terminal portion of the predicted protein (11,12).

The process of assembling a genetic pathway typically begins with the isolation of intergenic revertants of a chosen mutation. For example, mutagenic reversion of *unc-112(r367)* could provide information on the nature and identity of genes that interact with *unc-112*. Both new alleles of the *unc-112* gene and mutations in interacting genes may be obtained after ethylmethane sulfonate (EMS) mutagenesis of *unc-112(r367)* followed by selection of *unc-112(r367)* progeny that move better or are older when paralysis occurs. The number of intergenic suppressor mutations identified by reversion is related to the number of animals examined in the reversion screen and the nature of the mutation being reverted. As a rough estimate, examination of 50,000 to 100,000 animals are predicted to yield between 25 and 50 mutations based on the rate of 1 in 2000 for knock-out mutations in "average sized" genes between 1 kb and 3 kb (49).

New mutations arising from EMS mutagenic reversion of *unc-112(r367)* may be located either in another gene (intergenic suppressor) or in the same gene as the original mutation (intragenic suppressor).

Assigning the intergenic mutations to complementation groups indicates the number of genes that have been isolated by reversion. Genetic mapping and complementation testing of a member of each group also can identify new mutations in previously identified genes as well as new genes. The identity of the gene may suggest the nature of the suppression of the mutant phenotype. For example, mapping a suppressor mutation to a known transcription factor would suggest that suppression results from altering either the timing or the amount of transcription of the original mutant allele. A developmental pathway involving the original mutant gene and several other genes then can be assembled by constructing a series of double mutants and determining which mutants are epistatic to others (49,58).

Knowledge of the class of mutation (null, hypomorph or hypermorph) and the resulting phenotype can be used to formulate and evaluate hypotheses for the function of the normal gene product and mutant alleles. The Pat (embryonic paralysis and death) phenotype of the null allele *unc-112(st562)* (123) indicates that functional UNC-112 is essential for early muscle development and viability of the embryo. Determining the class of mutation to which *unc-112(r367)* belongs will provide information concerning the relationship of the genotype and resulting phenotype. If *unc-112(r367)* is a hypomorphic allele, then the resulting phenotype could be explained as a consequence of UNC-112 functioning at a reduced level that would enable muscle development to proceed further than in null mutants. An understanding of the phenotypic consequences of a given class of alleles is particularly useful for interpreting the interactions of the intergenic and intragenic revertants of *unc-112(r367).* For example, intergenic revertants of a hypomorphic allele could be viewed as mutants that are more sensitive to the remaining activity of UNC-112; intergenic revertants of a null allele may act by recruitment of alternative proteins to function in place of UNC-112.

In summary, using standard genetic methods the null state of *unc-112* has been determined, an intergenic suppressor has been identified, and both genes have been placed in a sarcomere assembly pathway. The effects these mutations have on sarcomere structure have been elucidated using several imaging methods. Sites within sarcomeres requiring UNC112 and suppressor proteins were determined using these methods and from these results a model of how *unc-112* contributes to the formation of has been proposed.

2. Materials and Methods

A. Nematode strains and culture conditions.

Nematodes were grown and maintained at 15 °C or 20 °C on NGM plates streaked with *E. coli* strain OP50 according to (17). (NGM plates: 0.3% NaCl, 0.25% peptone, 1.7% agar, 5 mg/L cholesterol, 1 mM $CaCl_2$, 1 mM $MgSO_4$, and 25 mM KPO_4 pH 6). Genetic nomenclature follows the recommendations of Horvitz et al, (55). Some of the strains used in this study were obtained from L. DeLong and B. Meyer (University of California at Berkeley, CA, USA), K. Kemphues (Cornell University, Ithaca, NY, USA), A.M. Rose (University of British Columbia, Vancouver, BC, Canada), B.D. Williams and R.H. Waterston (Washington University School of Medicine), and the *Caenorhabditis* Genetics Center (University of Minnesota, St. Paul, MN, USA).

B. Organismal mutant phenotype

Near-starvation cultures of N2 and *unc-112(r367)* animals were grown on seeded 10 cm NGM plates at 20 °C. Animals were washed from the plates in M9 buffer and sedimented at 1500 rpm. (M9 buffer: 0.3% KH_2PO_4, 0.6% Na_2HPO_4, 0.5% NaCl, and 1 mM $MgSO_4$). The eggs were isolated using the alkaline-hypochlorite protocol (52). The eggs of N2 and *unc-112(r367)* were plated on separate seeded 10 cm NGM plates and grown at 20 °C. The plates were examined under a dissecting microscope at approximately 24, 41, 70, 89, and 104 hours post-plating. Observations were made concerning movement from hatching through adulthood, time to maturation, relative length of the mutant to N2 animals of the same age, and egg-laying ability. Brood sizes were determined by scoring the number of progeny produced from 30 individually plated *unc-112(r367)* hermaphrodites after 3 to 4 days growth at 20 °C, or after 15 days growth at 15 °C.

C. Phenotypic characterization of gonads and bodywall muscles

Organization, position, and shape of adult hermaphrodite gonads were determined using differential interference contrast (Nomarski) optics. The birefringent pattern of wild type and mutant sarcomeres in bodywall muscle cells was determined using polarized light microscopy. Nematodes were mounted on slides in M9 buffer with Sephadex G-100 (Pharmacia) beads to support the coverslip and retain as much of the fragile muscle structure as possible (47). Nomarski and polarized light images were obtained with a Zeiss Axiophot Photomicroscope (D-7082 Oberkochen) and photographed on Kodak TMAX-100 35 mm film and Kodak TMAX-400 35 mm film, respectively.

D. Phenotypic characterization of pharyngeal and bodywall muscles.

The preparation and permeabilization procedure is a variation of the protocol of D. Albertson [2]. Plain glass slides were coated at one end with 0.1 mg/ml poly D-lysine (Sigma) in water and dried at 37 °C. Alternatively, slides were dipped for 10 seconds into a 2% 3-aminopropyltriethoxysilane (APT) in acetone and dried at 37 °C. Worms were washed off one or more 70 mm NGM plates with EN buffer into conical glass centrifuge tubes and sedimented at 1500 RPM for 2 minutes in a tabletop centrifuge. (EN buffer: 0.1M NaCl, 0.01M EDTA pH 7.4). The EN buffer was aspirated and the worms washed in sucrose buffer and sedimented as before (sucrose buffer: 4% sucrose, 1 mM EDTA pH 8). All but 0.2 ml of buffer was aspirated. Between 15 and 20 μl of worm-containing buffer was transferred to a coated slide and covered with a 24 mm x 50 mm coverslip. Slides were chilled on an aluminum block placed at -80 °C from 10 minutes to overnight. Coverslips were quickly removed with a razor blade, the slides immersed in 100% methanol and chilled to -20 °C for 4 minutes. Slides were then transferred through a series of acetone baths: -20 °C 100% for 4 minutes, then at room temperature in 75%, 50%, and 25% for 1 minute each. The final bath was for 1 minute in TBS-Tween (TBS-Tween: 25 mM Tris, 2.7 mM KCl, and 137 mM NaCl pH 8, plus 0.1% Tween-20). Slides were kept in TBS-Tween at 20 °C until needed.

Primary antibodies used in this study were mouse monoclonals to *C. elegans* vinculin (MH24), β-integrin (MH25), and perlecan (MH3) described by Francis and Waterston (36,119). The antibodies were

diluted in TBS-Tween 1:200 for MH24, 1:250 for MH3 and MH25. Goat-anti-mouse secondary antibodies coupled to Texas Red (TRSC) (Jackson Scientific) were diluted 1:200 in TBS-Tween. Slides were incubated in primary or secondary antibody solutions for 2 hr minimum at room temperature in humid chambers. Following each incubation, slides were washed in TBS-Tween for 1 hr at room temperature. Coverslips were mounted using 90% glycerol and 2.5% 1,4-diazabicyclo-[2.2.2] octane (DABCO) anti-bleaching agent in TBS-Tween. Patterns were visualized with a Nikon Optiphot-2 microscope with epifluorescence and laser scanning confocal microscopy. Confocal images were captured using the BioRad MRC 600 system with a krypton/argon laser. Raw projections were imported to a Macintosh Quadra 840 AV computer and manipulated using NIH Image 1.59b and Adobe Photoshop 3.0. Images were photographed with Kodak TMAX-100 35 mm film.

E. Genetic deficiency mapping *unc-112(r367).*

Hermaphrodites heterozygous for one of the following deficiencies were crossed to *+/unc-112(r367)* males: *unc-42(e270)yDf9/nT1(m435)(IV:V), yDf11*/nT1(m435)(IV:V), *unc-42(e270)yDf8/dpy-11(e224)unc-76(e911),* and *itDf2/nT1(n754)(IV:V).* F1 hermaphrodites were scored for the presence of paralyzed progeny which indicated that the deficiency failed to complement *unc-112(r367).* The nonparalyzed progeny resulting from *yDf8*- and *itDf2*-bearing animals were progeny tested to determine their genotype and frequency in the F1 brood since no paralyzed animals were observed. The presence of either *yDf8* or *itDf2* was confirmed by the segregation of 25-30% dead embryos from one-third of the viable F1 hermaphrodites. The crosses with *yDf9*- or *yDf11*-bearing animals also resulted in nonparalyzed F1 animals. At least 20 non-Unc F1 hermaphrodites, which lacked the dominant Unc-bearing *nT1(m435)* chromosome used to balance *yDf9* and *yDf11*, were picked for progeny testing to determine genotype and frequency. In addition, the presence of each deficiency was confirmed by crossing individual F1 hermaphrodites to *+/unc-76(e911)* or *+/unc-61(e228)* males. The resulting progeny were examined for Unc behavior, indicating failure to complement one of these mutations.

F. Isolation of *unc-112(r367)* intragenic and intergenic revertants

Homozygous *unc-112(r367)* hermaphrodites from five 100 mm seeded NGM plates were mutagenized in 50 mM EMS (ethylmethane sulfonic acid) in M9 buffer for 4 hours at 23 °C. Nematodes were then plated on fifty 100 mm seeded NGM plates and grown at 20 °C for one generation to isolate dominant revertants. Between 50,000 and 100,000 progeny were examined. Two well-moving hermaphrodites were picked from each plate for a total of 100. Fifty-nine homozygous well-moving strains were established. One spontaneous dominant revertant, DM1262, was isolated while maintaining *unc-112(r367)* animals at 20 °C. A second revertant, DM1261, was isolated while attempting to place *unc-112(r367)* into a *mut-6 unc-22* background for Tc1 tagging. These two non-EMS induced revertants were added to the collection. The revertant strains, except DM1262, were classified as either intragenic or intergenic classes after outcrossing homozygous hermaphrodites to N2 (wild type) males and examining the F2 progeny for the frequency of paralyzed Unc animals. Those failing to segregate paralyzed *unc-112(r367)* animals were classified as intragenic suppressors. Those segregating paralyzed animals were classified as intergenic suppressors. For one intergenic suppressor strain, DM1243, several well-moving siblings from the F2 generation were picked and allowed to segregate progeny by self-fertilization. Their progeny were followed until a homozygous *dim-1(ra102)* suppressor strain was established.

G. Two and three factor mapping of two intergenic suppressor mutations

Rubber Band phenotype of the suppressor strain DM1243 is linked to chromosome IV. DM1243 hermaphrodites were outcrossed to N2 males. Several wild-type F1 hermaphrodites were allowed to self-fertilize. If the F2 generation exhibited only Wild type and Rubber Band phenotypes (none were Paralyzed), then several Rubber Band hermaphrodites were allowed to self-fertilize. Those segregating only Rubber Band progeny (no Wild type or Paralyzed progeny) were shown to be homozygous and the mutant allele named *ra201*. For two-factor mapping analysis, *ra201* hermaphrodites were crossed to *dpy-5(e61) I /+*, *dpy-10(e128) III/+*, *dpy-18(e364) IIII/+*, or *dpy-13(e184) IV/+* males. Ten Wild type hermaphrodite progeny from each outcross were allowed to self-fertilize. If Dpy F2 progeny were observed, the numbers of progeny exhibiting Wild type, Dpy, Rubber Band, and Dpy-Rubber Band were

determined. If Dpy F2 progeny were not observed then the brood was discarded because the parent (F1) did not acquire the Dpy mutation from the male parent (P0).

For cis two-factor mapping, *ra201dpy-4(e1166)* hermaphrodites were outcrossed to N2 males. Fifteen Wild type hermaphrodites were allowed to self-fertilize. The numbers of progeny exhibiting Wild type, Dpy-Rubber Band, Dpy and Rubber Band were determined. The recombination frequency was determined from the sum of the Dpy and Rubber Band phenotypic classes divided by the total number of progeny examined as given by Brenner (17).

For three-factor mapping analysis, *dpy-4(e1166)unc-22(e66)* hermaphrodites were crossed to *ra201*/+ males. Ten Wild type progeny were allowed to self-fertilize. The broods of four hermaphrodites segregated Rubber Band progeny. From these broods 9 Twitcher and 20 Dpy hermaphrodites were allowed to self-fertilize. The phenotypes of the progeny of the Dpy hermaphrodites were determined by examination. From this the number of broods segregating only Dpy or both Dpy and Rubber Band were determined.

Dim phenotype of the suppressor strain DM1243 is linked to the X chromosome. Linkage of the mutation causing the Dim phenotype was tested by outcrossing *ra201dpy-13(e184)* hermaphrodites to N2 males. Wild type F1 hermaphrodites were allowed to self-fertilize. Seventeen Rubber Band and 37 Wild type F2 progeny were examined for Dim muscles using polarized light microscopy. A homozygous Dim strain, DM2, was isolated from the progeny of 1 of 2 hermaphrodites that only segregated Dim progeny and did not segregate Dpy or Rubber Band progeny. The DM2 strain containing the *ra102* allele was used in all subsequent mapping experiments.

To determine linkage of the *ra102* mutation, *ra102* hermaphrodites were crossed to *dpy-5(e61) I*/+, *dpy-10(e128) II*/+ or *dpy-18(e364) III*/+ males. Several Wild type F1 hermaphrodites from each cross were allowed to self-fertilize. Seventeen Dpy F2 progeny were examined using polarized light microscopy and the number of animals with Wild type or Dim muscles determined. Linkage of the Dim phenotype with the X chromosome was tested by outcrossing DM2 hermaphrodites with N2 males. More than 20 outcrossed male and hermaphrodite progeny were examined using polarized light microscopy for Dim muscles.

To position the *ra102* mutation on the X chromosome, *dpy-7(e88)unc-3(e151)*, *unc-1(e719)dpy-7(e88)* or *lon-2(e678)dpy-6(e14)* hermaphrodites were crossed to DM2 males. Several Wild type F1 hermaphrodites were allowed to self-fertilize. Several Unc, Dpy or Lon progeny were examined using polarized light microscopy for Dim muscles.

H. Complementation testing of two intergenic suppressor mutations

Complementation testing of chromosome IV Rubber Band *ra201* mutation with *unc-30* and *unc-26* alleles. Hermaphrodites homozygous for *unc-30(e318)* or *unc-26(e205)* were crossed to *ra201*/+ males. The phenotype of all male progeny was determined. For confirmation, 10 Wild type F1 hermaphrodites were allowed to self-fertilize. The numbers of Wild type and Unc progeny were determined. In addition, *ra201* hermaphrodites were crossed to *unc-30(e318)*/+ or *unc-26(e205)*/+ males and the phenotype of all male F1 progeny was determined.

Complementation testing of the *ra102* mutation with *unc-97* and *unc-98* alleles. Hermaphrodites homozygous for *unc-97(su110)* or *unc-98(su130)* were crossed to *ra102* males. The phenotype of all hermaphrodite progeny was determined by examining animals for movement and organization of bodywall muscles under polarized light.

I. Genetic deficiency mapping of *ra102*

Male *ra102* animals were crossed to *uDf1/szT1[lon-2(e678)] (X;I)*, *yDf2/szT1[lon-2(e678)] (X;I)*, or *yDf3/szT1[lon-2(e678)] (X;I)*, or *unc-32(e189) III;yDf5/szT1[lon-2(e678)] (X;I)* hermaphrodites. All hermaphrodite progeny were examined for Dim muscle under polarized light.

J. Suppression analysis of *unc-112(r367)* by *unc-26*, *unc-97*, *unc-98*, *uDf1*, and *dim-1*

Hermaphrodites homozygous for *unc-112(r367)* were crossed to *unc-26(e205)*/+ males. Ten Wild type F1 hermaphrodite progeny were allowed to self-fertilize and the frequency of each mutant phenotype was determined for the F2 progeny.

unc-97(su110) or *unc-98(su130)* hermaphrodites were crossed to +/*unc-112(r367)* males. Several non-Unc F1 hermaphrodites were picked to self-fertilize and the frequency of each mutant phenotype was determined for the F2 progeny.

uDf1/szT1[lon-2(e678)] (X:I) hermaphrodites were crossed to +/*unc-112(r367)* males. Twenty F1 hermaphrodites were picked and the F2 broods were examined for absence of Lon males and the frequency of Paralyzed and Bag-of-worms animals. Several Bag-of-worms animals were further examined for the presence of Dim muscle under polarized light.

Hermaphrodites homozygous for *unc-112(r367)* were crossed to *dim-1(ra102)* males. Several Wild type F1 hermaphrodites were allowed to self-fertilize. The F2 progeny were examined for the ratio of Paralyzed to Wild type and Slow animals. In addition, several animals from each brood were also examined for Dim muscle under polarized light.

K. Identification of additional alleles of *dim-1* by complementation test

Each of the 17 homozygous intergenic revertants were tested for complementation by crossing heterozygous revertant males to homozygous *dim-1(ra102)* hermaphrodites. At least 10 F1 hermaphrodites and all males were examined under polarized light for the presence of Dim muscle.

3. Results

A. Characterization of *unc-112(r367)*.

Phenotype of unc-112(r367).

Prior to this study the phenotype of the *unc-112(r367)* allele had not been described in detail. The phenotype of a mutant strain may provide clues to the function of *unc-112*, and a careful examination of the behavioral and morphological phenotypes was undertaken at the organismal, organ system, and cellular levels. When viewed under the dissecting microscope, *unc-112(r367)* animals raised at 20 °C appear to move well as young larvae but lose coordination as they progress through larval stages L3 and L4, becoming small, thin and paralyzed as adults. In addition, *unc-112(r367)* animals grow more slowly and produce smaller broods of progeny than N2 animals.

It was noted previously that viability of the *unc-112(r367)* strain was affected by the temperature at which it was maintained, surviving better at 15 °C than at 20 °C (P. Anderson, per. comm.). This may be either due to temperature-delayed onset of paralysis and hence improved feeding and egg-laying, or due to more larvae surviving to adulthood. In order to differentiate between these possibilities, synchronous cultures of N2 and *unc-112(r367)* raised at 15 °C and 20 °C were observed at each stage of development (see Table 2). The L1 and L2 stage larvae of *unc-112(r367)* exhibited wild type body morphology and movement but were slower than N2 larvae of the same age and grown at the same temperature. Paralysis occurred about L3 stage for animals raised at 20 °C, and about L4 stage for animals raised at 15 °C. The average length of mutant animals was also temperature dependent. *unc-112(r367)* animals raised at 20 °C were about 30% smaller at all stages than their N2 counterparts but were the same size when raised at 15 °C. Temperature also affected the ability of *unc-112(r367)* animals to lay eggs and also median brood size. All eggs were retained inside hermaphrodites raised at 20 °C, and the median brood size was 16 progeny (n=30). In contrast, many eggs were laid by hermaphrodites raised at 15 °C, and the median brood size increased to 23 progeny (n=30). The length of time to maturity also was temperature-dependent and about 1.5-3 fold longer for *unc-112(r367)* than N2 animals. Embryos isolated from N2 hermaphrodites and raised at 20 °C reached adulthood in about 70 hours, whereas *unc-112(r367)*

TABLE 2

TEMPERATURE SENSITIVE AND INSENSITIVE PHENOTYPES OF *unc-112(r367)*

PHENOTYPES	15 °C	20 °C	TS EFFECT
MOVEMENT	Slow	Slow	no
COORDINATION: Good Poor Paralyzed	 L1, L2, L3 good L4 poor adults paralyzed	 L1 and L2 good L3 poor L4 and adults paralyzed	yes
RELATIVE LENGTH vs WILDTYPE	100%	70%	yes
MEDIAN BROOD SIZE	23 viable progeny (n=30)	16 viable progeny (n=30)	yes
GONAD MORPHOLOGY	One or both arms malformed, some ruptures	One or both arms malformed, many ruptures	no
GROWTH RATE vs WILDTYPE	15 days mutant/ 5 days + (1.5 x)	~100 hr mutant/ 70 hr + (1.5 x)	no

embryos required 100 hours. N2 embryos raised at 15 °C required 5 days to mature, while *unc-112(r367)* animals required 15 days. In summary, the muscles and gonads of *unc-112(r367)* animals raised at 15 °C function longer and produce more progeny per generation, but eventually deteriorate.

Muscle phenotype of unc-112(r367).

Polarized light and immunofluorescence microscopy were used to examine and compare the sarcomere structure in bodywall muscles of *unc-112(r367)* and N2 adult animals. Under polarized light, N2 bodywall muscles (see figure 5a) contain alternating light and dark gray stripes running the length of the cells. Closely paired light bands correspond to the region of overlap of thin and thick filaments (A-bands). The dark bands correspond to the thin filaments (I-bands) and the row of light spots within them correspond to the dense bodies. This pattern is not seen in *unc-112(r367)* bodywall muscle cells (see figure 5b). In these animals bright fibrous clumps fill the bodywall muscle cells with no apparent organization.

Immunofluorescence microscopy using antibodies to vinculin, β-integrin and perlecan was used to determine in what part of the structure a break had occurred that led to the collapse of the sarcomeres seen under polarized light in *unc-112(r367)*. Vinculin antibodies stain the base of the dense bodies and the adhesion plaques at the ends of the rows of thin filaments. In N2 bodywall muscles this forms a pattern of punctate stripes the length of the cells (rows of dense bodies) with bright spots at the ends of the cells (adhesion plaques) (see figure 6a). In *unc-112(r367)* muscle only short rows and clumps of spots are seen. Furthermore, it is not possible to distinguish dense bodies from adhesion plaques or the ends of individual cells (see figure 6d). β-integrin antibodies stain the transmembrane portion of dense bodies, M-lines and adhesion plaques. The N2 pattern of β-integrin staining forms alternating rows of bright, punctate (dense bodies) and less bright, unbroken (M-lines) stripes and very bright adhesion plaques (see figure 6b). β-integrin staining in *unc-112(r367)* bodywall muscle shows only short rows and clumps of spots of indeterminate origin (see figure 6e). Perlecan antibodies stain the basement membrane overlying the muscle quadrants in a diffuse blanket with brighter staining in spots overlying the dense bodies and unbroken stripes overlying the M-lines and ends of the cells. The N2 pattern of

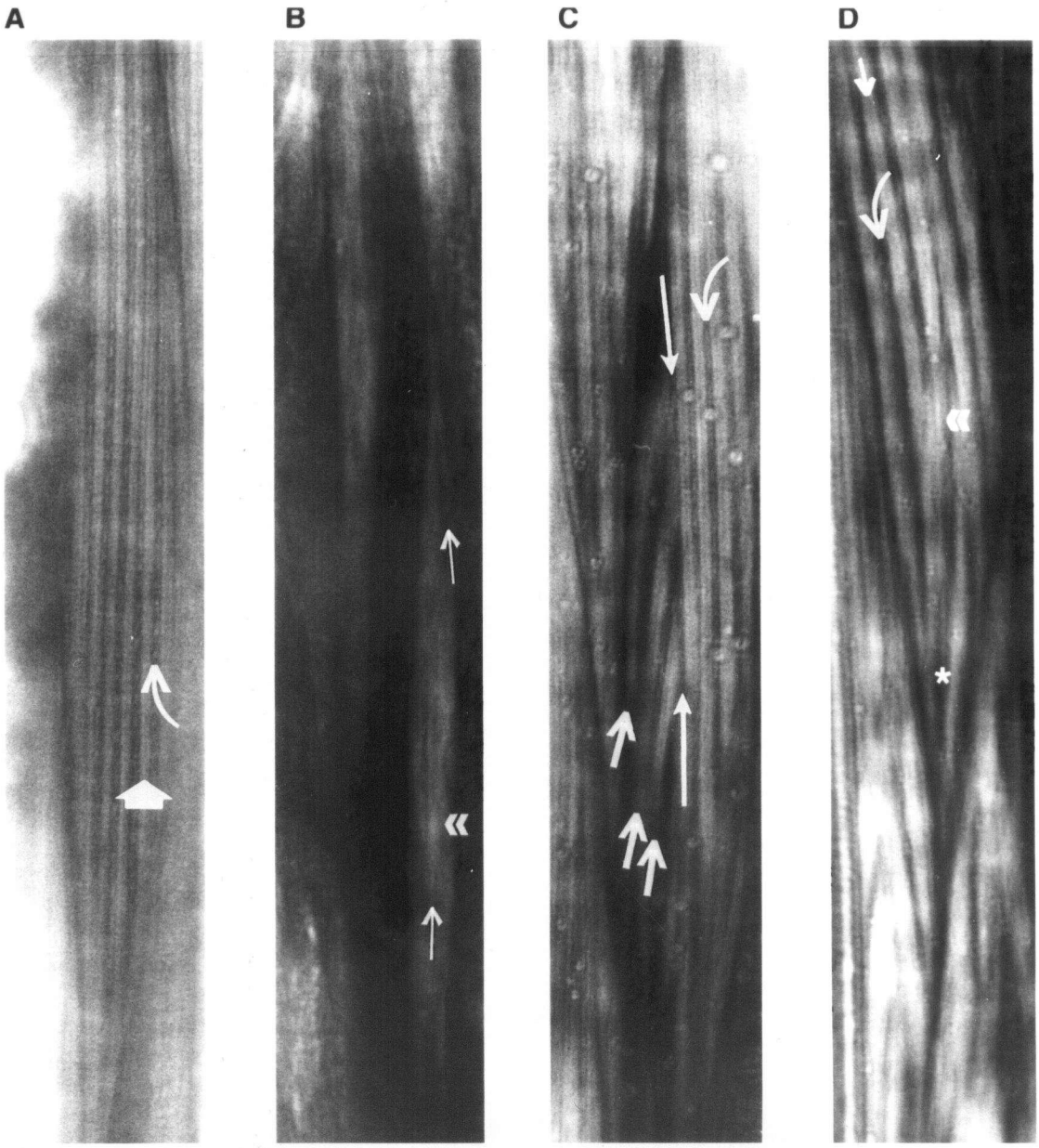


Figure 5: Comparison of birefringent muscle structures in N2, unc-112(r367), dim-1(ra102) and unc-112(r367);dim-1(ra102).

Panel **A**) shows a single wildtype body wall muscle cell. The bright A-bands are the regions where thick and thin filaments overlap (stubby arrow). These alternate with darker bands containing the non-overlapping portions of the thin filaments connected to the bright dense bodies (curved arrow). Panel **B**) shows *unc-112(r367)* body wall muscle to the same scale as A). No organization of the filaments is apparent only bright clumps (double arrowhead) and short fibers (arrows). **C**) most muscle cells in dim-*1(ra102)* animals appear wildtype, but a few cells per animal have the chevron pattern shown here. Three short A-bands on the left side of the panel (short arrow) appear to terminate near an A-band of the opposite orientation (long arrows). Dense bodies are seen in the dark normal length bands and in the abnormal dark bands (curved arrow). **D**) the *unc-112(r367);dim-1(ra102)* muscle cells show the partially suppressed phenotype. The A-bands are not clearly defined and dense bodies are not easily seen (curved arrow). Portions of some A-bands have clumped (double arrowheads) at one end while the other end appears somewhat normal with M-lines still visible (stubby arrow). In places it appears that the ends of two A-bands have fused together (*).

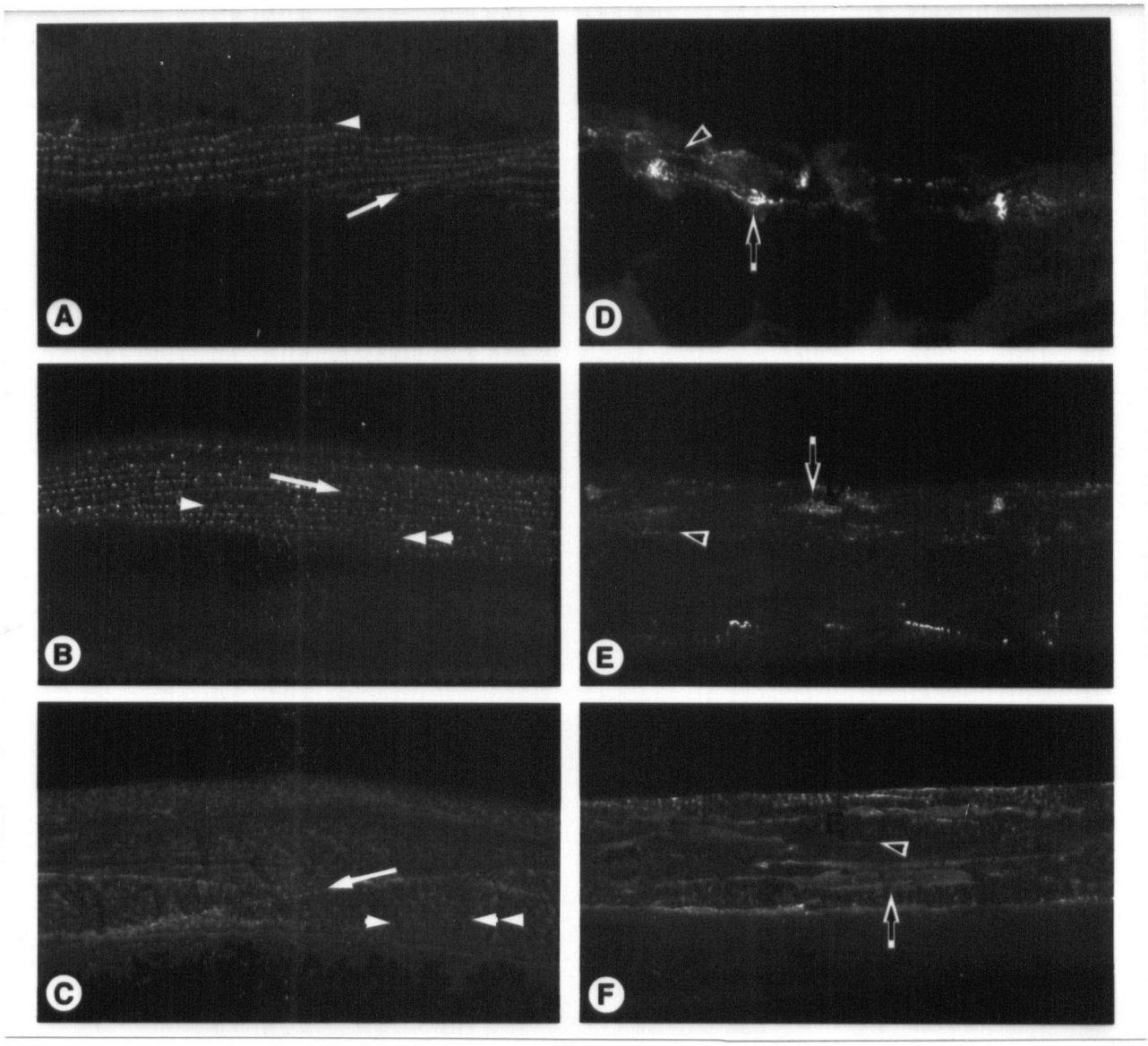


Figure 6: Comparison of vinculin, β-integrin and perlecan distribution in the body wall muscles of N2 and unc-112(r367) animals.

The immunofluorescent antibody staining patterns for N2 are shown in A-C and for *unc-112(r367)* are shown in D-F. In **A**) the vinculin antibody MH24 stains rows of dense bodies (solid arrowhead) and the attachment plaques along cell margins (solid arrow). In **D**) clumps (open arrow) and short rows (open arrowhead) of punctate MH24 staining are seen. It is not possible to distinguish the attachment plaque staining along the cell margins. In **B**) β-integrin antibody MH25 stains rows of dense bodies (solid arrowhead), M-lines (double arrowhead) and attachment plaques (arrow). In **E**) MH25 stains clumps (open arrow) and short rows (open arrowhead) of punctate material and it is difficult to distinguish the cell margins. In **C**) the perlecan antibody MH3 diffusely stains the basement membrane and is particularly concentrated over muscle cell margins (arrow), dense bodies (solid arrowhead) and M-lines (double arrowhead). In **F**) MH3 stains diffuse patches (open arrow) and in short punctate rows (open arrowhead).

perlecan staining appears as a diffuse stripe spanning the length and width of the muscle quadrants with each cell brightly outlined and patterned with alternating punctate rows and solid stripes overlying the dense bodies and M-lines, respectively (see figure 6c). In contrast, the *unc-112(r367)* pattern of perlecan staining shows patches of diffuse staining, scattered clumps, and short stripes (6f). In addition, the outlines of the cells cannot be determined. All of these results appear to be similar in form to the polarized light image of disorganized fibrous clumps. This suggests that it is the stability of the placement of the sarcomeres in the membrane and not the stability of the association of all of these components that is defective in *unc-112(r367)* bodywall muscles.

The distribution of vinculin was also examined in both N2 and *unc-112(r367)* pharyngeal muscles using immunofluorescence and MH24 antibodies. The pharynx, as described by Albertson and Thomson (2), consists of five parts: the buccal cavity, procorpus, metacorpus, isthmus, and terminal bulb. All but the buccal cavity contain muscle cells (see figure 7a). Vinculin antibodies did not stain the lumenal surface of the muscle cells, but did stain the apical muscle membranes of all of the muscle cells except for the three cells of the terminal bulb closest to the isthmus in both N2 and *unc-112(r367)* (see figure 7b,c). The non-staining layer of muscle cells interdigitate with the three cells in the staining layer closest to the intestinal valve. This appeared as a hand-shaped pattern of vinculin staining. The density of the vinculin staining, however, appeared to be half as much in *unc-112(r367)* than in N2. At the base of the terminal bulb that opens into the intestine was a saucer-shaped muscle cell which also stained with vinculin antibodies, but left a gap in staining at the interface between this cell and the hand-shaped cells. This gap was wider in *unc-112(r367)* than in N2. In addition, the region of staining just below the hand shape was narrower in *unc-112(r367)* than in N2. These differences in pharyngeal vinculin antibody staining are less dramatic than the differences seen in bodywall muscle staining; however, they may be sufficient to reduce the ability of *unc-112(r367)* animals to obtain sufficient nourishment.

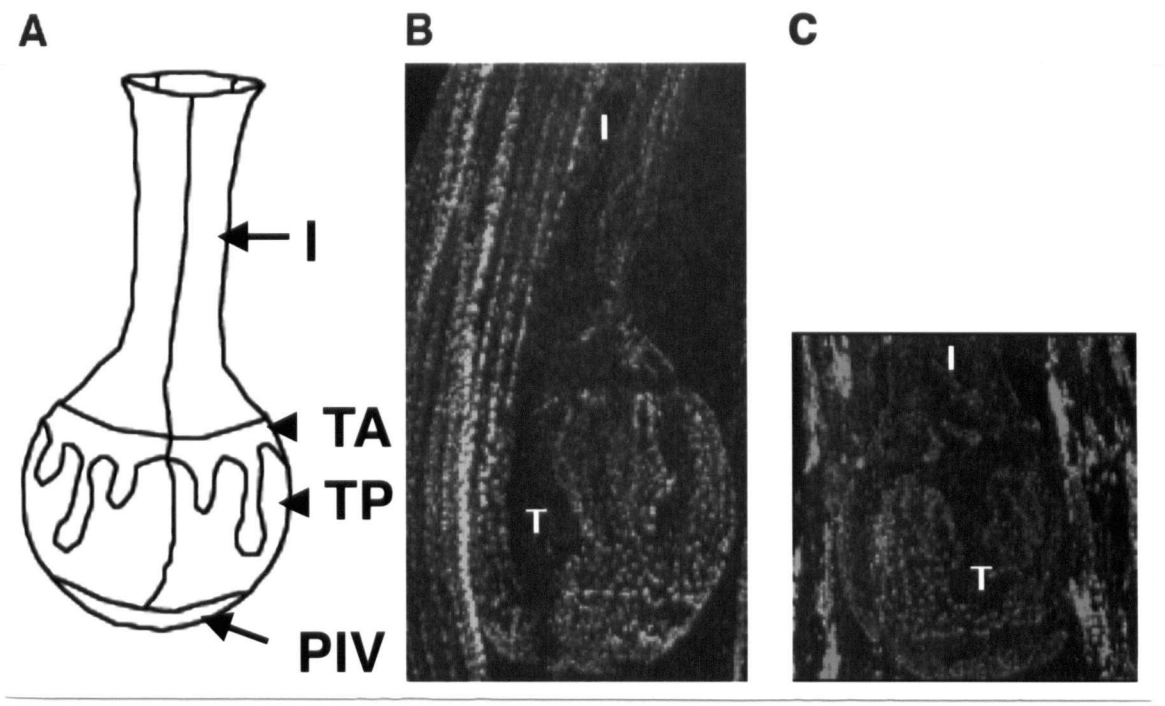


Figure 7: Comparison of vinculin distribution in the pharyngeal muscles of N2 and unc-112(r367) animals.

The bucal opening of the pharynx is at the top and the opening into the intestine is at the bottom of the images. Panel **A**) shows a cartoon of the lower portion of the pharynx from the isthmus (I) to the posterior intestinal valve (PIV). Two layers of muscle cells interdigitate to form the terminal bulb: termed TA (for terminal bulb anterior) and TP (for terminal bulb posterior). Panel **B**) shows the pharyngeal bulb of a wildtype adult, flanked by bodywall muscle cells, stained with MH24. The "hand-shaped" patches of adhesion plaques on the terminal bulb muscle cells (T) are located near the basal surface of the cells. There are no adhesion plaques near the lumenal surface of the cells. The regions of no staining form an outline of the muscle cell shapes (3). The MH24-stained adhesion plaques are in rows down the length of the isthmus (I) analogous to body wall muscle organization. Panel **C**) shows the pharyngeal bulb of an *unc-112(r367)* adult flanked by bodywall muscle cells and stained with MH24. The terminal bulb muscle cells (T) have "hand-shaped" patches of adhesion plaques but there does not appear to be as many as in panel A. Likewise the isthmus muscle cells (I) do not have as many MH24-staining adhesion plaques. The staining however, is not organized in rows down the length of the cells.

Deletion mapping and analysis of unc-112(r367).

The *unc-112(r367)* allele was mapped to the interval between *par-1* and *unc-76* on chromosome V by 2- and 3-factor mapping (P. Anderson, per. comm.). Additional mapping placed *unc-112* to the right of *egl-1* (M. Hengartner, per. comm.). However, a 0.5 map unit region was too large to attempt cloning the *unc-112* gene. To more precisely map *unc-112*, genetic complementation tests were conducted between *unc-112(r367)* and 4 deficiencies extending from the right of *par-1* through *unc-76*: *itDf2*, *yDf8*, *yDf9*, and *yDf11* (see figure 8). The results show that *unc-112(r367)* complements *yDf9* and *yDf11*, but fails to complement *itDf2* and *yDf8*. Consequently, *unc-112(r367)* lies to the left of *yDf9* and *yDf11* and is located in the region right of *par-1* and close to the right breakpoint of *yDf8* (see figure 8).

In addition to refining the map position of *unc-112*, the deficiency complementation results with *itDf2* and *yDf8* also indicate that *r367* is not a null allele. The phenotype of *unc-112(r367)/yDf8* or *itDf2* was embryonic Lethality and not Paralysis. This phenotype is similar to the Pat phenotype of *unc-112(st562)* which is an embryonic lethal. The change in severity of phenotype when an allele is in trans to a deficiency, in this case from larval paralysis to embryonic lethality, is a defining characteristic of hypomorphic alleles (49). Furthermore, the similarity of the Pat phenotype of *unc-112(st562)* and the embryonic lethal phenotype of *unc-112(r367)/yDf8* or *itDf2* suggests that *unc-112(st562)* is a null allele.

Isolation of intragenic and intergenic revertants.

In order to identify intragenic modifications and interacting genes, motile progeny were isolated from paralyzed *unc-112(r367)* hermaphrodites mutagenized with 50 mM EMS. Fifty-nine dominant F1 revertants were isolated from between 50,000 and 100,000 homozygous *unc-112(r367)* animals. One spontaneous revertant (DM1261) was identified during an attempt to introduce the Tc1 transposon mutator, *mut-6*, into *unc-112(r367)*. A second spontaneous revertant (DM1262) was isolated while maintaining the strain on plates. DM1262 was subsequently determined to be homozygous lethal. Both spontaneous alleles were added to the collection bringing the total to 61 revertants. With the exception of DM1262, homozygous strains of each revertant were established. Each homozygous strain was then outcrossed to N2 males and the F2 progeny examined for the reappearance of the paralyzed phenotype

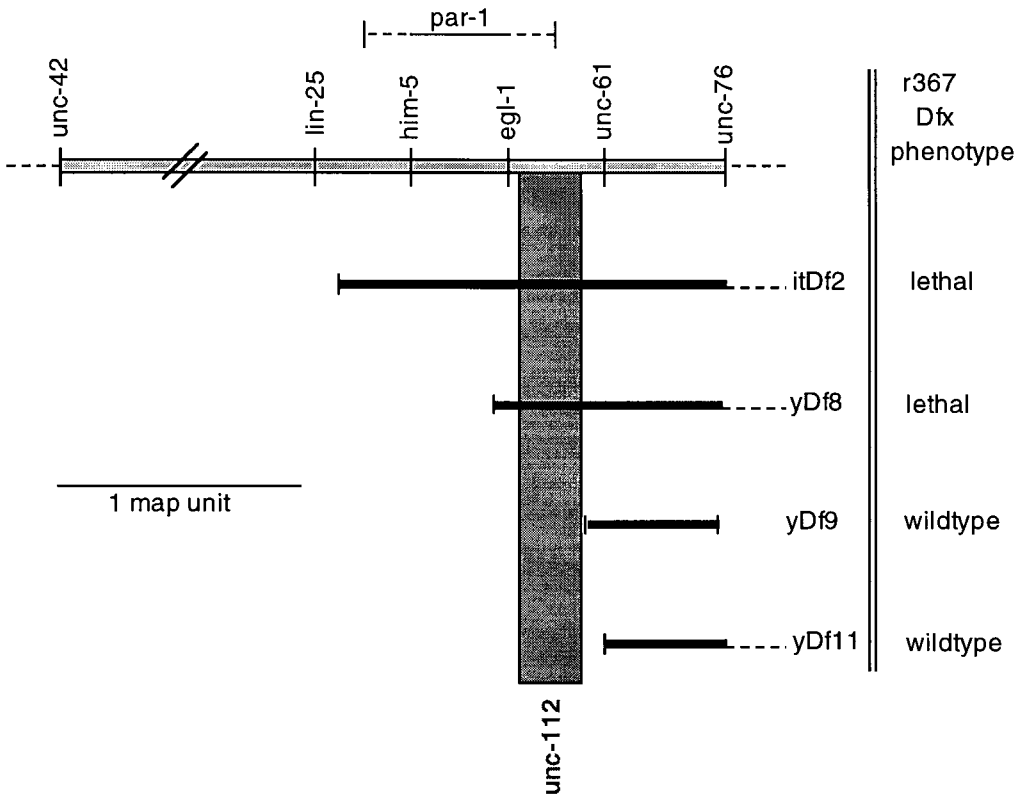


Figure 8: Refinement of the genetic map position of unc-112(r367) by deficiency complementation.

Approximately 2.5 map units of chromosome V are shown to scale, and the relative positions of genes in the region are shown either on the line or above the line with the map range indicated by an H-shaped bar. The center of the chromosome is to the left, and the telomere end is to the right. The deficiencies *yDf8, yDf9* and *yDf11* complement (do not knock out) *par-1*, whereas *itDf2* fails to complement (does knock out) *par-1* (L. DeLong, pers. comm.). The genetically defined breakpoints of the four deficiencies within this region are shown below the genetic map. The phenotype of *unc-112(r367)* over each deficiency is given to the right of each deficiency. The lethal phenotype of *unc-112(r367)/yDf8* indicates that this deficiency knocks out function by breaking within or deleting this essential gene. The result places *unc-112* either very close and to the left of *egl-1* or between *egl-1* and *unc-61*. Three factor mapping of *unc-112(r367)* relative to *egl-1* placed *unc-112(r367)* to the right of *egl-1* (M. Hengartner, pers. comm.). The resulting map position of *unc-112* is shown in the hatched box extending down from the genetic map.

of *unc-112(r367)* animals. The frequency with which the original phenotype arose indicated whether the suppressor mutation and *unc-112(r367)* were linked. Of the 58 revertant strains tested, 9 outcrossed revertant strains did not segregate any paralyzed F2 animals, indicating the suppressor mutations were either second-site alterations within *unc-112* (called intragenic suppressors) or very tightly linked intergenic suppressors. The 49 remaining revertant strains tested segregated paralyzed F2 animals, indicating no or very little linkage between the suppressors and *unc-112(r367)* (called intergenic suppressors); (see Table 3). Four of the intragenic revertants were used to isolate cDNAs that may encode *unc-112* (Chapter 3).

B. Characterization of *dim-1(ra102)*

Genetic analysis of intergenic suppressors.

The reversion of the *unc-112(r367)* mutation yielded 49 intergenic strains that delayed the onset of paralysis or reduced the severity of the uncoordinated phenotype for the life span of the animals. Characterization of these suppressor mutations may identify factors that interact with *unc-112* in embryonic muscle differentiation, or in larval sarcomere growth or in both processes. However, an extensive characterization of all 49 mutations was beyond the scope of this thesis; therefore, one intergenic revertant strain was selected for genetic mapping. The revertant strain DM1243 was selected owing to its visible phenotype of spontaneous bursts of simultaneous contraction in all four quadrants of bodywall muscle and a muscle phenotype of fragility and in some muscle cells aberrantly organized A-bands dubbed the "Chevron" phenotype. The visible phenotype resembles the "Rubber Band" phenotypes of *unc-93*, *sup-9* and *sup-10* (44). While the muscle phenotype has been seen only rarely in a bodywall muscle in the tail of in wild type animals, in DM1243 several bodywall muscles anywhere along the length of the quadrants exhibit the Chevron pattern.

The Rubber Band and Chevron phenotypes exhibited by the DM1243 suppressor strain were individually mapped to determine whether the phenotypes were a consequence of one or two mutational events. Two factor mapping was used to link the Rubber Band phenotype, designated *ra201*, with chromosomes I, II, III or IV. It was unnecessary to test chromosomes V and X because the Rubber Band

TABLE 3

STRAIN NAMES AND PHENOTYPES OF INTRA- AND INTERGENIC SUPPRESSORS

STRAIN NAME	GENOTYPE	PHENOTYPE	INTRA/INTERGENIC
DM2208	*unc-112(r367);dim-1(ra102)*	wildtype	intergenic
DM1202	*unc-112(r367);(ra210)*	wildtype	intergenic
DM1204	*unc-112(r367)* + ?	wildtype	not determined
DM1205	*unc-112(r367)* + ?	wildtype	not determined
DM1206	*unc-112(r367);(ra212)*	wildtype	intergenic
DM1208	*unc-112(r367ra202)*	wildtype	intragenic
DM1209	*unc-112(r367);(ra213)*	coiler	intergenic
DM1210	*unc-112(r367);(ra214)*	coiler	intergenic
DM1211	*unc-112(r367);(ra211)*	slow	intergenic
DM1212	*unc-112(r367ra206)*	slow	intragenic
DM1213	*unc-112(r367);(ra215)*	slow	intergenic
DM1215	*unc-112(r367ra216)*	slow	intragenic
DM1217	*unc-112(r367)* + ?	wildtype	not determined
DM1218	*unc-112(r367)* + ?	wildtype	not determined
DM1219	*unc-112(r367);(ra217)*	wildtype	intergenic
DM1220	*unc-112(r367);(ra218)*	wildtype	intergenic
DM1221	*unc-112(r367);(ra219)*	wildtype	intergenic
DM1223	*unc-112(r367)* + ?	wildtype	not determined
DM1224	*unc-112(r367ra207)*	very slow to paralyzed bag-of-worms	intragenic
DM1225	*unc-112(r367);(ra220)*	wildtype	intergenic
DM1227	*unc-112(r367);(ra208)*	slow	intergenic
DM1229	*unc-112(r367)* + ?	wildtype	not determined
DM1230	*unc-112(r367);(ra221)*	wildtype	intergenic
DM1231	*unc-112(r367);(ra222)*	slow	intergenic
DM1233	*unc-112(r367ra223)*	wildtype	intragenic
DM1234	*unc-112(r367);(ra224)*	wildtype	intergenic
DM1235	*unc-112(r367ra205)*	slow to paralyzed bag-of-worms	intragenic
DM1236	*unc-112(r367);(ra225)*	wildtype	intergenic
DM1237	*unc-112(r367);(ra226)*	slow	intergenic
DM1238	*unc-112(r367);(ra227)*	slow	intergenic
DM1240	*unc-112(r367);(ra228)*	wildtype	intergenic
DM1242	*unc-112(r367);(ra229)*	wildtype	intergenic
DM1243	*unc-112(r367);dim-1(ra201)*	wildtype	intergenic
DM1244	*unc-112(r367);dim-1(ra204)*	L3 transient paralysis	intergenic
DM1246	*unc-112(r367);(ra230)*	wildtype	intergenic
DM1247	*unc-112(r367);dim-1(ra203)*	wildtype	intergenic
DM1248	*unc-112(r367);(ra231)*	wildtype	intergenic
DM1250	*unc-112(r367);(ra221)*	wildtype	intergenic
DM1252	*unc-112(r367ra233)*	wildtype	intragenic
DM1254	*unc-112(r367);(ra234)*	wildtype	intergenic
DM1256	*unc-112(r367);(ra235)*	deformed tail, wildtype movement	intergenic
DM1257	*unc-112(r367ra236)*	wildtype	intragenic
DM1258	*unc-112(r367)* + ?	wildtype	not determined
DM1259	*unc-112(r367);(ra237)*	slow	intergenic
DM1261	*unc-112(r367);dim-*	wildtype	intergenic

	1(ra209::Tc1?)		
DM1262	*unc-112(r367)* + ?	wildtype (possibly homozygous lethal)	not determined

phenotype segregates independently of *unc-112(r367) V* (see above), and outcrossed males move normally. The phenotypes of the progeny from 5 *dpy-13(184) IV/+;unc-112(r367)/+; ra201/+* hermaphrodites (for a total of 1184 progeny) were determined. The ratio of the phenotypic groups was 48.6: 21.5: 20: 1 (Wild type: Dpy: Rubber Band: Dpy + Rubber Band respectively). This deviates significantly from the predicted ratio of 9: 3: 3: 1 if the two mutations were unlinked and demonstrates that the Rubber Band phenotype is linked to *dpy-13(184) IV*. Recombination frequency was measured between the mutation producing the Rubber Band phenotype and *dpy-4(e1166)* by outcrossing the double mutant strain. The progeny of 15 hermaphrodites were examined, and the two mutations were recombinationally separable at a frequency of 4.5% (151/ 3369 progeny were either Dpy or Rubber Band). The result positions the *ra201* mutation 4.5 map units either left or right of *dpy-4* on chromosome IV. Three factor mapping with *dpy-4(e1166)unc-22(e66)* was used to position the Rubber Band mutation near *unc-30* and *unc-26,* 4.5 map units to the left of *dpy-4(e1166).* Twenty Dpy progeny from triple heterozygous hermaphrodites were selected and allowed to self-fertilize. Eleven of the Dpy animals yielded progeny exhibiting the Rubber Band phenotype, and the progeny of nine Dpy animals did not exhibit the Rubber Band phenotype. These results place *ra201* between *unc-22* and *dpy-4* and approximately 4.5 map units to the left of *dpy-4*. The Rubber Band mutation was then tested for allelism with *unc-30* and *unc-26* by complementation analysis. The ratio of Wild type to Unc male progeny from crossing *ra201/+* males with *unc-30(e318)* or *unc-26(e205)* hermaphrodites is predicted to be 1: 0 if the two mutations complement and 1: 1 if the two mutations fail to complement. No Unc males resulted from the complementation test with *unc-30(e318),* therefore the *ra201* mutation is not allelic with this gene. However, half of the males resulting from the complementation test with *unc-26(e205)* were Unc, therefore the *ra201* mutation is an allele of *unc-26.*

Two factor mapping was performed by D. Moerman to link the Chevron phenotype, designated *ra102*, to either chromosome I, II, III or IV. The Chevron phenotype segregated independently of markers on each of these chromosomes, demonstrating that two genes were mutagenized in DM1243: *unc-26(ra201) IV* and another gene *ra102.* Confirmation that *ra102* was linked to the X chromosome was obtained from the observation that all outcrossed males from the suppressor strain exhibited the Chevron phenotype. Furthermore, outcrossed hermaphrodites had wildtype muscles and therefore, were

heterozygous for the mutation. Since males carry a single X chromosome and hermaphrodites have two X chromosomes, these results demonstrate that the *ra102* mutation is X-linked and recessive.

Three factor mapping performed by D. Moerman placed *ra102* to the right of *dpy-7,* and to the left of *dpy-6* on the X chromosome (see figure 9). Deficiency complementation tests with *raDf2, raDf9,* and *uDf1* was used to further refine the position of *ra102* and placed it between *mec-7* and *xol-1* (see figure 9). The *ra102* allele was tested for complementation with two muscle-affecting genes, *unc-97* and *unc-98,* which also map in this region (128). The results showed that *ra102* is not an allele of either gene. Taken together, these results demonstrate that the mutation causing the Chevron phenotype is an allele of a new gene which has been named *dim-1* (for disorganized muscle).

Positioning *dim-1(ra102)* between the breakpoints of the deficiencies also enabled the genotypic nature of the mutation to be determined. The *dim-1(ra102)* allele appears to be null since the phenotype of *dim-1(ra102)/uDf1* animals was the same as that of homozygous *dim-1(ra102)* animals (D. Moerman, per. comm.). Moreover, the viability of both *dim-1(ra102)/uDf1* and *dim-1(ra102)* homozygous animals also suggests that *dim-1* is not essential for survival in *C. elegans*.

Suppression of unc-112(r367) is a property of dim-1(ra102) and not unc-26(ra201).

Mapping results revealed that a single suppressor strain contained two new mutations, *unc-26(ra201) IV* and a new gene *dim-1(ra102)* X. In order to determine which of these mutations was the dominant suppressor of *unc-112(r367)* both *unc-26(ra201)* and *dim-1(ra102)* were tested for suppression of paralysis in an *unc-112(r367)* genetic background. The ratio of phenotypes of the progeny from double heterozygote hermaphrodites is predicted for dominant suppression to be 12 Wild type: 3 Rubber Band: 1 Paralyzed with *unc-26(e205).* Since the Dim phenotype cannot be distinguished from Wild type except under polarized light, the ratio of phenotypes of the progeny from double heterozygote hermaphrodites is predicted to be 15 Wild type: 1 Paralyzed for dominant suppression with *dim-1(ra102).* The ratio of progeny phenotypes from *unc-26(e205)/+;unc-112(r367)/+* hermaphrodites was 8 wildtype: 3.5 Rubber Band: 3 Paralyzed, which was closer to the predicted ratio of recessive suppression (9 Wild type: 4 Rubber Band: 3 Paralyzed (9)) than for dominant suppression. In contrast, the ratio of phenotypes from

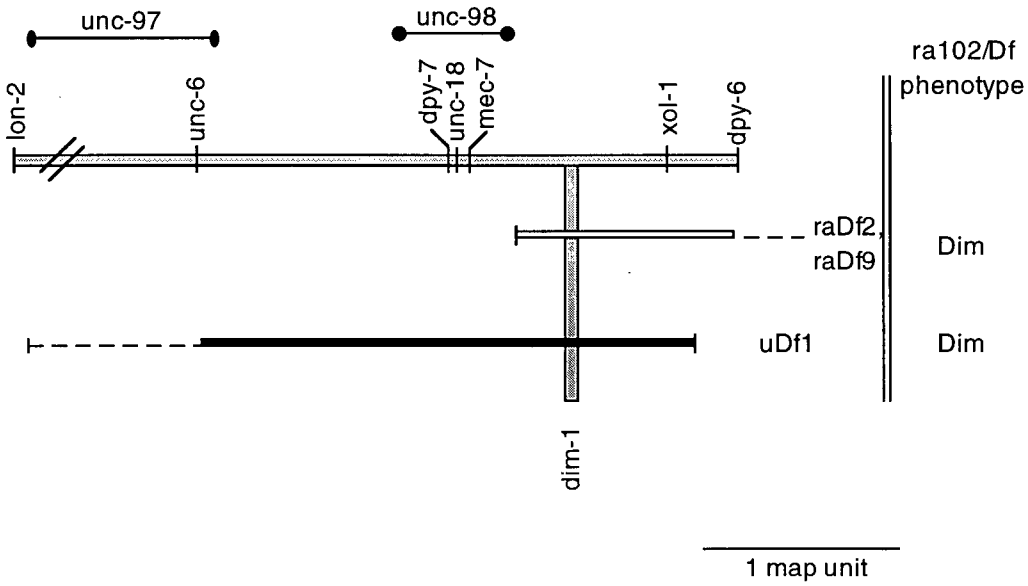


Figure 9: Refinement of the genetic map position of dim-1 by deficiency complementation.

The region of the X chromosome containing *dim-1* is shown (black bar) and neighboring genes are indicated on the chromosome or above it if they have not been precisely mapped. Below the areas deleted by the deficiencies are shown (I-bars) with their names to the right as well as the phenotype of the *dim-1/Df* strains. The position of *dim-1* is shown (gray bar).

the progeny of *unc-112(r367)/+;dim-1(ra102)/+* hermaphrodites was 14.5 wildtype: 1 Bag-of-worms: 0.5 Paralyzed. This ratio is closer to the expected ratio for dominant suppression. However, the Paralyzed class is under represented, and the Bag-of-worms class does not represent the visible phenotype of *dim-1(ra102)*. The Bag-of-worms phenotype resembled nearly-paralyzed, starved, full-sized hermaphrodites that have retained eggs and hatched larvae internally. The appearance of this phenotypic category indicates that *dim-1(ra102)* is not a strong dominant suppressor. The Bag-of-worm category is most likely to be of the genotype *unc-112(r367);dim-1(ra102)/+* and only partially able to suppress the Paralyzed phenotype. This prediction was tested by determining the ratio of phenotypes of the progeny of two Bag-of-worms hermaphrodites. Twice as many Wild type progeny were produced as either Paralyzed or Bag-of-worms. This is the expected ratio if half of the *unc-112(r367);dim-1(ra102)/+* animals looked Wild type and half looked Bag-of-worms. Because the mutagenic reversion screen selected for dominant suppressors in F1 progeny of *unc-112(r367)* hermaphrodites, the *unc-26(ra201)* allele is not the cause of the observed suppression of *unc-112(r367)* paralysis. The source of suppression of paralysis in *unc-112(r367)* is due to the semi-dominant *dim-1(ra102)* allele.

Loss-of-function alleles of dim-1 are unc-112(r367) suppressors.

Suppression of *unc-112(r367)* Paralysis by *dim-1(ra102)* may be due to the loss of function of *dim-1(ra102)*. To test this possibility *uDf1* was assayed for the ability to suppress *unc-112(r367)* paralysis. The *uDf1/+;unc-112(r367)* hermaphrodites were not paralyzed, and their bodywall muscles exhibited the Chevron phenotype. These results indicate that *uDf1* can act as a suppressor of *unc-112(r367)* and confirms that *dim-1(ra102)* is likely to be a null mutation. However, neither *uDf1* nor *dim-1(ra102)* are completely epistatic to *unc-112(r367)* since the bodywall muscles are easily disrupted and do not exhibit wild type sarcomere placement in *uDf1/+;unc-112(r367)* and *dim-1(ra102);unc-112(r367)* animals. This suggests that the absence of DIM1 protein may not completely bypass the requirement for UNC112 function. This hypothesis was confirmed when *dim-1(ra102)* was shown to be incapable of suppressing the lethality of *unc-112(st581)* (D. Moerman, per. comm.).

Identification of additional dim-1 alleles.

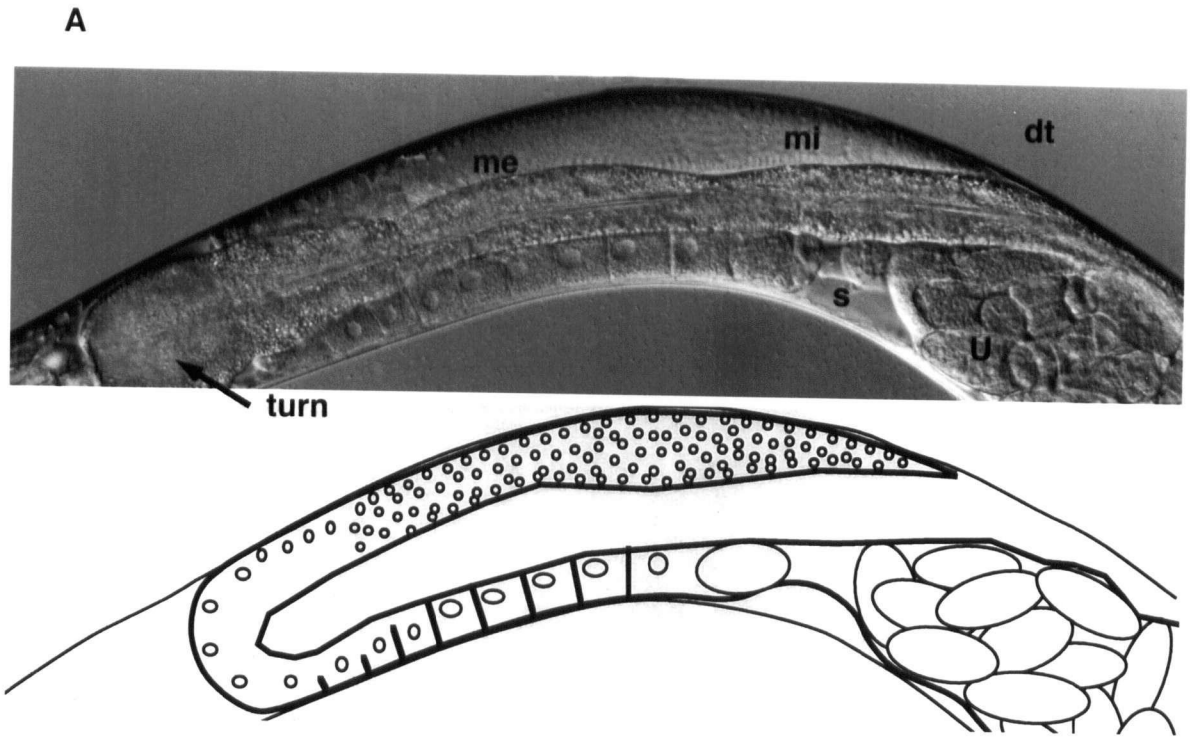
The 48 intergenic revertants obtained from the mutagenesis of *unc-112(r367)* animals provided a pool of mutants from which additional alleles of *dim-1* may be identified. Seventeen intergenic *unc-112(r367)* suppressors were tested by complementation with *dim-1(ra102)*. The Chevron phenotype was found in the bodywall muscles of all 17 sets of suppressor/*dim-1(ra102)* animals using polarized light microscopy. Therefore, all 17 suppressors tested are alleles of *dim-1*. The remaining 30 intergenic revertants have not been examined. The new *dim-1* alleles exhibited a range of phenotypes from normal movement displayed by *ra102* animals to transient immobility displayed by L3-L4 staged *ra204* animals. This suggests that some of these alleles are not null mutations. It remains possible that the complete range of mutations for *dim-1* is not represented in this group of 18 alleles since these revertants are from a suppressor screen for nonparalyzed *unc-112(r367)*-bearing animals. In the future it may be useful to isolate new alleles of *dim-1* using other strategies.

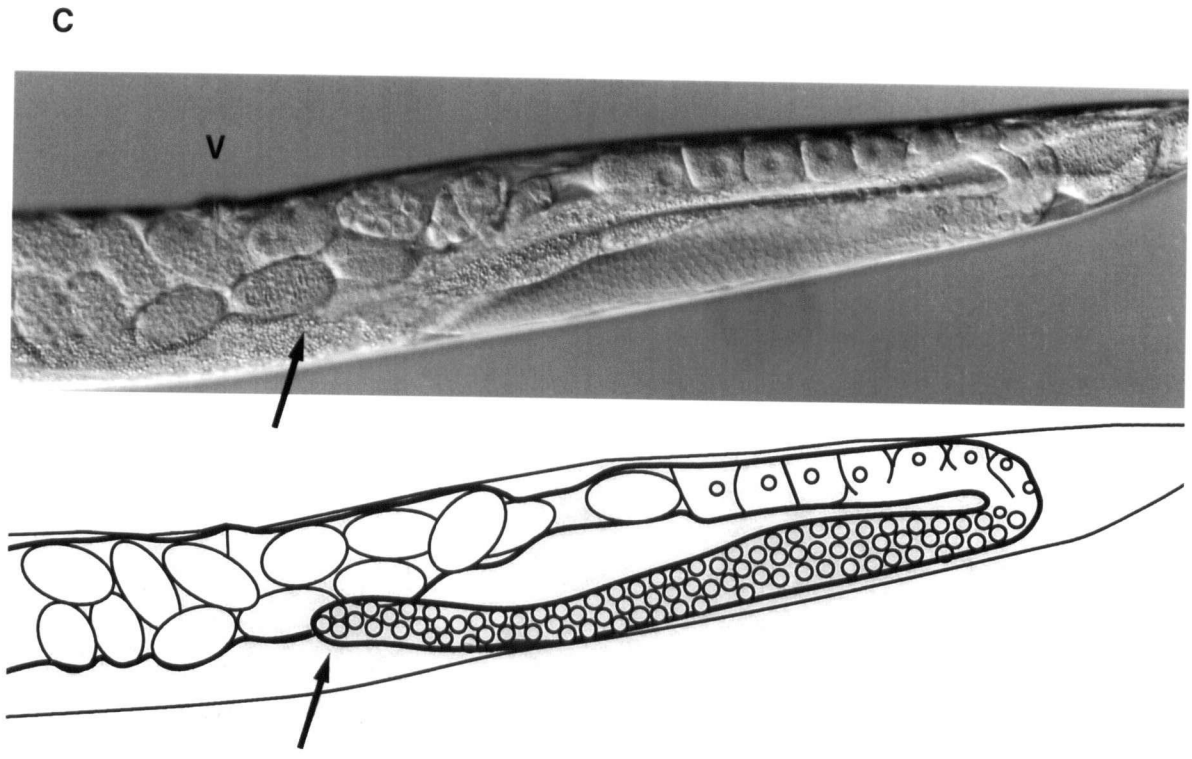
Visible phenotype of dim-1(ra102).

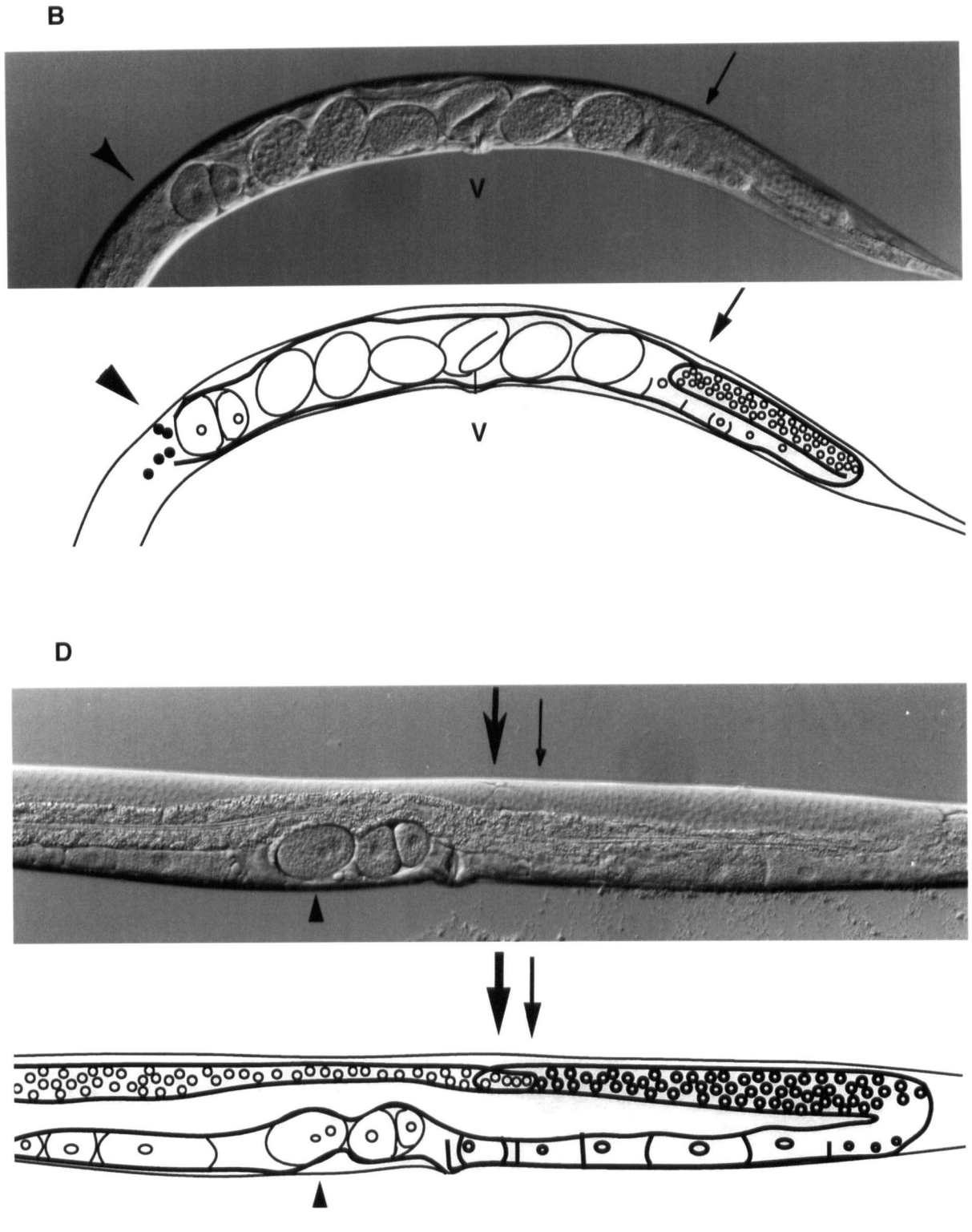
EMS reversion of *unc-112(r367)* resulted in the isolation of several suppressor mutations in a new gene, *dim-1*. The revertant strain DM1243, containing *dim-1(ra102),* was selected for further characterization owing to its unusual non-suppressor phenotype. When viewed under the dissecting microscope, *dim-1(ra102)* animals appeared indistinguishable from N2 animals in appearance and movement. *dim-1(ra102)* animals mature in about 70 hours at 20 °C, male *dim-1(ra102)* animals are capable of mating successfully with hermaphrodites (a trait not common to mutations linked to the X chromosome), and *dim-1(ra102)* hermaphrodites produce approximately the same number of progeny as their N2 counterparts. Furthermore, the gonads of more than 20 hermaphrodites appeared normal when examined with Nomarski optics. However, the distal tip of one gonad arm in one hermaphrodite was slightly deflected toward the vulva (see figure 10). Therefore, at low levels of magnification, *dim-1* mutant animals are indistinguishable from wildtype animals.

Figure 10: Comparison of gonadal morphology of N2, unc-112(r367), unc-112(r367);dim-1(ra102) and dim-1(ra102).

Nomarski images are oriented with the heads to the left. Cartoons beneath each image outline the shape of the gonad and the positions of nuclei, developing oocytes and fertilized eggs. **A**) the N2 gonad arm forms a reflexed tube with the distal tip (dt) terminating on the dorsal side of the uterus above the vulva (V). Mitotic nuclei (mi) are found in the distal portion of the gonad arm. Meiotic nuclei (me) are found in the distal arm closer to the turn (arrow). In the proximal arm, from the turn to the spermatheca (*), cellularization occurs. The fertilized eggs are stored in the uterus for a short time before being expelled through the vulva. **B**) shows an *unc-112(r367)* hermaphrodite posterior gonad distal arm terminating short of its normal position (arrow). The distal arm of the anterior gonad is behind the uterus but the left end of the proximal arm has ruptured releasing several nuclei (arrow head) into the body cavity. **C**) *dim-1(ra102)* hermaphrodites normally form and position the gonad arms but occasionally the distal tip is deflected ventrally (arrow). **D**) shows a young *unc-112(r367);dim-1(ra102)* adult. Notice that the uterus contains two recently fertilized eggs and that the egg and sperm pronuclei in one egg are about to fuse (arrowhead). Frequently, the gonads form normally, but as shown here the tips sometimes overlap (large arrow indicates the tip of the posterior distal arm and the small arrow indicates the tip of the anterior distal arm).


A
me
mi
dt
s
U
turn


C
V


B
V
V
D

Muscle phenotype of dim-1(ra102).

The birefringence pattern of *dim-1(ra102)* bodywall muscle was compared with N2 using polarized light microscopy. Whereas N2 bodywall muscle showed a pattern of alternating parallel rows of birefringent A-bands and dense bodies and darker I-bands and M-lines (see figure 5), *dim-1(ra102)* bodywall muscles showed either parallel rows of bands or a novel nested chevron pattern of bands, of which both patterns were prone to collapse under the weight of the coverglass. The nested chevrons were formed by the intersection at an acute angle of A-bands and I-bands that extended the full length of the muscle cell with much shorter A-bands and I-bands (see figure 5). The Chevron pattern has been observed previously in N2 bodywall muscles near the tail on rare occasions (R.H. Waterston, per. comm.). However, the *dim-1(ra102)* Chevron pattern is novel because it is seen in several cells along the length of a muscle quadrant and is associated with fragility of sarcomeres in all bodywall muscle cells.

Immunofluorescence microscopy using antibodies to vinculin, β-integrin and perlecan was also used to characterize the muscle phenotype of *dim-1(ra102).* The vinculin antibody staining pattern in bodywall muscle cells showed either Chevron-shaped or parallel rows of spots corresponding to dense bodies and brighter spots at the ends of cells corresponding to adhesion plaques (see figure 11). Likewise, the β-integrin antibody staining pattern in bodywall muscle cells showed either Chevron-shaped or parallel rows of spots and stripes corresponding respectively to dense bodies and M-lines and brighter spots at the ends of the cells corresponding to adhesion plaques (see figure 12). Perlecan antibodies diffusely stained the basement membrane overlying the muscle quadrants in *dim-1(ra102)* and also showed brighter punctate and striped staining overlying the dense bodies, adhesion plaques, and M-lines in either parallel rows or Chevron patterns (see figure 13). It therefore appears that the Chevron pattern extends from β-integrin in the muscle cell membrane out to the basement membrane overlying the bodywall muscle cells and into the dense bodies, adhesion plaques, M-lines and filaments.

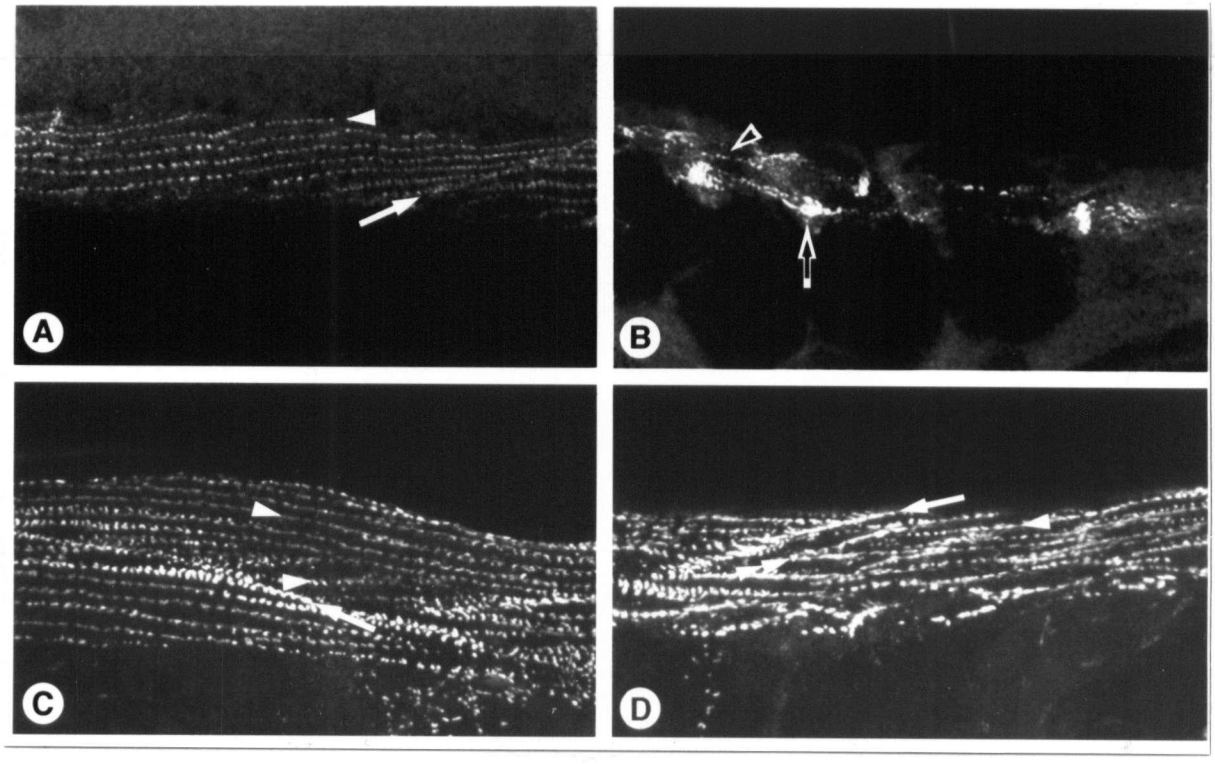


Figure 11: Vinculin distribution in N2, unc-112(r367), dim-1(ra102) and unc-112(r367);dim-1(ra102) with MH24 monoclonal antibody.

Panel **A**) shows the normal staining pattern of a wildtype body wall cell. Vinculin staining is found in both dense bodies (arrowhead) and attachment plaques anchoring half I-bands at the ends of the muscle cell (arrow). **B**) shows the *unc-112(r367)* staining pattern. Only a few short rows are seen (open arrowhead) and clumps of spots (open arrow). It is not possible to distinguish muscle cell boundaries. (For clarity magnification of this panel is two thirds lower than the other three panels.) In **C**) the *dim-1(ra102)* staining pattern is well organized. Rows of dense bodies (arrowhead) run the length of cell and attachment plaques (arrow) are seen at the ends of the cell. However, some cells also contain short rows of dense bodies that intersect with the longer rows, reminiscent of the nested chevron pattern seen with polarized light microscopy (notice the angle of rows of dense bodies indicated by the arrowheads). **D**) shows the staining pattern of *unc-112(r367);dim-1(ra102)* muscle. The organization is quite poor, although not as bad as in B). Only one row of dense bodies appears to terminate on the row of attachment plaques in the cell on the right (arrow and the row of dense bodies beneath the arrow), whereas several rows from the left cell terminate normally. The rows of dense bodies are not aligned in parallel rows but appear to intersect with each other haphazardly (double arrowhead).

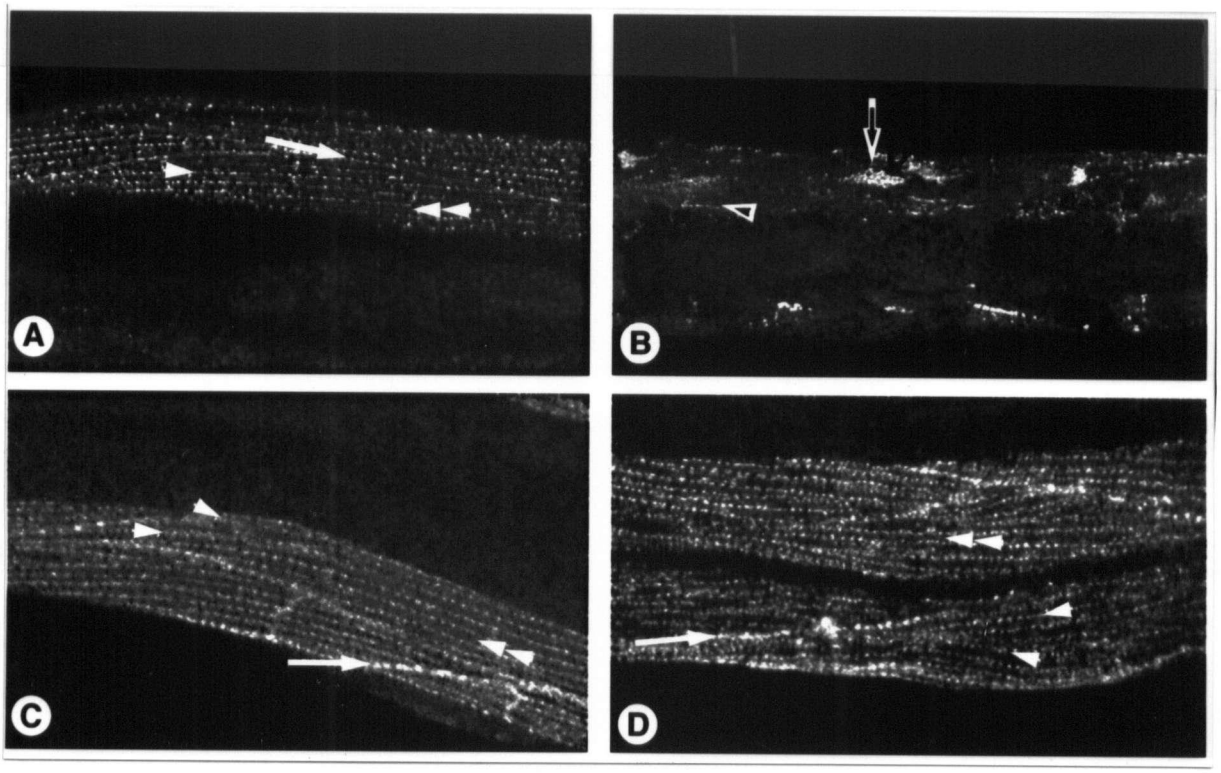


Figure 12:* β-*integrin distribution in N2, unc-112(r367), dim-1(ra102) and unc-112(r367);dim-1(ra102) with MH25 monoclonal antibody.

Panel **A**) shows a wildtype β-integrin muscle staining pattern. β-integrin is found in attachment plaques along cell margins (arrow), dense bodies (arrowhead) and M-lines (double arrowhead). **B**) shows β-integrin distribution in *unc-112(r367)* muscles. Only short rows of punctate material (open arrowhead) or clumps (open arrow) are seen. It is not possible to identify cell boundaries, or distinguish dense bodies from M-lines. This panel is shown at one third of the magnification of the other three for clarity. **C**) shows β -integrin staining *dim-1(ra102)* muscle is well organized but not entirely normal. Attachment plaques at cell margins (arrow) and rows of M-lines (double arrowhead) and dense bodies (arrowheads) are easily identified. Notice the convergence of the indicated short and long rows of dense bodies forming chevrons however. **D**) shows β-integrin staining *unc-112(r367);dim-1(ra102)* muscles. Two muscle quadrants are in focus. Rows of attachment plaques at cell margins (arrow) can be seen but are less uniform than in panels A or C. M-lines are very hard to discern but are present (double arrowhead). Rows of dense bodies (arrows) appear to form chevrons randomly. Notice the indicated rows and the two rows between them all appear to converge at one point just short of the cell margin.

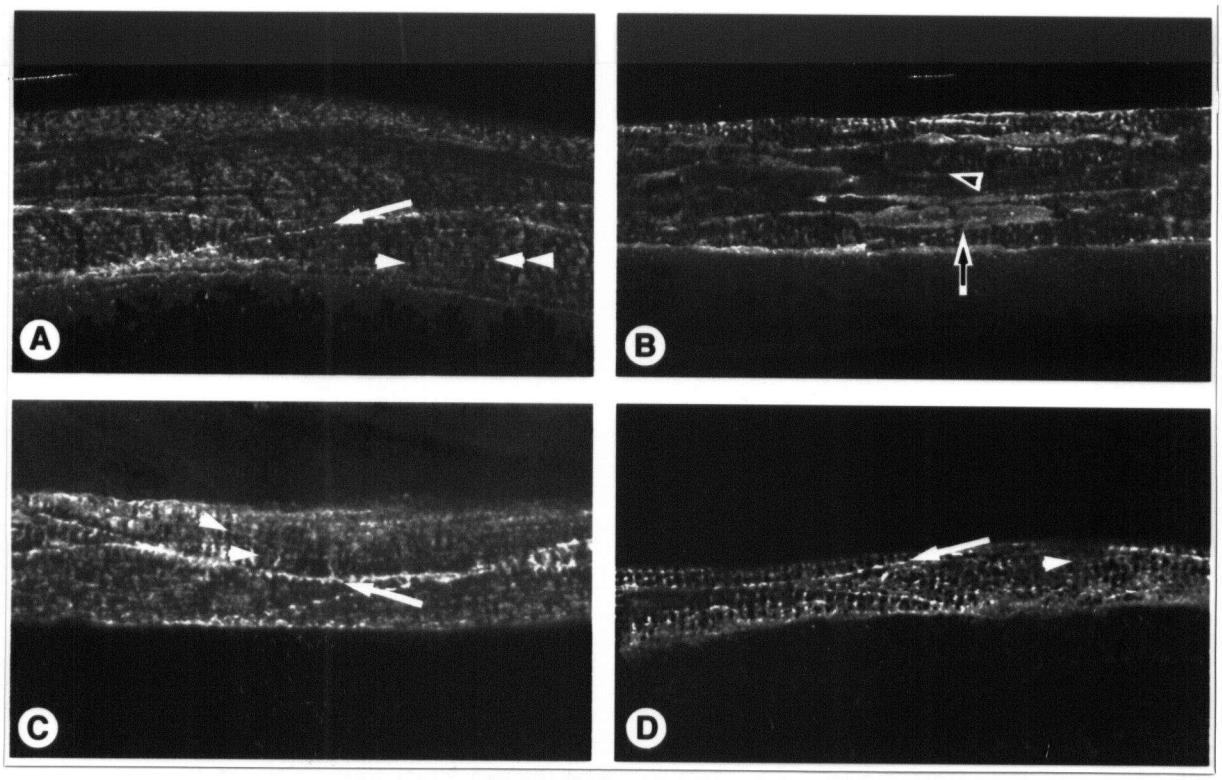


Figure 13: Perlecan distribution in N2, unc-112(r367), dim-1(ra102) and unc-112(r367);dim-1(ra102) with MH3 monoclonal antibody.

Panel **A**) shows wildtype perlecan distribution in the basement membrane as a uniform haze restricted to the area overlying the muscle quadrant, with higher concentrations of staining of the muscle cell margins (arrow), M-lines (double arrowhead) and dense bodies (arrow). **B**) shows perlecan distribution in *unc-112(r367)* animals. Patches of stained basement membrane with no discernible striations (open arrow) and stripes (open arrowhead) are reminiscent of the short rows and clumps seen with MH24 and MH25 staining. No cell margins, dense bodies or M-lines could be distinguished. Notice that MH3 staining remains strictly above the quadrant and does not diffuse into the rest of the basement membrane. **C**) *dim-1(ra102)* staining resembles the wildtype pattern of concentrations over cell margins (arrow), dense bodies (arrowheads) and M-lines and less dense overall staining of the muscle quadrant area of basement membrane. (In this image the intensity of the diffuse perlecan staining makes it very difficult to see the M-lines.) However, the striations form chevrons as seen with MH24 and MH25 staining muscle cells (follow the angle of the two arrowheads). **D**) *unc-112(r367);dim-1(ra102)* perlecan distribution is less uniformly distributed above cell margins (arrow) and the striations (arrowhead) form random chevrons throughout the membrane.

C. Cellular level interactions of *dim-1* and *unc-112*.

Suppression of gonadal defects in unc-112(r367);dim-1(ra102) hermaphrodites.

While the *dim-1(ra102)* allele was originally isolated on the basis of its ability to restore movement to otherwise paralyzed *unc-112(r367)* animals, this was only one aspect of the phenotype manifested by *unc-112(r367)*. Gonadal structure and temperature-sensitivity of fecundity, rate of growth and average size were also affected in *unc-112(r367)* animals. It was possible that *dim-1(ra102)* could also suppress these aspects of the *unc-112(r367)* phenotype. To test this hypothesis, *dim-1(ra102)* was crossed into a homozygous *unc-112(r367)* background, and animals of *dim-1(ra102) /+;unc-112(r367)* and *dim-1(ra102);unc-112(r367)* genotypes were compared to wild type, *unc-112(r367)*, *dim-1(ra102)* and each other for gonadal structure, fecundity, average size, and paralysis (see Table 4). A single copy of *ra102* was found to partially restore gonad morphology, fecundity, and average adult size. However, approximately 50% of the animals moved slowly and produced normal numbers of progeny, while the remaining animals gradually became paralyzed and retained a significant portion of their eggs which then hatched internally and ate their way out of the afflicted hermaphrodites (the "Bag-of-worms" phenotype). In contrast, *dim-1(ra102);unc-112(r367)* animals very rarely exhibited the Bag-of-worms phenotype. Instead these animals moved more slowly than wild type and produced normal numbers of progeny. Examination using Nomarski optics showed that the morphology of one, or both, distal gonad arm(s) of a few *unc-112(r367)* hermaphrodites were short or twisted and that embryos and oocytes in most hermaphrodites from both groups were often mislocalized due to a breach in the membrane of one, or both, gonad arms (see figure 10). Examination of the gonads of *dim-1(ra102);unc-112(r367)* animals showed most were normal and only a few showed defects in morphology or structural integrity. One type of defect seen was of a gonad distal tip extending beyond the normal termination point dorsal to the vulva and overlapping with the distal tip of the second gonad (see figure 10). The gonads of *dim-1(ra102)* animals were indistinguishable from their wild type counterparts except for the rare downward displacement of the distal tip (see figure 10).

TABLE 4

PHENOTYPES OF *dim-1(ra102)* AS A SUPPRESSOR OF *unc-112(r367)*

Genotype	+/+;+/+	unc/+; dim/+	unc/+;dim/dim	unc/unc;dim/+	unc/unc;dim/dim
Overall Phenotype	wildtype	wildtype	wildtype	50% wildtype and 50% bag-of-worms	wildtype
gonad morphology	wildtype	wildtype	wildtype	most are wildtype, some with twists in one arm	most wildtype and a few with distal tip overlaps
fecundity	wildtype	wildtype	wildtype	~95% of wildtype	wildtype
size	wildtype	wildtype	wildtype	wildtype	wildtype
movement	wildtype	wildtype	wildtype	slow	wildtype

Note 1: Abbreviations used are "unc" for *unc-112(r367)*, and "dim" for *dim-1(ra102)*.

Note 2: Average frequencies of progeny phenotypes arising from *unc-112(r367)/+;dim-1(ra102)/+* parent: 92% wildtype, 5% paralyzed, 3% bag-of-worms (number of broods examined = 2; total number of progeny = 653). The same frequencies were obtained from 3 broods of *unc-112(r367)/+;dim-1(ra237)/+* parents (total number of progeny = 749).

Suppression of muscle phenotypes.

Comparison of the sarcomere structure in wild type and each mutant strain may provide clues to the site(s) that are weakened or disrupted by the mutations and provide insight into the functional relationship between *dim-1(ra102)* and *unc-112(r367)*. Adult hermaphrodites of each genotype were examined using polarized light microscopy to reveal the thick filament morphology and arrangement. In wild type *C. elegans* bodywall muscles polarized light generates a pattern of dark bands punctuated with bright dashes (thin filaments and dense bodies) and closely paired bright bands (A-bands which include most of the thick filaments length except for the bare zone and the M-line) (see figure 5). The A-bands in *unc-112(r367)* muscle cells appeared as disorganized, fibrous clumps, and the intensity of the birefringent material appeared to be less in the mutant than in N2 animals (see figure 5). The appearance of the birefringent bands in most *dim-1(ra102)* bodywall muscles were indistinguishable from N2. The distinguishing features were fragility in all bodywall muscle cells and the Chevron pattern in some cells (see figure 5). The birefringent pattern of *dim-1(ra102);unc-112(r367)* resembled the Chevron pattern, but the bands appeared ragged and, in places, the filaments had broken or pulled away from the ends of the cells forming clumps and obscuring the dense bodies in some regions (see figure 5). (This was apparent even though Sephadex beads were supporting the coverslip to prevent stress-induced collapse of the structures.) The ranking of sarcomere morphology and fragility was wild type> *dim-1(ra102)*> *unc-112(r367);dim-1(ra102)*>> *unc-112(r367)*. Thus it appeared that *dim-1(ra102)* only suppressed some of the filament fragility of *unc-112(r367)*.

The M-lines and dense bodies were examined using indirect immunofluorescence to determine how they were affected in *unc-112(r367);dim-1(ra102)* animals compared to N2 and each individual mutant. Immunostaining with mouse monoclonal antibodies either to vinculin, β-integrin, or perlecan was used to examine these components in dense bodies, M-lines, and overlying basement membrane. Vinculin antibodies stained the ends of muscle cells where adhesion plaques anchor half I-bands to the membrane and rows of discrete spots corresponding to the dense bodies in N2 bodywall muscles (see figure 11). In contrast, vinculin antibodies stained clumps of spots in *unc-112(r367)* muscle cells, ragged chevrons and clumps of spots in *dim-1(ra102);unc-112(r367)*, and orderly chevron-shaped rows of spots in *dim-*

1(ra102). β-integrin antibodies stained the ends of muscle cells corresponding to adhesion plaques and rows of discrete spots and thin bands corresponding to dense bodies and M-lines in N2 muscle cells (see figure 12). However, β-integrin antibodies stained clumps of spots in *unc-112(r367)*, ragged chevrons of lines and spots and clumps of spots in *unc-112(r367);dim-1(ra102)*, and orderly chevrons of lines and spots in *dim-1(ra102)* muscle cells. In N2 animals, perlecan antibodies diffusely stained the basement membrane overlying the bodywall muscle quadrants, with higher intensity staining over muscle cell boundaries and rows of dense bodies and M-lines (see figure 13). In *unc-112(r367)* perlecan antibody staining was found in large patches containing various-sized spots over the muscle cells and in ragged bands over muscle cell boundaries. Perlecan antibody distribution was concentrated at the cell boundaries and in chevron-shaped striations overlying *dim-1(ra102);unc-112(r367)* muscle cells; however, the diffuse staining was less uniform and M-lines were indistinguishable from dense bodies. The distribution in *dim-1(ra102)* was similar to N2 except that the staining over the dense bodies and the M-lines was Chevron-shaped. These results generally paralleled the polarized light observations in that the changes in distribution of all three filament attachment components examined were mildest in *dim-1(ra102)*, more severe in *dim-1(ra102);unc-112(r367)*, and most severe in *unc-112(r367)*. All of the results taken together suggested that *dim-1(ra102)* partially suppressed the unstable placement of dense bodies and M-lines in an organized array but did not affect the formation of these structures or of the A-bands.

4. Discussion

A. Analysis of *unc-112(r367)*

The gene *unc-112* is involved in an early and essential step in embryonic muscle development (123) and is also required to maintain the integrity of the sarcomere array from larval stage 3 (L3) through adulthood (91,116). Further characterization of this gene and the identification of other genes with which it

interacts would contribute towards our understanding of the embryonic development of muscle cell-specific structures and provide novel information regarding the mechanisms involved in maintaining the sarcomeres in growing muscles.

In this thesis, the phenotype of the viable allele *unc-112(r367)* has been characterized at the organismal, organ system, and cellular levels. Many aspects of the visible phenotype of *unc-112(r367)* were found to be temperature-sensitive, appearing milder in animals raised at 15 °C and more severe in animals raised at 20 °C. The temperature sensitive characteristics included: embryo to adult growth rate, age of onset of paralysis, median brood sizes, egg-laying ability and size of the animals. This temperature-sensitive variability of phenotype suggests that a mutant protein is synthesized but the protein loses activity with increasing temperature. Often thermolabile proteins are the result of mutations that destabilize folding of the protein, and this may be true of *unc-112(r367)* as well. Consequently, *r367* is unlikely to be a null allele. These predictions could be tested by comparing the DNA sequences of the wild type allele and *r367*.

The small size and thinness of these animals and the slower growth rate appear to be partially the result of inefficient feeding. Since paralysis of the bodywall muscles in *unc-112(r367)* animals arises from the collapse of sarcomere arrays (116, and this study), I tested the possibility that perhaps a similar fate affects the sarcomeres of the pharynx and sex-specific muscles. Examination of vinculin antibody distribution in the pharynx using immunofluorescence microscopy showed approximately 50% fewer spots, particularly in the lower region of the muscle cells of the terminal bulb closest to the intestine. Such a reduction in the number of attachment sites and sarcomeres might be responsible for the thin and sickly appearance of *unc-112(r367)* animals. Similarly, a reduction in the number of sarcomeres in other single-sarcomere muscles in gonadal myoepithelia, uterine and vulva may be responsible for the observed fragility of the myoepithelial sheath and the progressive loss of egg-laying ability. The effects of *unc-112* on pharyngeal and bodywall muscles is uncharacteristic of the majority of muscle-affecting genes placing this gene in a small and select group of muscle genes.

Polarized light and indirect immunofluorescence microscopy were used to determine the regions of the sarcomere in bodywall muscles affected by *unc-112(r367)*. In bodywall muscle cells the birefringent

filaments were distributed in clumps and ragged rows in *unc-112(r367)* adults. Vinculin and β-integrin staining patterns of bodywall muscles consisted of clumps and short rows of spots distributed in a manner similar to the pattern of birefringent filaments. There appeared to be no significant perinuclear staining which might indicate that the proteins were not being incorporated into the sarcomeres. Perlecan antibody staining also included large, diffuse patches of staining over the punctate clumps and rows. If the mutant protein had disrupted a later step in assembly or an association with another structural protein, then the expected result would have been fragmentation at the weakened site in the sarcomere structure and not the observed aggregation of all parts of the sarcomere and overlying basement membrane. This observation is in agreement with earlier observations of unattached filaments in embryos (123), and recent results showing unpatterned β-integrin placement in the muscle membranes of *unc-112(st562)* embryos (G.P. Mullen and D.G. Moerman, unpub. results). These results suggest that *unc-112(r367)* affects either a very early step in the assembly of sarcomeres or is essential for keeping the sarcomeres in position within the muscle cell membrane.

Since the phenotype of *unc-112(r367)* is of mobile larvae that gradually become paralyzed as they mature, the sarcomeres are functional but gradually collapse. Within the detection limits of the microscopic techniques utilized, the results suggest that thin and thick filaments, dense bodies, M-lines and basement membrane components assemble properly into sarcomeres but are tenuously anchored in arrays across the muscle cell membrane. An implication of this would be that the UNC112 protein is not likely to be an structural component of either dense bodies or M-lines. One prediction stemming from this hypothesis is that UNC112 protein may be located either in the muscle cytoplasm or basement membrane very close to, or embedded in, the muscle cell membrane at the sites of dense body and M-line anchoring. A second prediction is that UNC112 may function as a regulatory protein to organize the dense bodies and M-lines into arrays within the muscle membrane (Mechanism 2 in the Introduction section A to this chapter). Both of these ideas could be tested by electron microscopic examination of the filament and attachment structures in the muscles of *unc-112(r367)* animals using UNC112-specific and other muscle- and basement membrane-specific antibodies. The UNC112 pattern of distribution could then be compared with vinculin, β-integrin and perlecan, for example, to determine whether UNC112 co-localizes with these proteins at the interface of muscle membrane with dense bodies and M-lines.

Class I Pat mutations, including *unc-52* and *unc-112* alleles, are determined by the stage of development attained when paralysis occurs and further development of muscle ceases and immunostaining patterns (123). In *unc-52* and *unc-112* Pat mutant embryos immunostaining patterns are very similar: anti-myosin and anti-actin staining reveal primarily fibrous clumps and some diffuse cytoplasmic material (123). However, anti-β-integrin antibody staining patterns on *unc-52* and *unc-112* Pat mutant embryos are quite different: β-integrin staining in *unc-52* Pat embryos is nearly normal, whereas β-integrin staining is in scattered spots over the surface of the muscle cells *unc-112* Pat embryos (G.P. Mullen and D.G. Moerman, unpub. results). Furthermore, the phenotype of *unc-112(st562)* appears to be more severe than *unc-52* Pat alleles since *unc-112(st562)* embryonic muscles were never seen contracting, whereas *unc-52* Pat muscles exhibit weak twitching (B Williams, per. comm.). Taken together, the Pat and embryonic lethal phenotypes indicate that *unc-112* is essential for one of the earliest steps in embryonic muscle development and is also required for maintaining the sarcomeres in their proper sites in larval and adult muscle cell membranes.

The genetic analysis of *unc-112(r367)* provided additional crucial information regarding its chromosomal position and the nature of the mutation. The position of *unc-112* on the genetic map was refined by deficiency complementation with *unc-112(r367)*. The results restricted the gene locus to the right of *par-1* and *egl-1* and in the area deleted by *yDf8* and left of both *yDf11* and *yDf9*. Instead of the expected L3 larval paralysis, the phenotypes of *unc-112(r367)/itDf2* and *unc-112(r367)/yDf8* animals were embryonic lethal. This indicates that not only did the deficiency fail to complement the mutation, but that *unc-112(r367)* is not a null mutation. In comparison, the mutation *unc-112(st562)* isolated by Williams and Waterston (123) exhibits paralysis of embryos at the 1.5-fold stage and developmental arrest by the 2-fold stage (dubbed Pat for paralyzed-and-arrested-at-two-fold). Both the deficiency complementation result and the isolation of lethal alleles of *unc-112* suggests that Pat is the null phenotype of *unc-112*. This was confirmed by demonstrating that both *unc-112(r367)/yDf8* or *itDf2* embryos and *unc-112(r367)/unc-112(st562)* embryos exhibited the same embryonic lethal phenotype.

B. Analysis of *dim-1(ra102).*

The gene *dim-1* is the first member of a novel class of muscle-affecting genes that perturbs sarcomere organization without affecting movement of the animal. Results in this thesis demonstrate that *dim-1* is involved in suppressing gonadal morphology and paralysis in *unc-112(r367)* animals and in the formation and maintenance of the pattern of sarcomeres in bodywall muscles. Further characterization of this gene would contribute towards our understanding of the development of muscle cell-specific structures and provide novel information regarding the mechanisms involved in maintaining the sarcomeres in growing muscles.

Hermaphrodite and male *dim-1(ra102)* animals were examined for defects in morphology, fecundity, and movement and were shown to be indistinguishable from N2 animals. Closer examination of hermaphrodite gonads using Nomarski microscopy showed normal configuration of the gonad arms with the exception of a deflected distal tip in the gonad arm of one animal. This may be an age-related phenomenon and of no significance (J.G. Culotti, per. comm.). A closer examination of the bodywall muscles was also undertaken using polarized light microscopy. Bodywall muscles of *dim-1(ra102)* animals were discovered to be easily disrupted by the weight of a coverglass. In contrast, wildtype animals can be rolled while under a coverslip without disrupting muscle integrity. Consequently, Sephadex beads were used to support the coverglass so that the muscle phenotype could be determined. While most bodywall muscle cells exhibited a Wild type birefringence pattern, some cells showed a Chevron pattern (see figure 5c). The Chevron pattern has been observed at a low frequency in N2 bodywall muscles, but only in cells near the tail (R.H. Waterston, per. comm.). However, the Chevron pattern of birefringence in *dim-1(ra102)* bodywall muscles is seen in cells near the head and midsection of the quadrants in multiple cells and is susceptible to disruption. These results indicate that the *dim-1(ra102)* muscle phenotype identifies a new class of muscle-affecting mutations which alters bodywall muscle organization but not movement.

Indirect immunofluorescence microscopy was used to determine the regions of the sarcomere in bodywall muscles affected by *dim-1(ra102).* The Chevron phenotype was seen using vinculin, β-integrin and perlecan antibodies indicating that *dim-1(ra102)* affects both dense bodies and M-lines from inside the

muscle cells to the basement membrane overlying the muscle cells and does not promote dissociation of any tested portion of these structures. These results suggest that *dim-1* may be involved in placement of dense bodies and M-lines in muscle cell membranes and not directly involved in sarcomere assembly.

Genetic analysis of *dim-1(ra102)* provided additional crucial information regarding its chromosomal position and nature of the mutation. The *dim-1(ra102)* mutation was one of two mutations isolated from the revertant strain DM1243. DM1243 exhibited both a visible "Rubber band" phenotype and a novel muscle phenotype of fragility and, in some muscle cells, a Chevron-shaped birefringence pattern. The two phenotypes were shown through mapping to be the result of a new allele in a known gene, *unc-26(ra201) IV,* and an allele in a new gene, *dim-1(ra102) X.* Semi-dominant suppression of *unc-112(r367)* was shown to be entirely due to *dim-1(ra102).* Complementation tests with 17 other revertant alleles indicated that all were alleles of *dim-1.* Refinement of the map position of *dim-1* with deficiency complementation also suggested that *ra102* was a null mutation. This result further demonstrates that *dim-1* is not essential for viability of the animal and is unlikely to interact directly with *unc-112.*

The Chevron phenotype of *dim-1(ra102)* bodywall muscles seen with polarized light and immunofluorescence with vinculin, β-integrin and perlecan suggests several explanations for the phenotype of the null allele: 1) the loss of a sarcomere structural protein whose function is not critical until the size and strength of the animal exceeds the limits of the defective structure; 2) the failure to modify a structural component for proper assembly of the sarcomere; 3) the loss of a factor that directs the placement of sarcomeres; or 4) the loss of a factor that responds to the placement signaling factor. Both options 1 and 2 favor a role for *dim-1* in the structural composition and assembly of dense bodies and M-lines. I do not believe this is a likely explanation of the data since structural components from inside the muscle cell such as birefringent A-bands or antibody stained components such as vinculin, from within the muscle cell membrane (β-integrin), and outside the muscle cell (perlecan) reveal similar distributions of sarcomere attachments in *dim-1(ra102)* muscles. If *dim-1* was a structural component, then a disruption would have been expected in the patterns of components on either side of the membrane. For example, if *dim-1(ra102)* affected a structural muscle component inside cells between vinculin and the membrane, then A-bands and vinculin would be located in clumps whereas β-integrin and perlecan would appear in near normal patterns. However, these options cannot be removed from the list of possibilities without

examining these structures with antibodies to two components simultaneously to confirm that neither the dense bodies or M-lines are disrupted. Options 3 and 4 favor a role for *dim-1* in patterning of dense bodies and M-lines within the muscle cell membranes. I believe that either of these options are plausible explanations of the data because the attachment structures can be formed and function in the absence of a pattern regulating protein, but will not necessarily be positioned in the membrane in a Wild type configuration.

C. Analysis of interaction between *unc-112* and *dim-1*

The effects of one or two copies of *dim-1(ra102)* on *unc-112(r367)* were determined for size of the animals, movement, number of progeny produced, gonad morphology and generation time. *dim-1(ra102)* animals were indistinguishable from wild type for size, movement and brood size. When viewed with Nomarski optics the distal tips of the gonads occasionally appeared to be slightly displaced. The visible phenotype of *unc-112(r367);dim-1(ra102)/+* animals was similar to wild type in size of the animals, number of progeny produced and generation time, but movement was distinctly slower than wild type in 50% of the population; the other 50% of the population became Bag-of-worms. The visible phenotype of *unc-112(r367);dim-1(ra102)* was indistinguishable from N2 in size, number of progeny produced and generation time, rarely exhibited the Bag-of-worms phenotype, and was slow moving.

Polarized light microscopy and immunostaining were used to compare the distribution of filaments and dense body and M-line attachment components in *dim-1(ra102), unc-112(r367);dim-1(ra102),* and *unc-112(r367).* The filaments in *dim-1(ra102)* bodywall muscles were easily displaced from their normal positions by the weight of the overlying coverslips and formed a nested chevron pattern in some cells but otherwise were Wild type in morphology. The nested chevron pattern, but otherwise normal morphology, of the dense bodies and M-lines was echoed in the immunofluorescent staining patterns of vinculin, β-integrin and perlecan antibodies. The filaments in *unc-112(r367);dim-1(ra102)* bodywall muscles were ragged in appearance and although most filaments were still arranged in chevrons some clumping was evident. The immunostaining results also revealed chevron patterning of dense body and M-line proteins as well as some clumping and some loss of contrast at the boundaries of the staining patterns. The

pattern of *unc-112(r367)* was of fibrous clumps. Similarly, the immunofluorescent patterns were primarily of clumps of spots containing vinculin and β-integrin and overlying patches of perlecan. Thus, the severity of sarcomere disruption from best to worst was: *dim-1(ra102)> unc-112(r367);dim-1(ra102), >>unc-112(r367).*

A number of possible mechanisms may account for the ability of *dim-1(ra102)* to suppress the *unc-112(r367)* phenotype, including informational suppression (14,53,124), gene duplication (70,77) of a multigene family member, and epistasis (84). However, the deficiency *uDf1* acts as a semi-dominant suppressor of the paralyzed phenotype when crossed into *unc-112(r367).* This result eliminates tRNA informational suppression and gene duplication mechanisms since both rely on production of a gene product. Furthermore, this result indicates that suppression is indirect and partially epistatic to *unc-112(r367)* and also confirms that *dim-1(ra102)* is a null mutation. That epistasis is only partial further suggests that the requirement for UNC112 protein cannot be bypassed completely. The viability of both *dim-1(ra102)/uDf1* and *dim-1(ra102)* animals also suggests that *dim-1* may not be essential for survival.

The mechanism by which *dim-1* interacts with *unc-112* is very likely to be indirect because the deficiency, *uDf1*, is a semidominant suppressor of both muscle and gonadal defects in *unc-112(r367).* If suppression of *unc-112(r367)* by *dim-1(ra102)* had been direct (i.e. that it contained a complementary change in its protein product that restored functional interaction with *unc-112(r367)*), then the deficiency would have been unable to suppress *unc-112(r367).* The epistatic suppressor activity of *dim-1(ra102)* was also shown to be only partial because 50% of *unc-112(r367);dim-1(ra102)/+* animals exhibited the Bag-of-worms phenotype. As well, *dim-1(ra102)* was incapable of suppressing the *unc-112(st562)* embryonic lethal mutation suggesting that it may not be expressed in embryogenesis. Thus, *dim-1* null mutations reduce the demand for UNC112 protein but are unable to entirely bypass the requirement for UNC112 protein. Taken together, these results suggest that these two genes may function in the same pathway but do not interact protein-to-protein (see figure 14).

The *dim-1(ra102)* allele has a strong effect on the myoepithelial sheath of the twisted and fragile gonad arms of *unc-112(r367)* animals. The distal tips of the gonads in some *unc-112(r367);dim-1(ra102)* animals extend past the normal termination site rather than twisting or terminating prematurely, the

myoepithelia are less prone to rupture, and the brood sizes are nearly wild type. Homozygous *dim-1(ra102)* animals have normal gonads and number of progeny, but in some instances the distal tips are slightly deflected in the ventral direction. The suppressive effects of *dim-1(ra102)* on the single sarcomere muscles of the pharynx have not been examined directly, but presumably occur since *unc-112(r367);dim-1(ra102)* animals are of wild type length and girth indicating normal feeding capability. Therefore, it appears that the sarcomeres in the myoepithelial sheath may not only move the developing oocytes through the gonad and spermatheca into the uterus, but the M-lines and dense bodies may also be the sites through which the gonad arms are attached to the surrounding basement membrane. This might be examined by immunostaining the gonads of the mutant strains discussed above with antibodies to vinculin, myosin, UNC112 and DIM1 (when available) to determine if there is a correlation between either distribution or amounts of these proteins with gonad morphology or structural resilience.

The *dim-1(ra102)* mutation confers a novel, recessive muscle phenotype. Examination under polarized light revealed some cells with a pattern of A-bands resembling a nested series of chevrons. Most of the bodywall muscle cells in *dim-1(ra102)* animals had a normal A-band phenotype, but the sarcomeres in every muscle were quite fragile and prone to collapse under the weight of the coverslip during polarized light or Nomarski examination. Consequently, the coverslip was supported with Sephadex beads to remove this side effect. An intermediate muscle phenotype was seen in *unc-112(r367);dim-1(ra102)* animals which consisted of ragged chevrons and partially collapsed muscle cells. The distributions of dense bodies, M-lines and filaments from polarized light examination are mimicked in the distributions of the immunostaining patterns seen in *dim-1(ra102), unc-112(r367)* and *unc-112(r367);dim-1(ra102)* animals with vinculin, β-integrin and perlecan antibodies. The normal distributions of these components is consistent with the idea that *dim-1* encodes a protein which may be found in proximity to the dense bodies and M-lines, rather than a structural component of them, analogous to the UNC112 protein.

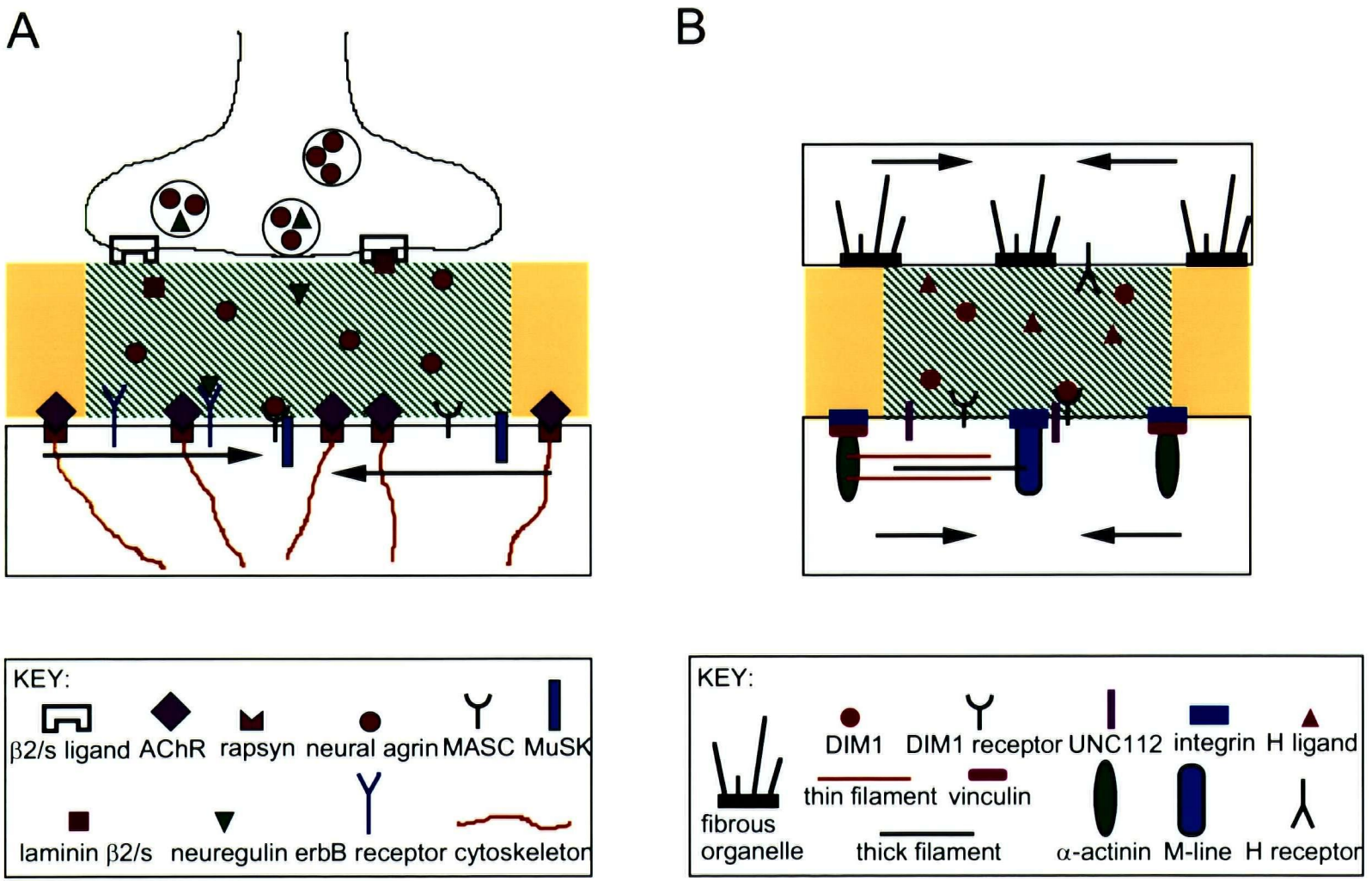
The assembly and placement of the Z-line and M-line complexes is crucial for determining both the orientation and the pattern of sarcomeres in vertebrates and nematodes. In embryos, all muscle types are a single sarcomere long with half-I-bands at the ends of the cells and an M-line in between. As the muscle cells grow in length and width in larvae, new sarcomeres are added to the embryonic sarcomeres

to increase the number of sarcomeres in a row, and at the same time new rows are aligned beside them. I have developed a model for the formation of the pattern of sarcomeres in nematode bodywall muscle quadrants based on the Agrin Hypothesis for the localization of pre- and post-synaptic proteins of the neuromuscular junction (NMJ) (63). There are three significant similarities in differentiation of the NMJ and the hypodermal-bodywall muscle association (HMA) (see figure 14). The first similarity is of a specialized basement membrane between the nerve terminal and the muscle as well as between the hypodermis and bodywall muscle quadrant. The basement membrane in the developing NMJ includes acetylcholinesterase, neuregulin and neural agrin (63,81). The basement membrane in the developing HMA includes isoforms of perlecan (97,98,123). In both cases, the unique basement membrane proteins are not found beyond the interface between the two cells. The second similarity is that both NMJ and HMA sites contain clusters of specialized membrane-bound proteins essential to either signal or force transduction. In the pre-synaptic nerve terminal synaptic vesicles and docking proteins are clustered while on the post-synaptic muscle terminal clusters of acetylcholine receptor/rapsyn/erbB receptor and MASC/MuSK form (18,42,63). Likewise, in the hypodermis, clusters of fibrous organelles (also known as hemi-desmosomes) attach to the cuticle and underlying basement membrane while dense bodies and M-lines are clustered in the muscle cell membrane in apposition to the hypodermis (37,56,123). The third similarity is that the position of the clusters in NMJs and HMAs is maintained by attachments to both the cytoskeleton and the specialized basement membrane (18,37,56,63,123). Based on these similarities, it appears likely that analogous mechanisms for the formation of each structure exist.

The formation of the vertebrate NMJ is an example of reciprocal subcellular differentiation that is initiated by neural agrin secreted into the basement membrane separating the nerve terminals from the motor endplates (63,81) (see figure 14a). Agrin appears to be bound in the basement membrane to heparan sulfate proteoglycans preventing its dispersal outside of the developing NMJ (18,63). Agrin interacts with a receptor on the muscle cell surface that includes MASC which in turn activates MuSK tyrosine kinase (42,63). Tyrosine phosphorylation then initiates clustering of acetylcholinesterase, acetylcholine receptors, erbB, and rapsyn (42,63). The postsynaptic complex organization is maintained by anchoring the clusters to both the overlying basement membrane, in part via heparan sulfate proteoglycans and agrin, and the underlying cytoskeleton via the dystroglycan complex, utrophin

Figure 14: Comparison of structural organization of neuromuscular junction and muscle-hypodermal association.

Panel **A** is based on (63) and shows basement membrane (yellow) and specialized basement membrane (green stripes) indicating the site of a developing NMJ. Clustering of acetylcholine esterase, receptors and other signaling components to the region of the muscle membrane destined to become the motor endplate is initiated by the binding of neural agrin to the MASC/MuSK receptor tyrosine kinase complex. Further development of the motor endplate is triggered by binding of neuregulin to another receptor tyrosine kinase, erbB receptor. The clusters of AChR complexes are held in position through rapsyn to the cytoskeleton on the cytoplasmic side of the membrane and to heparan sulfate-containing basement membrane proteins via disulfide bonds. The nerve terminal is then induced to differentiate by binding muscle-derived laminin β2/s which also attaches the nerve cell membrane to the specialized basement membrane. Panel **B** shows basement membrane (yellow) and specialized basement membrane (green stripes) indicating the site of a developing muscle-hypodermal association. Clustering of dense bodies and M-line structures may be initiated by a signaling molecule (possibly DIM1) binding to a DIM1 receptor complex (possibly including UNC112). Dense bodies and M-lines are held in position through actin-containing thin filaments and myosin-containing thick filaments which interdigitate and crosslink on the muscle cytoplasmic side of the membrane and through integrin complexes to heparan sulfate proteoglycans such as perlecan in the basement membrane. Components in the hypodermis such as fibrous organelles may be induced to cluster and bind to a muscle-derived signal secreted into the specialized basement membrane completing the bridge for transducing the force of muscle contraction to the exterior of the nematode.


A
B
KEY:
β2/s ligand
AChR
rapsyn
neural agrin
MASC
MuSK
laminin β2/s
neuregulin
erbB receptor
cytoskeleton
KEY:
DIM1
DIM1 receptor
UNC112
integrin
H ligand
thin filament
vinculin
fibrous organelle
thick filament
α-actinin
M-line
H receptor

and rapsyn (17,42,45). Rapsyn also associates with acetylcholine receptors and is part of the pathway responsible for directing laminin-β2/s to the synaptic basement membrane (63). Also released from the nerve terminal is neuregulin, which binds to erbB receptors on the motor endplate surface (63). Both MASC/MuSK and neuregulin/erbB complexes induce transcription of postsynaptic-specific and retrograde presynaptic signaling genes (42,63). The secreted muscle-derived signals such as laminin-β2/s in turn arrest axon growth and initiate presynaptic differentiation by binding ligands on the nerve terminal (42,63,81). Thus, the association of the presynaptic nerve terminal and motor endplate occurs via signaling pathways that are physically linked to the presynaptic nerve terminal, NMJ basement membrane and motor endplate.

The formation of the muscle-hypodermal association may also be the result of reciprocal subcellular differentiation (see figure 14b). Laser ablation of dorsal hypodermal cells results in posterior muscle cells that do not migrate to their normal terminal positions and myosin fails to organize into A-bands (M. Hresko, per. comm.). The orientation and spacing of the rows of sarcomeres may be established by an agrin-like signal from the hypodermis that activates a pathway extending from the overlying basement membrane to the interior of the muscle cells that would anchor newly formed dense bodies and M-lines in position in the muscle cell membrane via attachments to the cytoskeleton and a component of the basement membrane (see figure 15a). The rows of sarcomeres may increase in length by addition of new dense bodies and M-lines to the midpoint of the rows as the cell elongates with thin and thick filaments added as the distance between the new attachment sites can accommodate them. A candidate for the muscle equivalent of agrin is *dim-1* because of its unique effect on patterning rows of sarcomeres. Striated concentrations of DIM1 in the overlying basement membrane may induce clustering of preassembled dense bodies and M-lines in the membrane beneath the regions of greatest DIM1 concentration. The absence of a *dim-1* gradient to indicate the appropriate alignment of rows with half-I-bands may permit the rows of dense bodies and M-lines to attach to a half-I-band at one end and a dense body from an adjacent row at the other end (see figure 15b). DIM1 may interact with a protein in a MuSK-like role (perhaps *unc-112*) which associates with dense bodies and M-line complexes but is not an integral part of them. A double mutant lacking the agrin-like signal and with reduced MuSK-like function may have a phenotype similar to the bodywall muscles of *unc-112(r367);dim-1(ra102)*: ragged chevrons

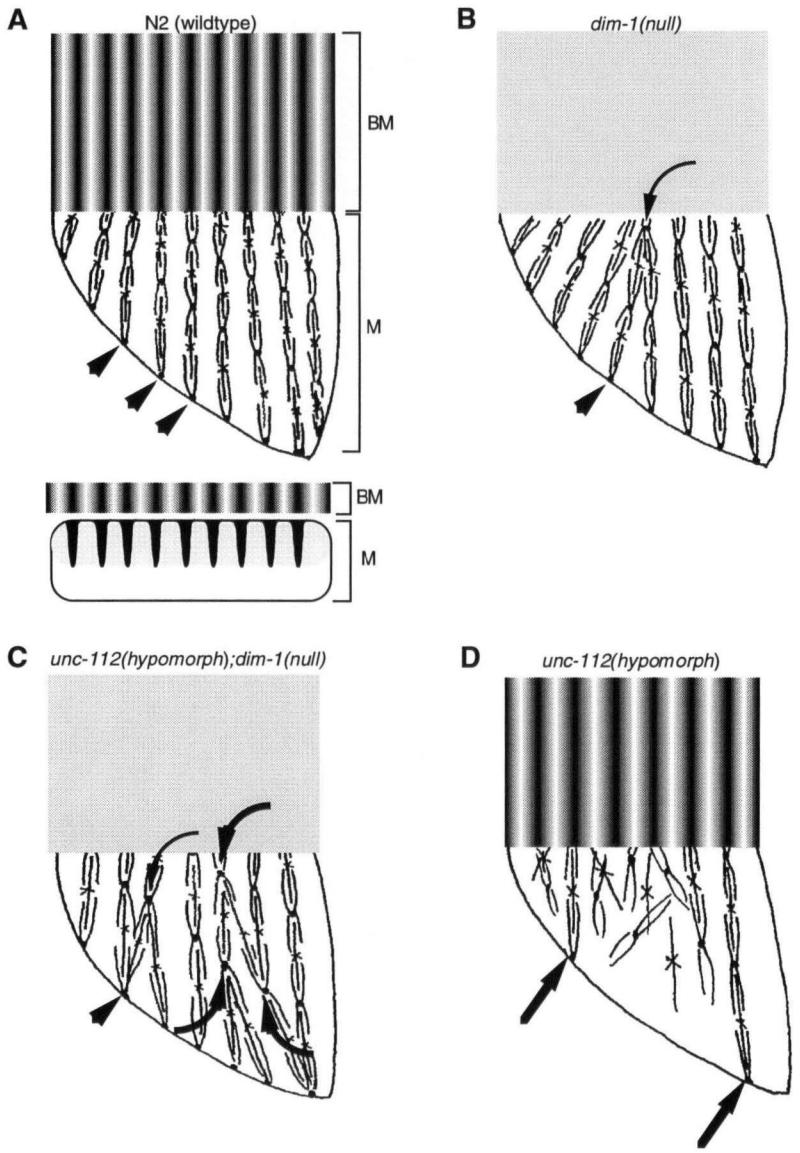


Figure 15: Model for the larval phase of muscle development.

Tangential section of a portion of a bodywall muscle cell is shown partially covered by basement membrane in all panels. **A)** In N2 animals, a wildtype bodywall muscle cell (M) is overlaid with a striped gradient of DIM1-containing basement membrane (BM). A transverse section showing the relationship of the DIM1 gradient with dense bodies and M-lines (solid projections) in the muscle cell is shown in the inset. The half-I-bands at either end of the cell (arrowheads) serve as reference points for connecting new M-lines and dense bodies to increase the length of a row of sarcomeres. The orientation of the new rows of sarcomeres may rely on DIM1 distribution. **B)** In *dim-1(null)* animals some of the growing rows of sarcomeres may align properly and some may align with a dense body (curved arrow) and a half-I-band at the end of a row (arrowhead), thus generating the chevron pattern. **C)** In *unc-112(hypomorph);dim-1(null)* animals which have partially functioning UNC112 receptor and lack the DIM1 signal, a mechanism of random orientation may result in rows of sarcomeres being joined with adjacent rows (bracketed by thick curved arrows) in addition to misalignment of rows with half-I-bands (bracketed by curved arrow and arrowhead). **D)** In *unc-112(hypomorph)* animals which have partially functioning receptor only a few successful rows of sarcomeres may be formed (arrows) even though the overlying DIM1 signal gradient is normal.

consisting of short rows of sarcomeres that appear to originate at a dense bodies of an adjacent row and terminate either at other dense bodies of another adjacent row of sarcomeres or at a half-I-band on the edge of the cell (see figure 15c). In contrast, reduced MuSK-like function with normal distribution of the agrin-like signal may result in a phenotype similar to the bodywall muscles of *unc-112(r367)*. DIM1 may bind to a MASC-like receptor and concentrate the receptor. A reduction of the strength of association of UNC112 with its MASC-like partner would lead to a diffuse distribution of UNC112 across the muscle cell membrane and a fixed distribution of the DIM1 receptor. Such a distribution would reduce the likelihood that the DIM1 signal would be turned into an intracellular response to anchor rows of sarcomeres (see figure 15d). This model may also explain the indirect suppression of *unc-112(r367)* by the loss of DIM1 function. The absence of DIM1 would allow both UNC112 and the MASC-like receptor to encounter each other more often because the MASC-like receptor would not be fixed in position by association with DIM1. Other factors involved in amplifying the clustering signal in erbB/receptor-like roles may be capable of activating the UNC112 complex, but with less efficiency than DIM1, enabling the formation of the ragged chevrons. Furthermore, the model also suggests why DIM1 cannot bypass the requirement for UNC112. The signaling molecule, DIM1, cannot replace UNC112 initiating the intramuscular clustering of dense bodies and M-lines. One way to test the notion that DIM1 functions in a role similar to agrin is to compare the position of sarcomeres in N2 and *dim-1(ra102)* embryos and larvae using green fluorescent protein-labeled MHC A to determine the earliest time that nonparallel sarcomeres can be found. The dynamics of sarcomere addition and intersection of the nonparallel myofibrils with a "normal" myofibril then can be followed with a video camera mounted on a fluorescent microscope.

In turn, signaling proteins may be secreted by the muscle cells into the basement membrane that induces the clustering of the fibrous organelles in the hypodermis to the region overlying the muscle quadrants (see H ligand; figure 14b). A candidate for this role is the *let-805* gene product (M. Hresko, per. comm.). The phenotype of *let-805* mutants is of developmental arrest at 2-fold and partial paralysis. In addition, bodywall muscle cells detach from the hypodermis, and hemidesmosomes fail to localize to the hypodermal membrane adjacent to the muscle quadrants, remaining scattered across the hypodermis. Furthermore, *let-805* mutants fail to stain with the MH46 monoclonal antibody (M. Hresko, per. comm.). The protein recognized by this antibody is synthesized by the hypodermis (M. Hresko, per. comm.). The

cDNA encoding the protein has been cloned and shown to be a very large protein containing 35 fibronectin III repeats and two potential transmembrane domains (M. Hresko, per. comm.). MH46 stains the basement membrane where the muscle quadrants and hypodermis contact each other and appears to associate with hemidesmosomes after the 2-fold stage when muscle contractions have begun (56). Taken together, these results suggest that *let-805* is an excellent candidate as the muscle-derived signal for clustering of hypodermal fibrous organelles over the muscle quadrants.

Chapter 3

Molecular Analysis of the *unc-112* region of chromosome V

1. Introduction

Molecular cloning and analysis of genomic clones and cDNAs support and refine the results of genetic analysis by providing information on the structure of the gene and the predicted protein sequence. The sequence of a cDNA enables the structure of the gene and the regulatory regions flanking the exons to be identified, the sites and sequences of mutant alleles to be determined, predictions to be made regarding the function of the wildtype and mutant proteins, and identification of related genes in *C. elegans* and other species (for examples see 11,12). Furthermore, the sequence information can be used to generate specific nucleotide and antibody probes to study the spatial and temporal distribution of the mRNA and protein in extracts or in situ.

The general process for cloning genes of unknown function and sequence in *C. elegans* at the present time is as follows: 1) the identification of proximal or intragenic molecular tags which serve as reference points to correlate the genetic and physical maps (breakpoints of deficiencies and several types of intragenic mutation are potential sources of molecular tags), 2) cosmid or YAC genomic clones from the region are then screened by hybridization to the tagged genomic DNA to select for the clones that overlap with the molecular tags or lie between tags, 3) the number and size of the transcripts derived from a selected genomic region are determined by Northern analysis, 4) cDNAs corresponding to these transcripts are obtained by selective hybridization of the cosmids to one or more cDNA libraries, 5) the size of the gene is estimated and structure of the gene reconstructed by correlating the sequences of the cDNAs with that of the genomic clones to locate the intron-exon boundaries and the 5' and 3' untranslated

sequences of the gene, and 6) direct proof that the cloned sequences are derived from the genetically characterized gene is obtained by either identifying the nucleotide change that results in the mutant phenotype or by rescuing the mutant phenotype with a cloned wildtype copy of the gene. The second approach involves microinjection of cloned DNA into the gonad syncytium of the parent and its subsequent incorporation into individual gametes (34,35,82). The information acquired at each step in the process contributes to a comprehensive picture of the genomic region surrounding the gene of interest as in addition to the gene itself. The nature and significance of the information derived from the cloning process is described below.

The sequence of unidentified cDNAs can be directly compared with a database of DNA or amino acid sequences from *C. elegans*. Databases such as ACEDB, Genbank and EMBL contain genomic sequences submitted by the *C. elegans* genome sequencing project and to date include a portion of chromosome V to which the *unc-112* gene has been mapped using deficiencies. Consequently, the gene structure can be readily determined by aligning the cDNA sequence with the genomic sequence. Furthermore, the predicted gene structures and sequence homologies of flanking genes contained on a sequenced cosmid are also available on ACEDB. This information is useful for determining whether the gene is a member of a clustered gene family (*act-1,2,3* for example, 70) or possibly is transcribed as a polycistronic message (see for example 102).

The location and type of mutation and the sequence in which the mutation is located provide the context for explaining the phenotype caused by a mutant genotype. For example, changing a tryptophan codon (UGG) to a stop codon (UGA) in exon seven of *unc-52* truncates the protein and results in a null (Pat) phenotype (97). In contrast, the same kind of change in alternatively spliced exon 18 of *unc-52* truncates only transcripts containing exon 18 and results in a Paralyzed phenotype (97). Mutations do not have to occur in coding sequences in order to cause a phenotype. Replacement of a G-nucleotide with an A-nucleotide in the splice donor site of an exon-intron boundary of *unc-52* prevents the removal of the intron between alternatively spiced exons 16 and 17 and animals bearing this mutation are paralyzed (97). If a nucleotide change had involved a splice donor site of an exon required in all perlecan isoforms, the result probably would have been lethal. Consequently, determining the genomic sequence of wild type and mutant alleles provides a molecular context for explaining lethal and viable mutant phenotypes.

Sequence similarity can be exploited to identify other genes containing regions of similarity with the probe gene in *C. elegans* and other species. For example, the presence of a family of highly similar genes in *C. elegans* suggests that isoforms of the protein may be differentially distributed either spatially or temporally. One or more members of a gene family may be recruited by mutagenesis to suppress a defect in another family member. For example, myosin heavy chain isoforms derived from four different genes are spatially distributed in different muscle types and positions within the thick filaments of each muscle type. Myosin heavy chain isoforms derived from *myo-1* and *myo-2* are found exclusively in pharyngeal muscle cells, and *unc-54* and *myo-3* isoforms are found in all other muscle cells (116). A class of *myo-3* mutant alleles is capable of dominant suppression of *unc-54* null alleles (116).

Two types of multigene family have been identified, dispersed and clustered, and different mechanisms are required to recruit them as suppressors of a defective member. The four members of the myosin family, *unc-54 I, myo-1 I, myo-2 X* and *myo-3 V,* exemplify a dispersed gene family (116). Three members of the actin family exemplify the second type, or clustered gene family, and are found in a cluster with two genes in tandem and one inverted on chromosome V (70,120). Mutations in one of the actin genes are pharyngeal muscle specific, while mutations in the other two genes are bodywall muscle specific. The mechanism by which *myo-3* mutant alleles suppress *unc-54* null alleles is duplication of the *myo-3* gene (116). Because *myo-3* and *unc-54* myosin heavy chain isoforms are expressed in the same muscle cells, overexpression of the *myo-3* isoform from multiple copies of the gene can compensate for the loss of functional *unc-54* isoform. In contrast, a mechanism by which members of the actin gene cluster suppress mutations in another member is by genomic rearrangement within the cluster. Reversion of homozygous-viable actin mutations results in loss-of-function revertants that map within the gene cluster. The reversion mutations delete the defective gene by fusing two of the three genes together into one functional gene, probably via unequal crossover (70).

The identification and cloning of cDNAs associated with polymorphisms in *unc-112* mutants will provide tightly linked genes and markers that can be used to further delimit the location of the *unc-112* gene. In addition, the characterization of the cDNAs may identify novel genes that may not have been identified by genetic mutation and merit further study independent of *unc-112.* Ideally, the combined results of these molecular studies on clones proven to be derived from *unc-112* would significantly

enhance the evaluation of the proposed mechanisms of *unc-112* function in bodywall muscles. Identification of sequences related to *unc-112* may suggest which mechanisms could be eliminated. Determination of the mutational changes and the context in which they are found in the various alleles of *unc-112* would enable more detailed versions of the remaining functional mechanisms to be proposed and tested.

2. Materials and Methods

A. Molecular mapping of deficiency breakpoints

Nematode strains were grown in complete S medium containing 5% OP50 (52) at 20 °C with shaking (200 rpm) until near starvation. Nematodes were collected by centrifugation at 1500 rpm in 15 ml conical glass tubes. Pellets were resuspended in M9 buffer and sedimented 3 times to remove debris and remaining bacteria. Pellets were resuspended in a 50 ml M9 and incubated at 20 °C with shaking for 30 min. to allow elimination of bacteria from within the nematodes. Collection was as above and the pellets were washed once in M9 . Pellets were resuspended in 5 volumes of lysis buffer (lysis buffer: 100 mM Tris pH 8.0, 50 mM EDTA pH 8.0, 200 mM NaCl, 1% SDS) and Proteinase K added to 200 μg/ml. The solutions were incubated at 55 °C until the nematode carcasses could no longer be seen under the dissecting microscope. The solutions were extracted 3 times with phenol:chloroform, once with chloroform, and centrifuged at 10 krpm for 10 minutes at 4 °C. Ethidium Bromide (EtBr) was added to 100 μg/ml and solid CsCl to 1.55 mg/ml. The samples were centrifuged in a VTi65 vertical rotor at 55 krpm and 25 °C for a minimum of 15 hr. Genomic DNAs were collected and the EtBr was extracted with NaCl-saturated butanol. The CsCl/DNA solutions were diluted with 2 volumes of H_2O and precipitated with 6 volumes of 95% ethanol for 1 hr or at -20 °C overnight. The DNA was centrifuged at 10 krpm for 20 min.

at 4 °C and precipitated a second time. The pellets were washed once in ice-cold 70% ethanol, dried in a Speedvac concentrator (Savant SC100) and resuspended in TE (TE: 10 mM Tris pH 8, 1 mM EDTA pH 8).

Approximately 10 μg each of N2, *yDf11/nT1(m435)(IV:V),* and *unc-42(e270)yDf9/nT1(m435)(IV:V)* DNAs were digested singly with Pst I and Xba I, and then separated on a 0.8% agarose- 1X TBE gel (TBE: 89 mM Tris-borate, 89 mM boric acid, and 2 mM EDTA). The gel was stained with 1μg/ml EtBr and photographed. The alkaline transfer of the DNA fragments to Nytran (Schleicher & Schuell) membrane was performed as per manufacturer instructions. The DNA was fixed to the membrane by exposure to UV light for 3 min.

Probes were synthesized by random hexamer extension incorporating [α-^{32}P] dATP, 3000 Ci/mmole, using kits from BRL or Boeringer Mannheim Biochemicals and following manufacturer instructions. The probes used were cosmids extending from one containing *unc-76* to one containing TCPAR1: C08C1, C25D7, W09D12, K01G8, W06A7, R15H1, C18G4, C54G10, T25B1, W09A6, F45A9, F08H9, C01G9, ZK868, C09G11, C35C4, ZC256, ZC376, T20C8, T02B5, D1086, T08G5, K10F6, T20G6, and EEEH7 (see Figure 16). Each probe was individually hybridized to the genomic Southern blots in hybridization solution (Hybridization solution: 5X SSPE (76), 5x Denhardts Solution (76), 1% SDS, and 100 μg/ml sheared herring sperm DNA). The Southern blots were prehybridized at 65 °C for at least an hour. The denatured probe in fresh hybridization solution was added and the incubation continued at 65 °C overnight. The blots were washed at room temperature once in 1x SSPE, 1% SDS for 10 min., then washed one to two times more at 65 °C in 0.1x SSPE, 1% SDS. The blots were then exposed to XAR film (Kodak) with an intensifying screen at -80 °C for 24-48 hr.

B. Mapping RFLPs in two intragenic revertant strains

DNA was prepared from liquid cultures of *unc-112(r367)* and four intragenic revertant strains as described in the previous section. The intragenic strains were *unc-112(r367ra202), unc-112(r367ra205), unc-112(r367ra206),* and *unc-112(r367ra207).* Approximately 10 μg of N2, and each of the 5 mutant DNAs were digested with Hind III or Kpn I and resolved on a 0.8% agarose-1x TBE gel. The gels were

EtBr-stained and photographed. The DNA fragments were then transferred to Nytran, fixed under UV light, and prepared for hybridization as described in the previous section. A 2.5 kb Kpn I-EcoRI subfragment of cosmid T02B5 was used to probe the Hind III-digested genomic filter. Hybridization and washing conditions were as described in the previous section. Cosmid T02B5 DNA was digested with Kpn I, and subsequently with both Kpn I and EcoRI, and the fragments separated on a 0.8% agarose-1x TAE gel (TAE: 40 mM Tris-acetate and 1 mM EDTA). Each of the fragments was purified using Geneclean II (BIO 101) followed by ethanol precipitation and resuspended in TE. Probes were synthesized from each fragment and then hybridized to the Kpn I -digested genomic Southern blots as described above.

C. RNA analysis of cosmid fragment

Near-starvation liquid cultures of N2s were subjected to the alkaline-hypochlorite protocol to isolate unhatched eggs (106). Some egg preparations were saved and some were used to start synchronous liquid cultures which were grown for 17 hrs, 2 days, 3 days, and 4 days to collect L1, L2, L3-L4, and adult populations. Mixed and staged populations were washed several times with M9 buffer to remove bacteria as previously described. The pellets were then resuspended in 5 volumes of ice-cold 4M guanidine isothiocyanate, 1% sarkosyl, 25 mM EDTA pH 7, 10 mM sodium phosphate pH 7, and 100 mM β-mercaptoethanol. The nematodes were then disrupted in a French Press at 10,000 psi followed by centrifugation at 10 krpm for 10 min. The cleared supernatants were transferred to fresh tubes and the RNA isolated by the method of Chirgwin et al (21).

Forty μg of each total RNA was denatured at 65 °C for 10 min. in 50 μl of 1x MOPS buffer, 7% formaldehyde and 50% formamide. The samples were separated on a 1% agarose gel containing 1x MOPS buffer and 7% formaldehyde at 70 V for 5 hours. The RNA was transferred to Hybond-N (Amersham) membrane according to Maniatis et al (76). The RNA was crosslinked to the filter with a 3 min. exposure to UV light. The filter was prehybridized in 5x SSPE, 5x Denhardts, 0.5% SDS, 100 μg/ml herring sperm DNA for 30 min. at 65 °C. The probe was a random hexamer labeled, 2.5 kb Kpn I-EcoRI

fragment of cosmid T02B5 in fresh hybridization solution. Incubation was at 65 °C for 2 days. The filter was washed and autoradiographed in the same manner as the Southern blots described above.

D. Isolation of cDNA clones

A cDNA library, cloned into λZAP (Stratagene) by R. Barstead, was plated to a density of 50,000 plaques on five 150 mm LB plates using bacterial strain BB4 as per manufacturer instructions. Plates were incubated at 37 °C for ~8 hrs and cooled to 4 °C. Duplicate sets of nitrocellulose filters (Schleicher & Schuell) were made. Filter-bound phage were then lysed and denatured in 0.1 M NaOH, 1.5 M NaCl and neutralized in 2x HPB (2xHPB: 1 M NaCl, 0.2 M Na_2HPO_4 pH 7, 0.01 M EDTA). Filters were air-dried then vacuum-dried at 70 °C for 2 hrs. to crosslink the DNA to the filters. The probe was random-hexamer labeled 2.5 kb Kpn I-EcoRI fragment of cosmid T02B5. Filters were pre-wet in a solution of 1x HPB, 1% sarkosyl, 100 μg/ml herring sperm DNA. Denatured probe was added to fresh hybridization solution (lacking sperm DNA) and hybridized to the filters at 65 °C overnight. Filters were washed as described for Southern blots above and autoradiographed. Twenty positive-containing plugs were transferred into λ dil buffer (λ dil buffer: 10 mM Tris pH 7.5, 10 mM $MgSO_4$). The 6 strongest positives were passed through secondary and tertiary screens to obtain pure strains. The cDNA-containing plasmids were excised from the λ arms according to manufacturer instructions. The clones obtained from this screen were called p2A2, p2B1, p3A1, p5A2, p10A2, and p10C2.

Two additional cDNA libraries constructed by R. Barstead were screened to obtain cDNA clones extending further 5' than pU112-1 (see below). The first library, λACT-RB1, consists of oligo dT-primed cDNA inserts. The second, λACT-RB2, consists of random primed cDNA inserts. The λACT vector is a yeast-bacterial shuttle vector described in Durfee et al (28). The bacterial host used is RB3E, a modified version of LE392, constructed by R. Barstead, which does not support the growth of non-lambda phage. Approximately 5 x 10^5 plaques of each library were screened as described above. The probe used in the screening was a 1 kb Kpn I-Hpa I fragment from pU112-1 covering the 5' third of the open reading frame. The lambda DNAs were purified by the method of Maniatis et al (76). The plasmids containing the cDNA clones were excised from the lambda DNAs by Not I digestion, followed by circularization with T4 DNA

ligase. Electroporation at 2.5 kV/ 25 mA was used to transform 40 μl of DH5α cells in water with ~300 ng of each ligation mix. The cells were allowed to recover in SOB for 1 hr at 37 °C (SOB: 2% Bacto-tryptone, 0.5% Bacto-yeast extract, 0.05% NaCl). The cell were plated on LB plus 100 μg/ml ampicillin plates and grown overnight at 37 °C (LB: 1% Bacto-tryptone, 0.5% Bacto-yeast extract, and 1% NaCl). The clones obtained from this screen were called pACT2-6, pACT2-7, pACT2-8a, pACT2-8b, and pACT2-9.

E. Test for suppression of paralysis with an extrachromosomal array

The strain BC5519 was obtained from D. Baillie, and constructed by D. Janke by microinjection of BC842, a male N2 strain, with a mixture of 20 ng/μl ZC376 and 80 ng/ μl pCes1943[*rol-6(su1006dm)*]. BC5519 hermaphrodites were crossed to *unc-112(r367)*/+ males and 30 individual Roller, hermaphrodite progeny were allowed to self-cross. Plates containing Paralyzed F2 progeny were scored for the frequency of Wild type, Roller, Paralyzed and novel phenotypes. To confirm the result that no plates contained only Roller and Paralyzed phenotypes, 34 Roller hermaphrodites were allowed to self-cross. The frequency of Roller/Wild type/no Paralyzed, Roller/Wild type/Paralyzed, Roller/Paralyzed/no Wild type, and Sterile-with-protruding-vulva phenotypes of F3 progeny was determined.

3. RESULTS

A. Molecular mapping of *unc-112* genomic region.

RFLP-mapping of the Df breakpoints.

Correlating the information from the deficiency complementation tests with molecular restriction fragment length polymorphisms (RFLPs), which mark the deficiency breakpoints in these chromosomes,

would restrict the region of DNA to be searched for the *unc-112* gene. Consequently, genomic DNA was isolated from N2, *yDf11/nT1(m435)(IV:V),* and *unc-42(e270)yDf9/ nT1(m435)(IV:V)* animals. The different DNA samples were digested with either Pst I or Xba I enzymes for Southern analysis. A series of genomic Southern blots were probed with 25 representative non-overlapping cosmids covering the approximately 1 Mb region from *unc-76* to a Tc1-generated polymorphism lying left of *par-1*, TCPAR1 (see Figure 16). Cosmids from the left half of the region hybridize strongly to a few bands and weakly to many bands, indicating an extensive distribution of repetitive sequences throughout the region. This phenomenon made definitive deficiency breakpoint RFLP identification difficult (see discussion). An RFLP was located with cosmid R15H1 in *yDf11* and another RFLP was located with cosmid ZC256 in *yDf9* (see Figure 17). These results suggest that *yDf9* and *yDf11* deficiencies may break within the regions covered by these cosmids. An RFLP identifying the left breakpoint of *yDf8* could not be located in any of the cosmids tested and therefore may lie in the gap between D1086 and T02B5. Since the genetic deficiency complementation tests demonstrated that both *yDf9* and *yDf11* did not uncover the *unc-112* gene, the DNA to the right of ZC256 was excluded from further analysis. The focus of the search for the *unc-112* locus was restricted to the region to the left of ZC256, covered by 8 cosmids and 2 gaps.

B. Identification of *unc-112*-associated polymorphisms.

RFLP-mapping of the intragenic revertants.

EMS occasionally causes small deletions in DNA, so the collection of intragenic revertants was screened for RFLPs with cosmid probes to the left of ZC256. Genomic DNA from *unc-112(r367), unc-112(r367ra202), unc-112(r367ra205), unc-112(r367ra206),* and *unc-112(r367ra207)* animals was digested with Kpn I. The 8 cosmids were used to probe genomic Southern blots. The cosmids hybridized to a few strong bands and many weak bands in both N2 and mutant DNA lanes. Two cosmids, T20C8 and T02B5, identified polymorphic bands of 2.5 kb and 1.5 kb in Kpn I-digested DNA from *unc-112(r367ra202)* and another of 4 kb in Kpn I-digested DNA from *unc-112(r367ra207)* (see figure 18, figure 24, and Appendix B). Since these two cosmids constitute a small "island" of clonable DNA flanked by gaps and because T20C8 was difficult to isolate intact (data not shown), it is possible that these RFLPs are rearrangements

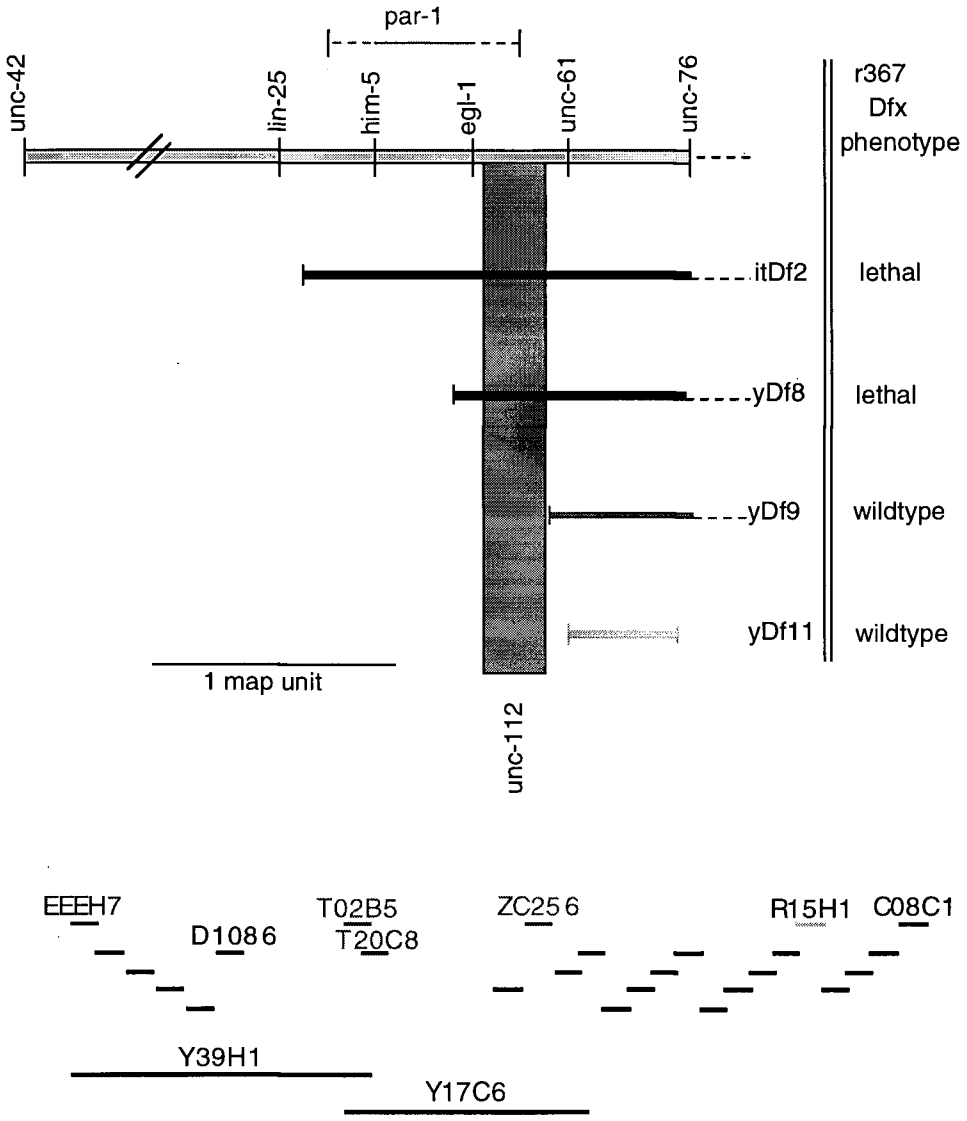


Figure 16: Physical map of the par-1 to unc-76 region of chromosome V.

The estimated amount of DNA from the TCPAR1 site to *unc-76* is 900 kb- 1,000 kb. The top line represents the physical map of the region containing *unc-112* and includes the position of the *unc-76* clone and a Tc1 polymorphism. The 25 cosmids used to span this region are positioned below the physical map. Each cosmid contains approximately 35 kb of genomic DNA. Below the cosmids are the YACs used to span the gaps on either side of the cosmid island including T02B5. Y39H1 is ~290 kb and Y17C6 is ~210 kb. The regions deleted by *yDf9* and *yDf11* are shown below the YACs. The cosmids containing the cloned gene or polymorphisms are identified by name.

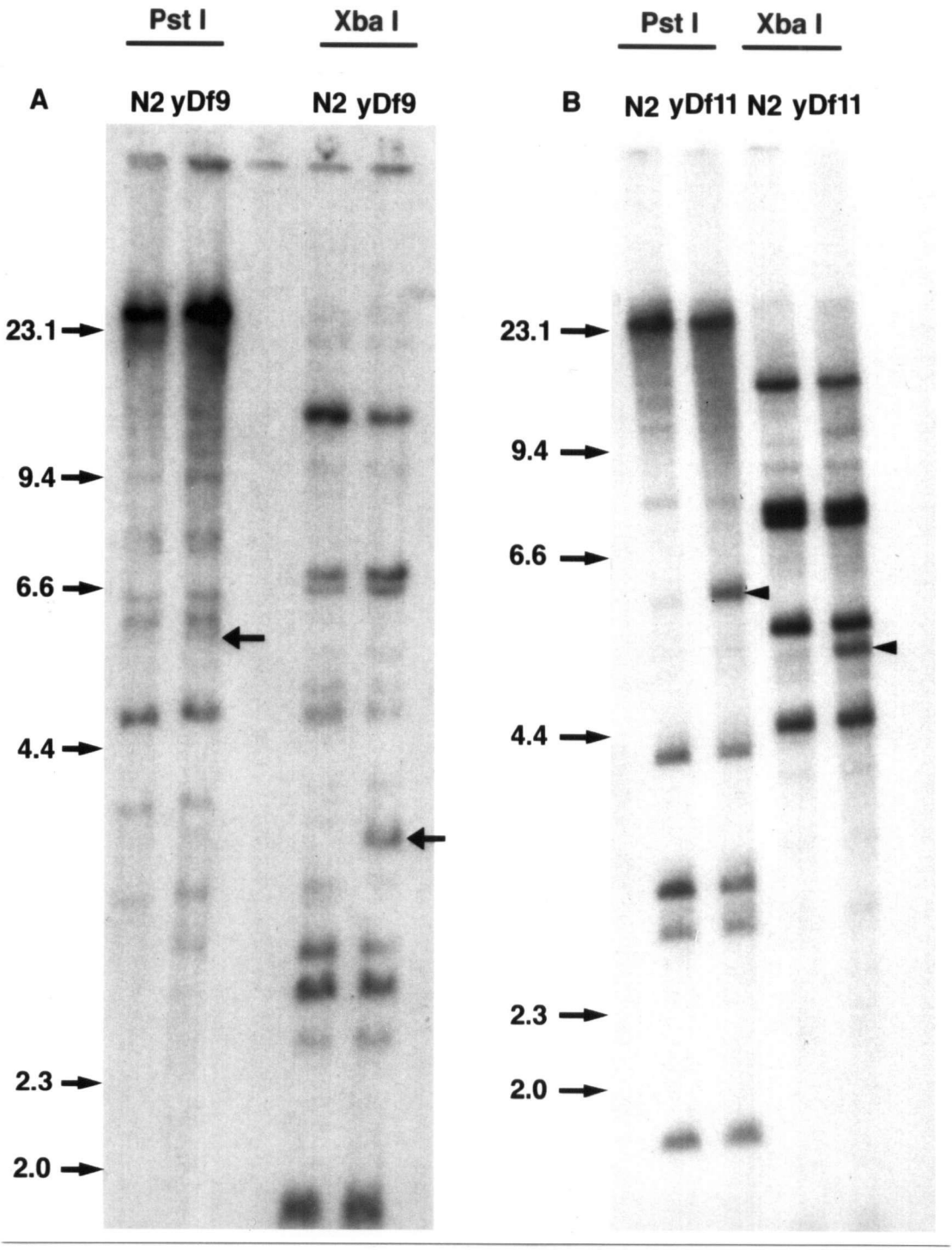


Figure 17: Cosmid identification of the left breakpoint RFLPs of yDf9 and yDf11.

Panel **A**) shows the hybridization pattern of the cosmid probe ZC256 of genomic DNA from N2 and each deficiency-bearing strain (*yDf9* or *yDf11*) digested either with Pst I or Xba I as indicated at the top of each lane. The polymorphic bands in *yDf9* DNA digests (arrows) are not present in the N2 lanes. Panel **B**) shows the pattern of cosmid probe R15H1 on N2 and *yDf11* genomic DNAs digested either with Pst I or Xba I. The polymorphic bands in *yDf9* DNA digests (arrowheads) are not present in N2 DNA digests. The molecular weight marker positions are given down the left side of the blots.

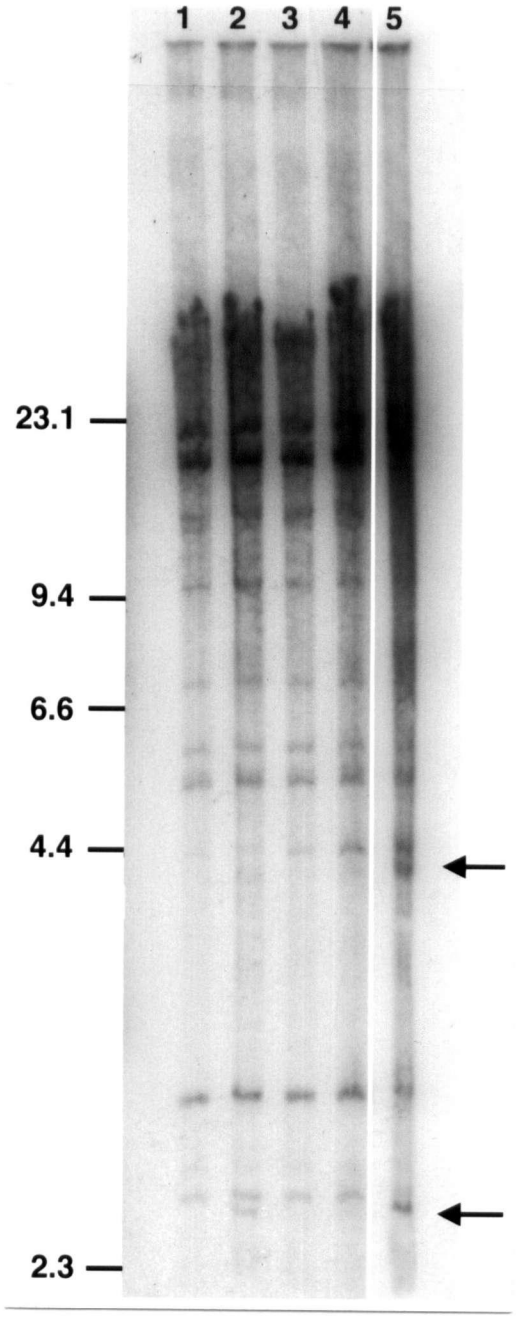


Figure 18: Identification of an unc-112(r367ra202) associated polymorphism with cosmid T20C8.

Genomic DNA from N2 (lane 1), *unc-112(r367ra202)* (lane 2), *unc-112(r367ra205)* (lane 3), *unc-112(r367ra206)* (lane 4), and *unc-112(r367ra207)* (lane 5) were digested with Kpn I and probed with cosmid T20C8. The molecular weight markers are given on the left side and the positions of the polymorphisms are indicated by arrows on the right side of the figure. Lane 5 is from a shorter exposure of the same Southern. The polymorphism in lane 2 is estimated to be 2.5 kb and in lane 5 is estimated to be 4 kb.

unrelated to the *unc-112* reversion events. However, these were the only polymorphisms identified within the 8 cosmid region defined by *yDf11* and TCPAR1 and therefore were good candidates as molecular tags for *unc-112* cloning and required further investigation.

Cosmid T02B5 is approximately 40 kb. In order to refine the position of the RFLPs within the region spanned by T02B5, the cosmid was digested into 6 smaller Kpn I fragments. Each fragment was used to probe a Southern blot of Kpn I-digested N2 and *unc-112(r367ra202)* DNA (see figure 19). The 7 kb, 3.7 kb, 2.3 kb, and 1.7 kb fragments hybridized primarily to genomic bands of the same size as the probe. Both the 16 kb and 9.5 kb fragments hybridized to a variety of genomic bands that totaled more than the probe size. This indicated that both of these regions contain repetitive DNA sequences. However, RFLPs were apparent between N2 and *unc-112(r367ra202)* DNA at 1.5 kb, 2.8 kb, and possibly a third at 6 kb in both the 16 kb and 9.5 kb hybridized panels. Due to hybridization with the repetitive sequences in both probes it was not possible to discern from this experiment whether the N2 16 kb and 9.5 kb genomic bands were of decreased sizes in *unc-112(r367ra202)* which would indicate the DNA in the new band was derived from mutation of the wildtype band.

The 16 kb and 9.5 kb Kpn I fragments from T02B5 were large enough that either could contain sequences from more than one gene. The 9.5 kb fragment was arbitrarily selected to be subdivided and tested for the ability to recognize the RFLPs in *unc-112(r367ra202).* The fragment was digested by EcoRI into 4 subfragments of 4 kb, 2.5 kb, 1.5 kb, and 1 kb. Each subfragment was used to probe a Southern blot containing N2 and *unc-112(r367ra202)* genomic DNA digested with Kpn I (see figure 20). Only the 2.5 kb EcoRI subfragment hybridized to bands of 9.5 kb, 1.5 kb, and fewer background bands. The 2.5 kb probe recognized a genomic Kpn I band of the same size as the cosmid band from which the probe was derived. This result indicated that this portion of T02B5 corresponds well with the genome and was unlikely to be rearranged (a concern due to the spontaneous deletion occurring in the overlapping cosmid T20C8). The result also focused attention on the 1.5 kb RFLP since it had the same intensity as the wild type band. Confirmation that the correct RFLP had been identified was obtained by probing a Southern filter including N2 and the four *unc-112* mutant DNAs digested with HindIII with the 2.5 kb EcoRI T02B5 subfragment (see figure 21). Only *unc-112(r367ra202)* and *unc-112(r367ra207)* lanes contain polymorphic bands in addition to the wild type bands. Furthermore, the polymorphic bands are of different

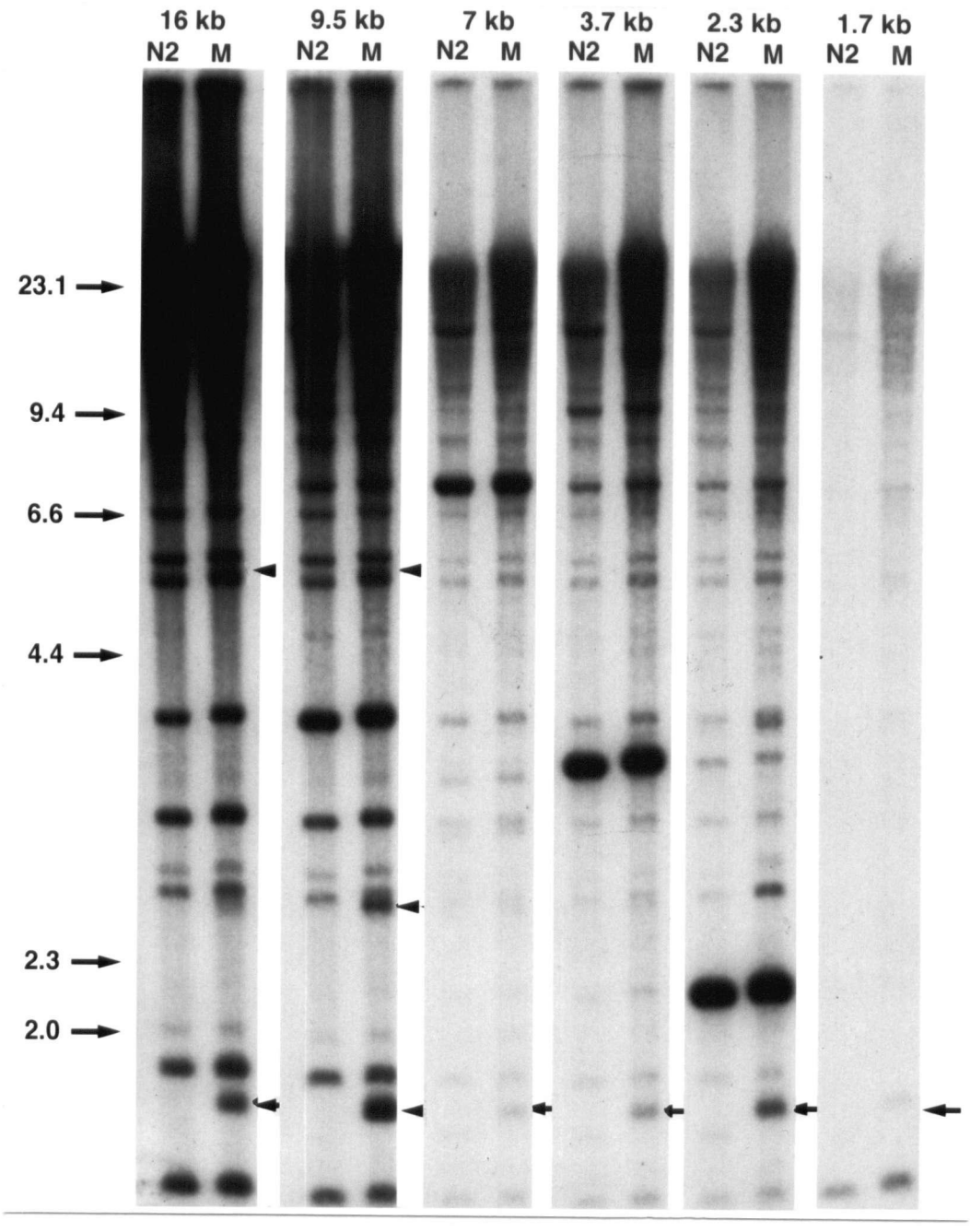


Figure 19: Identification of the unc-112(r367ra202) RFLP with Kpn I fragments of cosmid T02B5.

Genomic DNA from N2 and *unc-112(r367ra202)* were digested with Kpn I and probed with individual Kpn I fragments from cosmid T02B5. The molecular weight markers are given down the left side and the probing fragment is given above each pair of lanes. The N2 DNA is in the left lane and *unc-112(r367ra202)* DNA is in the right lane (M). The 7 kb, 3.7 kb, 2.3 kb and 1.7 kb probes identify bands of the same size in both (N2) and (m) lanes. The 16 kb and 9.5 kb probes identify multiple bands that do not add up to the probe size. The (M) lanes identify a polymorphic band of ~1.5 kb to varying intensities with all probes, the strongest in the 9.5 kb-probed (M) lane (arrows). Another polymorphic band is seen at 6 kb in both the 16 kb- and 9.5 kb-probed (M) lanes (arrowheads). A unique polymorphic band of ~2.8 kb is identified only in the 9.5 kb-probed (M) lane (double arrowhead).

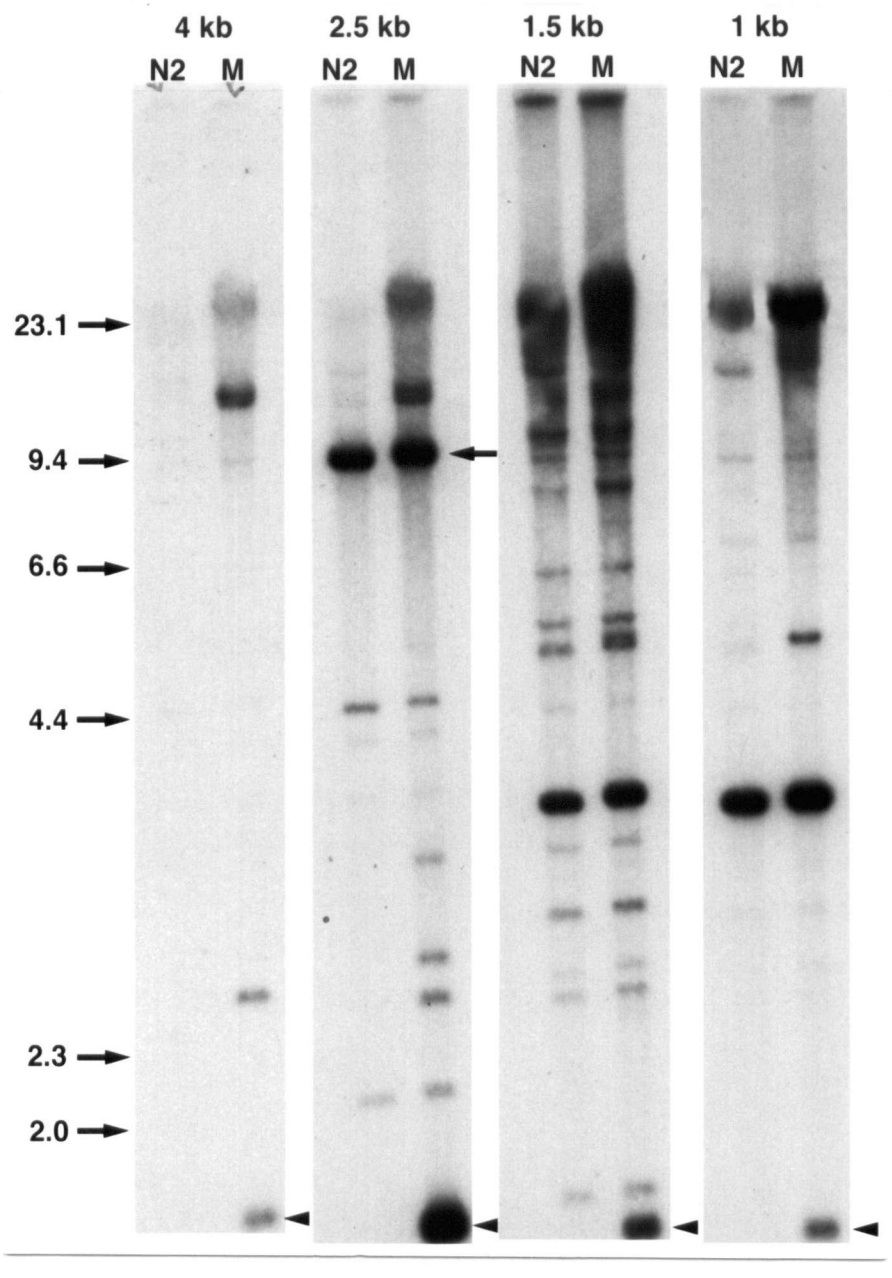


Figure 20: Localization of the unc-112(r367ra202) RFLP to a 2.5 kb Kpn I-EcoRI fragment of T02B5.

Genomic DNA from N2 and *unc-112(r367ra202)* were digested with Kpn I and probed with individual Kpn I-EcoRI fragments from the 9.5 kb Kpn I fragment of T02B5. The N2 DNA is in the left lane and *unc-112(r367ra202)* DNA is in the right lane (M). The molecular weight markers are given down the left side and the probing fragment is given above each pair of lanes. The 4 kb subfragment hybridized poorly to the genomic DNAs but did identify the 1.7 kb polymorphic band (arrowhead). The 1.5 kb and 1 kb subfragments hybridized only weakly to a 9.5 kb band and strongly to a 4 kb band in both N2 and M lanes. These two probes also hybridized to the 1.7 kb polymorphic band but with less intensity than to the 4 kb band. Only the 2.5 kb subfragment hybridized to a band the same size as the 9.5 kb Kpn I fragment from which it was derived (arrow) in both N2 and M lanes, and to a 1.7 kb polymorphic band in the M lane with equal intensity.

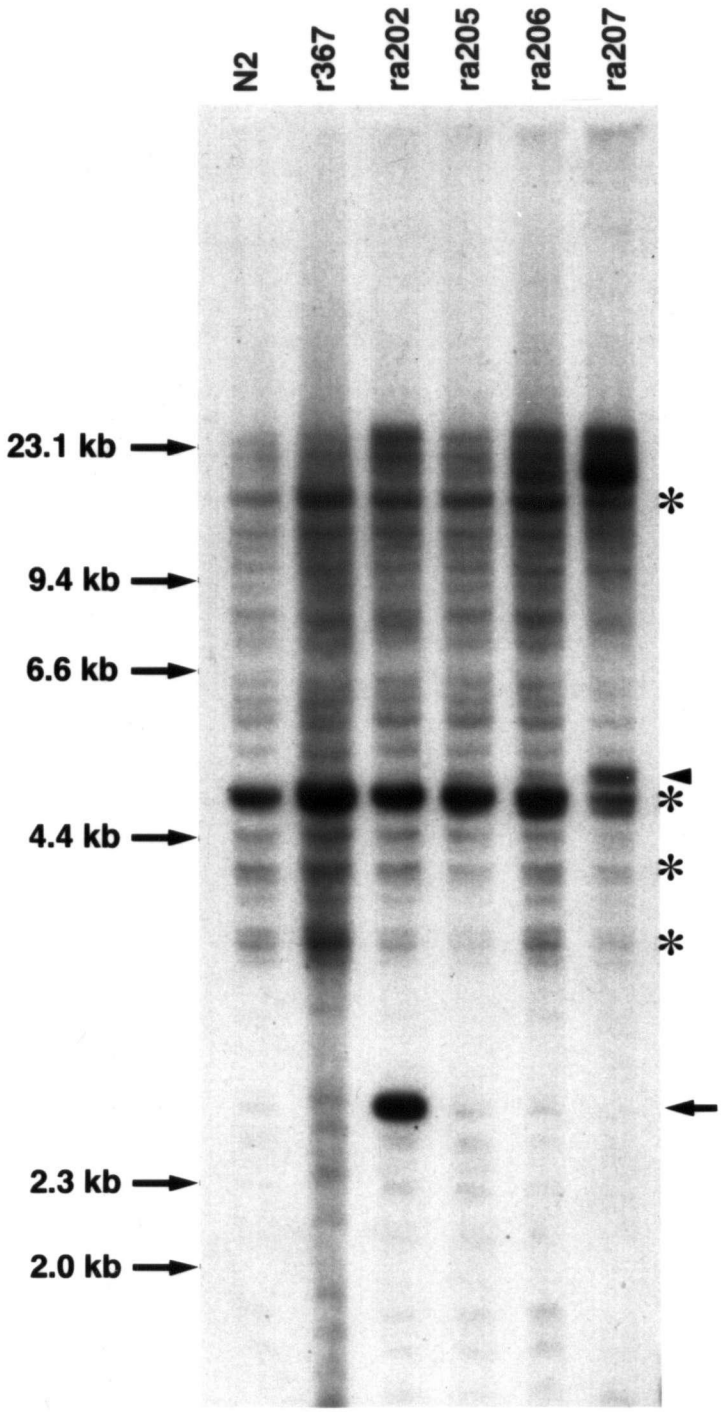


Figure 21: Confirmation of RFLPs associated with intragenic revertants.

Genomic DNA from N2, four intragenic and one intergenic revertant was digested with HindIII and probed with the 2.5 kb Kpn I-EcoRI fragment from cosmid T02B5. The positions of the molecular weight markers are given down the left side and the genotype of the DNA in each lane is given above. Both wildtype and mutant DNA digests show a prominent band of ~5 kb and crosshybridizing bands of 13 kb, 4 kb and 3.5 kb (*). In addition, *unc-112(r367ra202)* has a second band of ~2.8 kb (arrow) and *unc-112(r367ra207)* has a second band of ~5.3 kb (arrowhead).

sizes than the cosmid probe and the polymorphic bands seen with Kpn I-digested DNA. HindIII-digested *unc-112(r367ra202)* DNA generates an approximately 5 kb wild type band and a 2.8 kb mutant band, whereas the same DNA digested with Kpn I generates a 9.5 kb wild type band and a 1.5 kb mutant band (compare figure 21 with figure 20). Similarly, HindIII-digested *unc-112(r367ra207)* DNA generates an approximately 5 kb wild type band and a 5.5 kb mutant band, whereas the same DNA digested with Kpn I generates a 9.5 kb wild type band and a 4 kb mutant band (compare figure 18 with figure 24). These results eliminate both the possibility of probe DNA contaminating the genomic DNA samples and the likelihood of misidentifying a "background" band containing multiple repeats from another part of the genome as a polymorphism within the T02B5 region.

Characterization of the 2.5 kb cosmid fragment.

Prior to the isolation of possible *unc-112* clones, the 2.5 kb cosmid fragment was hybridized to a grid of Yeast Artificial Chromosomes (YACs) representing ~99% of the *C. elegans* genome. The cosmid probe only hybridized to the YAC, Y39H1, overlapping the region of chromosome V containing cosmid T02B5. This result suggested that the probe contained DNA that was not highly related to DNA located elsewhere in the *C. elegans* genome. The cosmid probe was also hybridized to a Northern blot of staged and mixed populations of N2 total RNA (see figure 22). A single mRNA of approximately 3 kb was revealed and appeared in greatest abundance in L3-L4 and mixed population RNAs. Therefore, the fragment had all the properties needed in order to serve as the probe to isolate candidate *unc-112* clones from a cDNA library: it was small and unrearranged; it identified RFLPs unique to two intragenic revertant strains *unc-112(r367ra202)* and *unc-112(r367ra207)*; it contained unique sequence that identifies only YACs from this region of chromosome V; and it hybridized to a single mRNA.

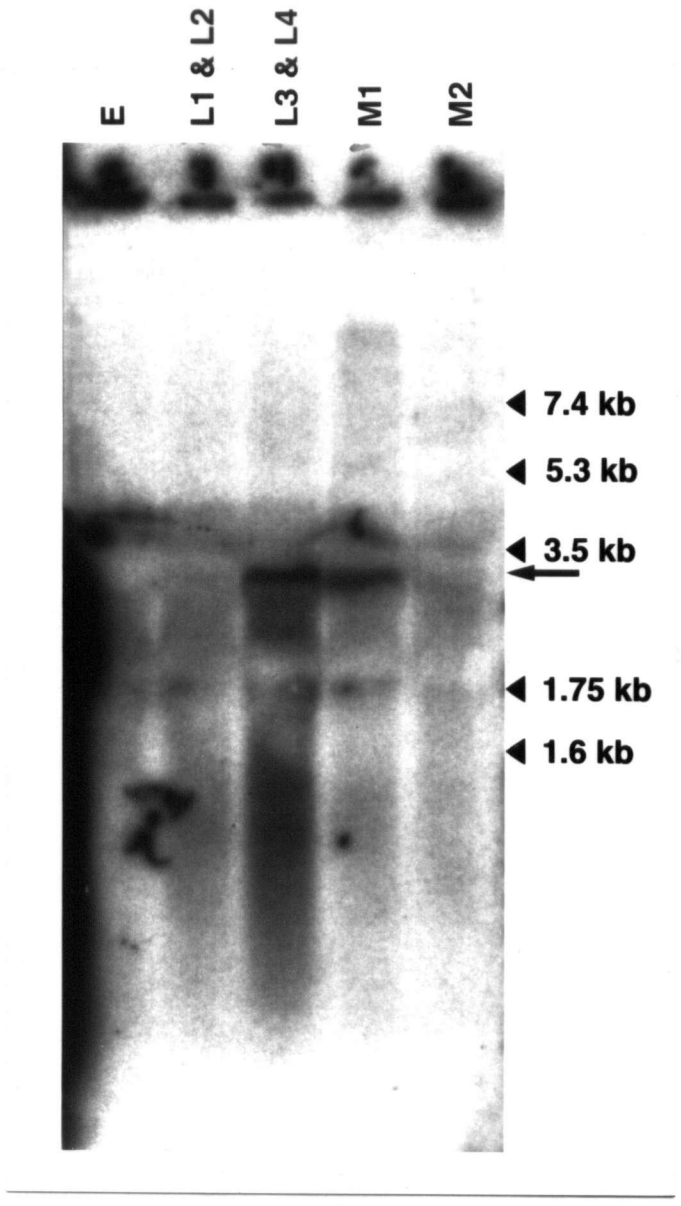


Figure 22: N2 developmental Northern profile of the 2.5 kb T02B5 fragment.

Total RNA from N2 embryos (E), larval stages I and II (L1 & L2), larval stages III and IV (L3 & L4) and two mixed populations (M1 and M2) were probed with the 2.5 kb Kpn I-Eco RI fragment from cosmid T02B5. The molecular weight markers are indicated down the right side. A single band sized 3 kb-3.5 kb (arrow) was identified in all lanes but is most abundant in L3 & L4.

C. Isolation and sequence determination of cDNAs.

Cloning the cDNAs.

The first cDNA library screened with the cosmid fragment consisted of unsized cDNA in the vector λZAP 1 (obtained from R. Barstead). Approximately 2.5 x 10^5 clones were screened with the 2.5 kb Kpn I-EcoRI T02B5 fragment to isolate six independent cDNAs (called p2A2, p2B1, p3A1, p5A2, p10A2, and p10C2).

The same cosmid fragment that identified RFLPs in two intragenic *unc-112* revertants also hybridized strongly to a 2 kb cDNA. This result indicated that the isolated cDNA may be derived from the gene identified by the RFLPs and Northern hybridization. However, it was possible that the cosmid fragment carried the coding sequence from an adjacent gene and some intervening sequences between the two genes (see figure 23). If the cosmid fragment only hybridized to cDNAs originating from the adjacent gene, then the cDNA would not be expected to recognize an RFLP located within the *unc-112* coding sequence. It was also possible that intervening noncoding sequences on the cosmid fragment were responsible for recognizing the RFLPs and that the cDNA would not recognize intra-*unc-112* RFLPs. These possibilities were tested using a pair of identical genomic Southern blots containing N2, *unc-112(r367); dim-1(ra102), unc-112(r367ra202), unc-112(r367ra205), unc-112(r367ra206),* and *unc-112(r367ra207).* One blot was probed with the 2 kb cDNA, and the second blot was probed with the 2.5 kb cosmid fragment. The cDNA recognized the same RFLPs as the cosmid fragment (see figure 24). Therefore, it is unlikely that the RFLPs are in either non-coding intervening sequence or a gene adjacent to the RFLP-containing gene.

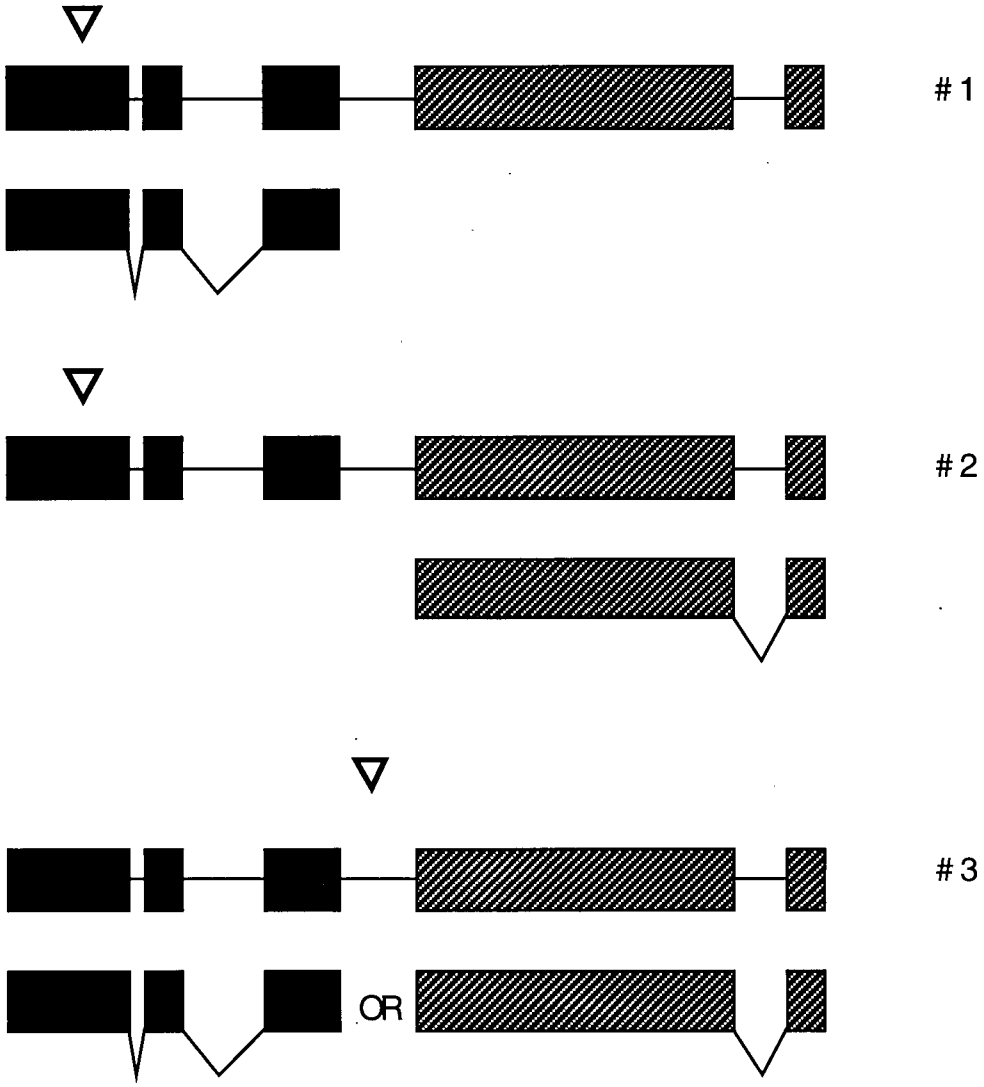


Figure 23: Model for the correlation of RFLPs with the cDNAs.

There are three possible models that relate the cDNAs to the genomic DNA (rectangles connected by a straight line) and the polymorphisms associated with the two intragenic revertants *unc-112(r367ra202)* and *unc-112(r367ra207)*. The first and second models are based on the assumption that the polymorphism (triangle) is located within a gene. In model **#1** the cDNA (rectangles joined by jagged lines) is derived from the same gene that contains the polymorphism and predicts that the cDNA will hybridize to the polymorphic genomic DNA of the intragenic revertants. In model **#2** the cDNA is derived from an adjacent gene and predicts that the cDNA will not hybridize to the polymorphic genomic DNA of the intragenic revertants. Model **#3** is based on the assumption that the polymorphism is located in a long stretch of noncoding sequence between two genes and predicts that cDNAs from either gene will not hybridize to the polymorphic genomic DNA of the intragenic revertants.

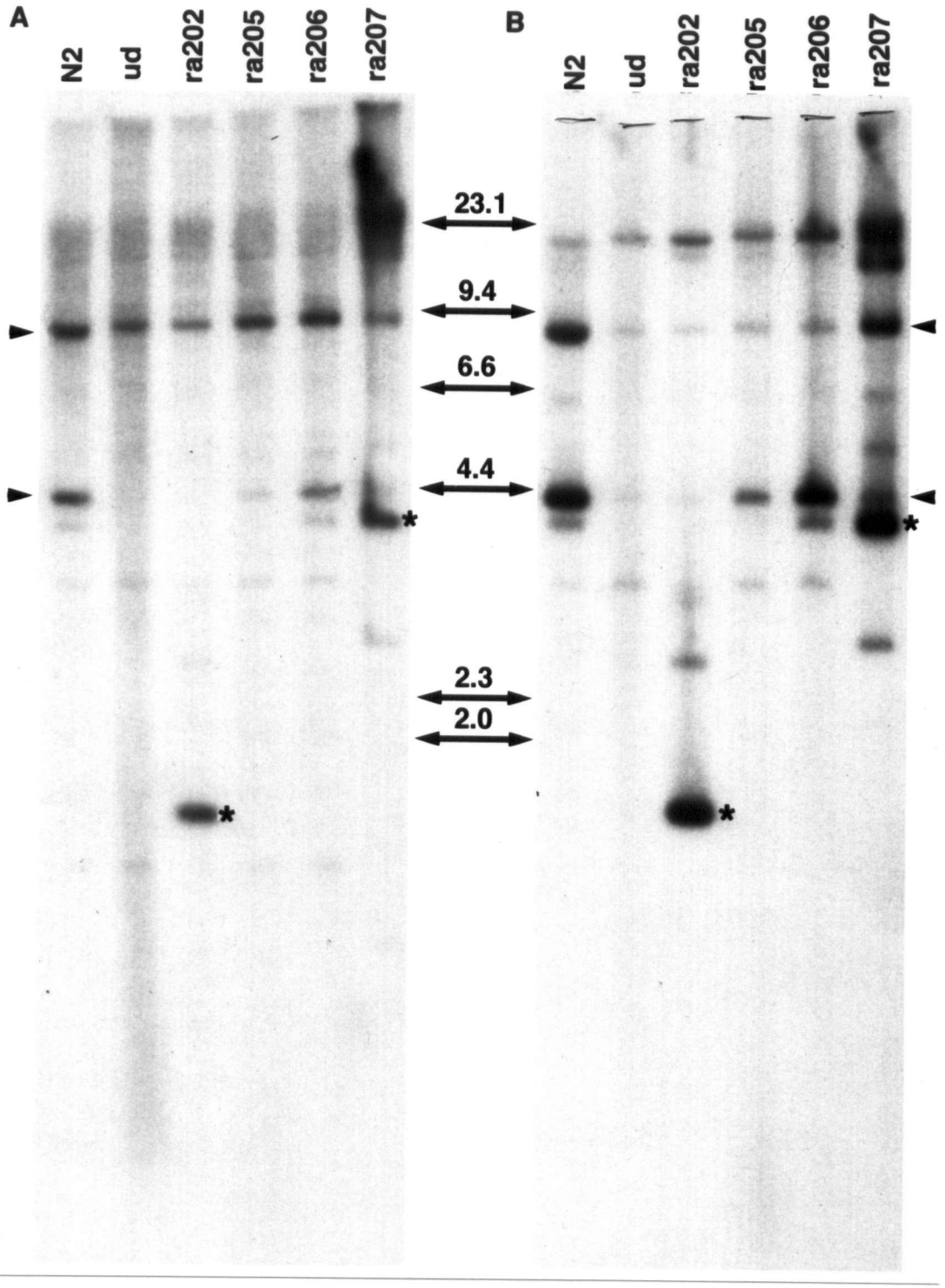


Figure 24: Comparison of the cDNA and cosmid fragment genomic patterns.

Genomic DNA from *yDf11/nT1(m435)(IV:V)*, which is wildtype for *unc-112, unc-112(r367);dim-1(ra102)*, and each of the four intragenic revertants were digested with Kpn I. They are labeled as N2, ud, ra202, ra205, ra206, and ra207 above each lane respectively. Panel **A**) was probed with the 2.5 kb Kpn I-Eco RI T02B5 fragment. Panel **B**) was probed with p2A2. Polymorphic bands seen in both panels are marked (*), wildtype bands in both panels are marked with arrowheads.

D. Attempted rescue of *unc-112(r367)* with ZC376 DNA

In order to determine whether ZC376.2 (whose exons are identical to pUNC112-1), ZC376.1 or the gene embedded within ZC376.1 were likely to encode the *unc-112* gene product, a *C. elegans* strain transformed with an extrachromosomal array consisting of ZC376 and *rol-6(su1006)*, a dominant phenotypic marker, was outcrossed to *unc-112(r367)/+* males. Frequencies of 52% Wild type, 30% Roller, and 18% Paralyzed phenotypes were obtained from the 2493 progeny of 10 Roller hermaphrodites. It was possible that *unc-112(r367);ZC376rol-6(1006)* animals had been missed when picking Roller hermaphrodites because too few Roller animals were picked, therefore 34 Roller progeny were picked and their genotypes determined by the phenotypes of the progeny. Sixteen plates consisted of Wild type, Roller and Paralyzed animals, 10 plates consisted of Wild type and Roller animals; 8 plates consisted of sterile hermaphrodites (a property of the parent strain carrying the extrachromosomal array); no plates consisted of Paralyzed and Roller animals. The absence of Paralyzed progeny in both tests suggests that the cosmid ZC376 is unable to compensate for the defects resulting from the *unc-112(r367)* mutation.

4. Discussion

A. Molecular mapping of the genomic region and identification of *unc-112* associated polymorphisms.

Proximal and intragenic molecular tags were sought to serve as reference points for correlating the genetic and physical maps, and as aids in obtaining both cDNAs and genomic clones of *unc-112*. The deficiencies with which *unc-112(r367)* was mapped genetically provided molecular tags for correlation with the physical map. The left half of the genomic region between TCPAR1 and *unc-76* was shown to hybridize to many bands that totaled more than the estimated size of the cosmid probe. This phenomenon is characteristic of repetitive DNA sequences and made definitive deficiency breakpoint RFLP identification difficult. The two mechanisms by which repetitive DNA interferes with interpretation of

hybridization patterns are: 1) by potentially masking polymorphic bands due to the high number of hybridizing bands, or 2) by confounding deficiency-generated RFLP results with bands generated by spontaneous genomic rearrangements between repetitive sequences. It is likely that the polymorphic band corresponding to *yDf8* was not discovered either due to masking of the band by the great number of bands containing repetitive sequences or that it was in the region to the left of T02B5 and T20C8 not covered by cosmids. The RFLP identified in *yDf11* with cosmid R15H1 and the RFLP identified in *yDf9* with cosmid ZC256 are associated with the deficiencies. Recently, PCR-based confirmation that R15H1 contains the left breakpoint of *yDf11* has been obtained (T. Rogalski and D.G. Moerman, pers. comm.), but the breakpoint of *yDf9* has yet to be confirmed. In addition, the egg-laying defect of *egl-1* mutants has been rescued using a cosmid that lies to the right of the *yDf9*-associated polymorphism (B. Conradt, per. comm.). This result conflicts with genetic mapping data that shows *egl-1* is to the left of *unc-112* and deficiency complementation data indicating that *egl-1* is outside the region deleted by *yDf9,* and *yDf11* (M. Hengartner, per. comm.), whereas *unc-112* is complemented by *yDf9* and *yDf11* but not by *yDf8*. With the *C. elegans* genome sequencing groups working in this region of chromosome V and the advent of PCR, the sites of the RFLPs now can be analyzed using oligonucleotides generated to DNA flanking both sides of the deficiency to amplify a bridging fragment in single, homozygous deficiency embryos. If the RFLP sites do not correspond to the deficiencies, they are probably polymorphisms that arose by independent and spontaneous rearrangement between nearby repetitive sequences.

Genetic complementation tests show that the deficiencies *yDf8, yDf9* and *yDf11* all complemented *par-1*, which indicates that a Tc1 polymorphism mapping one genetic map unit to the left of *par-1* and within cosmid EEEH7 was used as a left-most physical marker of the *unc-112* region (A. Telfer, per. comm.). Since the polymorphisms identified in *yDf9* and *yDf11* with cosmids ZC256 and R15H1 are the only consistent polymorphisms found between ZC256 and C08C1 (which rescues *unc-76*), the *yDf9*-associated polymorphism was used as a right-most physical marker of *unc-112*. These results suggest that the region likely to contain *unc-112* consists of 8 cosmids and two gaps between EEEH7 and ZC256.

In order to identify intragenic molecular tags to be used in cloning *unc-112*, four of the intragenic revertants were probed with the cosmids from the EEEH7 and ZC256 region. Unique polymorphisms were uncovered in revertants *unc-112(r367ra202)* and *unc-112(r367ra207)* with cosmid probes T02B5

and T20C8. These two cosmids are from a small "island" of clonable genomic DNA flanked by gaps. This is another indication of repetitive sequences and suggests that the polymorphisms in *unc-112(r367ra202)* and *unc-112(r367ra207)* may be the result of reversion of the *r367* mutation or a coincidental and unrelated rearrangement between repetitive sequences. If these polymorphisms are unrelated to reversion of *unc-112(r367)*, they are closely linked to *unc-112* and therefore can be used to further refine the region containing the gene. Since no wildtype bands are lost due to an increase or decrease in size, the polymorphisms are probably the result of duplication by an unequal crossover event (see Appendix B). If these polymorphisms are related to reversion of *unc-112(r367)*, they suggest that suppression of paralysis may be due to recruitment of a gene family member near *unc-112*.

B. Isolation and sequence determination of cDNAs.

The polymorphisms identified in DNA from *unc-112(r367ra202)* and *unc-112(r367ra207)* were not found in either wild type or *unc-112(r367)* genomic DNAs. Therefore they appear to be unique to these two suppressor strains and might be located within the *unc-112* sequence. The smallest fragment of T02B5 that identified the two putative intragenic RFLPs (a 2.5 kb EcoRI fragment) was used to isolate 5 cDNAs, including pUNC112-1, and to identify a single 3 kb RNA. Since the mRNA is estimated to be 3 kb and the insert of pUNC112-1 is about 2 kb and possibly short of full length by 1 kb, two more cDNA libraries were screened and five more clones were isolated. The sequences of two of the new cDNAs terminated within 15 nt of the 5' end of pUNC112-1, demonstrating that pUNC112-1 is a full length clone. Furthermore, pUNC112-1 was shown to exhibit the same genomic hybridization pattern as the 2.5 kb cosmid fragment. This result indicates that pUNC112-1 is derived from the same region as the genomic fragment and that the polymorphisms are associated with one or more transcribed genes and not associated with a region devoid of function.

C. Genetic rescue of *unc-112(r367)* attempted using ZC376 DNA

A search for polymorphisms in deficiency-bearing strains yielded a polymorphism in cosmid R15H1 associated with *yDf11* and another polymorphism in cosmid ZC256 associated with *yDf9*. The genetic complementation data indicate that both deficiencies are located to the right of *unc-112*. Therefore, the region between ZC256 and TCPAR1 was searched for polymorphisms associated with intragenic revertants of *unc-112*. The polymorphisms identified in the strains containing *unc-112(r367ra202)* and *unc-112(r367ra207)* were located in a small cosmid island within the 8 cosmid region defined by *yDf9* and TCPAR1. Furthermore, a cDNA was isolated that was derived from a clustered five gene family in the cosmid island (see Appendix A). The site of unequal crossover generating the polymorphism in the strain containing *unc-112(r367ra202)* was shown to be the result of fusion between two of the five genes by PCR amplification (see Appendix B). All of these results together suggest, but do not prove, that *unc-112* is encoded by one member of the five gene family. However, the evidence suggesting that *unc-112* is not in this gene cluster is that the cosmid that rescues *egl-1* lies to the right of this region (B. Conradt, per. comm.). This result would appear to conflict with genetic mapping which places *egl-1* to the left of *unc-112* and *yDf9* (M. Hengartner, per. comm.). Considering the amount of repetitive DNA in the *unc-112* region, it appears more likely that the breakpoint of *yDf9* does not correspond to the polymorphism in ZC256, and the clustered gene family may lie to the left of *egl-1* and *unc-112*.

To test the possibility that the RFLPs identified sequences that were unrelated to *unc-112,* two direct strategies were attempted: 1) microinjection of cosmid T02B5 into *dpy-11(e224)unc-112(r367)* animals raised at 15 °C then shifted to 20 °C and the progeny examined for suppression of the paralyzed phenotype, and 2) male-mediated transfer of *unc-112(r367)* into a ZC376/*rol-6(e1006)*-bearing strain. The first strategy yielded very few Dpy-Rol transformants and only one hermaphrodite that may have been rescued but was sterile. This experiment was attempted before the structure of the genome in the region had been determined. Since T02B5 contains only a 3' portion of ZC376.3 and all of T02B5.A and T02B5.B it is possible that essential regulatory sequences in ZC376 were needed or that rescue could only be effected by ZC376.1,2 or 3. However, ZC376 also was shown not to rescue *unc-112(r367)* from paralysis since no Roller hermaphrodite tested segregated only Rollers and Paralyzed progeny. There are three possible explanations for these results: 1) that an intact ZC376.3 gene is required for rescue; 2)

that either T02B5.A or T02B5.B can rescue if the regulatory regions are intact; or 3) that none of the five tandem esterase-like genes are capable of rescue and instead are close neighbors of *unc-112*. If either of the first two possibilities are correct, then rescue of *unc-112(r367)* from paralysis could be accomplished by introducing *unc-112(r367)* into a strain carrying an extrachromosomal array consisting of T02B5 and ZC376 plus the Roller marker (as described in section 2G of this chapter). If the third possibility is correct, then the transformation rescue experiment with T02B5 and ZC376 will fail to rescue *unc-112(r367)* from paralysis.

Recently, transformation rescue of *unc-112(r367)* from paralysis has been accomplished (T. Rogalski and D. Moerman, per. comm.). The rescuing cosmid, T10E3, lies to the left of the cosmid overlapping the yDf11 breakpoint and to the right of the *egl-1*-containing cosmid. Consequently, the polymorphisms identified in *unc-112(r367ra202)* and *unc-112(r367ra207)* and the corresponding cDNA are closely linked, but otherwise unrelated, to *unc-112*. However, the polymorphisms and cDNA identify a unique and ancient clustered gene family belonging to the cholinesterase subfamily of serine esterases (see Appendix A).

Chapter 4

Summary and Conclusions

A major aim of this thesis was to learn more about the essential principles of bodywall muscle differentiation in the nematode *Caenorhabditis elegans*. Along these lines I have characterized the gonadal and muscle phenotypes of the viable allele of *unc-112*, refined the map position of the gene genetically and molecularly, and determined the class of mutation. In addition, I have isolated cDNAs associated with polymorphisms in two intragenic revertants of *unc-112* that identify a unique and ancient clustered gene family belonging to the cholinesterase subfamily of serine esterases closely linked and to the left of *unc-112*. Furthermore, I have described a new gene, *dim-1*, which indirectly interacts with *unc-112*; characterized the interaction of *dim-1* with *unc-112*; genetically mapped *dim-1* and determined the class of mutation of the *dim-1(ra102)* allele. From these studies, I have developed a model that involves the *unc-112* and *dim-1* gene products in roles that signal and initiate the processes of assembly and organization of sarcomeres in muscle cells.

The gene *unc-112* is involved in an early and essential step in embryonic muscle differentiation (123) and is also required to maintain the integrity of the sarcomere array from larval stage 3 (L3) through adulthood (91,116). I have examined the effects of *unc-112* viable mutations on gonad and non-bodywall muscle morphology and shown that *unc-112(r367)* hermaphrodites are small and thin as well as defective in egg laying. These phenotypes as well as the slower growth rate are probably due to insufficient numbers of sarcomeres in the uterine, vulva and pharyngeal muscles. These results indicate that *unc-112* is expressed in all types of muscle in *C. elegans* and places this gene in a small and unique group of muscle-specific genes.

The organization of both sarcomeres and individual components in bodywall muscles of *unc-112(r367)* were examined. The highly birefringent A-bands form fibrous clumps in contrast to the paired rows of A-bands in N2 animals. Similarly, clumps of vinculin- and β-integrin-specific antibody staining

spots, and patches of perlecan-specific antibody staining are seen in the bodywall muscles of *unc-112(r367)* animals. These results are consistent with the idea that *unc-112* is involved in regulating the placement of the sarcomeres. This notion is further supported by the fibrous aggregate staining patterns of actin and myosin antibodies (123) and of randomized placement of β-integrin-specific staining spots (G.P. Mullen and D.G. Moerman, unpub. results) observed in bodywall muscles of *unc-112(st562)* embryos.

In order to understand the effects that mutant *unc-112* alleles have on muscle development it was important to identify other genes with which *unc-112* interacts. Genetic reversion of *unc-112(r367)*-induced paralysis yielded 9 intragenic suppressors and 49 intergenic suppressors. Seventeen of the intergenic suppressors were shown to be alleles of the new gene, *dim-1*. *dim-1* was mapped to a small and well characterized region of the X chromosome between *mec-7* and *xol-1*. Deficiency mapping also indicated that *ra102* is a null mutation and demonstrates that *dim-1* is not essential for viability. *dim-1* is the first member of a novel class of muscle-affecting genes that perturbs sarcomere organization without affecting movement, morphology or fecundity of the animal. The loss-of-function allele *dim-1(ra102)* has no adverse effects on single-sarcomere muscles, nor does it appear to affect bodywall muscle function. Alleles of *dim-1* confer a novel phenotype of easily disrupted birefringence in all bodywall muscle cells and a Chevron pattern in some cells. Thus, *dim-1* identifies a new class of muscle-affecting mutations which alter bodywall muscle organization but not movement. *dim-1* may be involved in placement of dense bodies and M-lines in muscle cell membranes either by directing the placement of sarcomeres, or functioning in response to a placement signaling factor. The loss of either type of factor would not prevent sarcomeres from forming or being inserted into the muscle cell membrane but would likely lead to an aberrant configuration of sarcomeres.

The assembly and placement of the Z-line and M-line complexes is crucial for determining both the orientation and the pattern of sarcomeres in vertebrates and nematodes. During the larval phase of bodywall muscle development the first row of sarcomeres may increase in length and width by addition of new dense bodies and M-lines to the midpoint of the rows and the formation of new rows. As the distance between the new attachment sites can accommodate them, thin and thick filaments are added. The orientation and spacing of the sarcomere attachments within rows and between rows may be established

by a signaling pathway originating in the overlying basement membrane and affecting components within muscle cells. The intramuscular response to the extramuscular signal would anchor newly formed dense bodies and M-lines in position in the muscle cell membrane via attachments to the cytoskeleton and the basement membrane. I have developed a model describing how this might occur based on analogy to the the localization of pre- and post-synaptic proteins of the NMJ (see figure 14) (63). Three significant similarities have been identified in the structure of the NMJ and bodywall muscle-hypodermis association which enable the analogy to be made: 1) a specialized basement membrane, 2) clusters of signal or force transduction proteins, and 3) components of the cytoskeleton and basement membrane anchor the clusters of signal or force transduction proteins. If *dim-1* functions in a role analogous to MASK or agrin, then the effect of loss of function of *dim-1* would be strongest early in the larval phase of muscle differentiation since the placement of the first new row on either side of the midline would be crucial for aligning the rest of the rows on that side of the midline. The absence of *dim-1* function may permit the first rows to align randomly from the midline instead of parallel to it. DIM1 may interact with a transmembrane protein capable of intracellular signaling, such as integrin (5) or another protein, in a MuSK-like role (perhaps *unc-112*) which is associated with dense bodies and M-line complexes but not an integral part of them.

The genetic analysis of *unc-112(r367)* provided additional information regarding its chromosomal position. The position of *unc-112* on the genetic map was restricted to the area deleted by *yDf8* which is right of *par-1* and *egl-1* and left of both *yDf9* and *yDf11*. The deficiency breakpoints were sought as proximal molecular tags to correlate the genetic and physical maps of the region and aid in cloning the *unc-112* gene. Polymorphisms were identified in *yDf11* with cosmid R15H1 and in *yDf9* with cosmid ZC256. These are the only consistent polymorphisms found between ZC256 and *unc-76*. Therefore the *yDf9*-associated polymorphism in ZC256 places *unc-112* in the region between EEEH7 and ZC256.

Unique polymorphisms were uncovered in revertants *unc-112(r367ra202)* and *unc-112(r367ra207)* with cosmids T02B5 and T20C8 which lie between EEEH7 and ZC256. These two cosmids correspond to a small "island" of genomic DNA flanked by gaps of unknown size and are an indication of repetitive DNA. Consequently, the polymorphisms in *unc-112(r367ra202)* and *unc-112(r367ra207)* may be due to a coincidental and unrelated rearrangement between repetitive sequences or reversion of the *r367*

mutation. To test these possibilities, attempts were made to rescue the paralyzed phenotype of *unc-112(r367)* using extrachromosomal arrays of either ZC376 or T02B5. Successful rescue would offer direct proof that this region contains the *unc-112* gene. However, T02B5 arrays yielded only one hermaphrodite that may have been rescued but was sterile and ZC376 also did not to rescue *unc-112(r367)* paralysis. In view of the discrepancy between the cosmid rescue of *egl-1* (B. Conradt, per. comm.) and the multi-factor and deficiency mapping results with *egl-1* and *unc-112*, cosmids between *egl-1* and *yDf11* were tested for ability to rescue *unc-112(r367)* paralysis (T. Rogalski and D. Moerman, unpub. results). Successful rescue of paralysis was obtained with cosmid T10E3 confirming that the polymorphisms in the revertant strains are linked, but otherwise unrelated, to *unc-112.*

In this dissertation I have attempted to address the identity and behavior of two muscle-affecting genes in *C. elegans*, *unc-112* and *dim-1*, and to integrate these results into the larger context of how muscle proteins organize into a structure capable of movement, feeding, processing oocytes and laying eggs. Although the identities of the genes have not been directly proven, both *unc-112* and *dim-1* are likely to be signaling molecules or receptors and not structural components of sarcomeres. In addition, results from analysis of a candidate cDNA and a five gene cluster associated with polymorphisms in two *unc-112* intragenic revertants encode cholinesterase-like proteins one or more of which may also have a role in sarcomere organization or function.

Bibliography

1. **Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson**. 1983. Molecular Biology of the Cell. Garland Publishing, Inc. New York.

2. **Albertson, D. and J. N. Thomson**. 1976. The pharynx of *Caenorhabditis elegans*. Philos. Trans. R. Soc. London **275**:299-325.

3. **Alfonso, A., K. Grundahl, J. R. McManus, J. M. Asbury, and J. B. Rand**. 1994. Alternative splicing leads to two cholinergic proteins in *Caenorhabditis elegans*. J. Mol. Biol. **241**:627-630.

4. **Anderson, P.** 1989. Molecular genetics of nematode muscle. Ann. Rev. Genetics **23**:507-525.

5. **Arcangeli, A., A. Becchetti, A. Mannini, G. Mugnai, P. DeFilippi, G. Tarone, M. R. DelBene, E. Barletta, E. Wanke, and M. Olivotto**. 1993. Integrin-mediated neurite outgrowth in neuroblastoma cells depends on the activation of potassium channels. J. of Cell Biology **122**:1131-1142.

6. **Auld, V. J., R. D. Fetter, K. Broadie, and C. S. Goodman**. 1995. Gliotactin, a novel transmembrane protein on peripheral glia, is required to form the blood-nerve barrier in Drosophila. Cell **81**:757-767.

7. **Baldwin, T. J. and S. J. Burden**. 1989. Muscle-specific gene expression controlled by a regulatory element lacking a MyoD1-binding site. Nature **341**:716-720.

8. **Barstead, R. J. and R. H. Waterston**. 1991. Vinculin is Essential for Muscle Function in the Nematode. The Journal of Cell Biology **114**:715-724.

9. **Barthalay, Y., R. Hipeau-Jacquotte, S. de la Escalera, F. Jimenez, and M. Piovant**. 1990. *Drosophila* neurotactin mediates heterophilic cell adhesion. The EMBO Journal **9**:3603-3609.

10. **Bejsovec, A. and P. Anderson**. 1988. Myosin heavy-chain mutations that disrupt *Caenorhabditis elegans* thick filament assembly. Genes & Development **2**:1307-1317.

11. **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 **342**:45-50.

12. **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 *Caenorhabditis elegans*. Genetics **134**:1097-1104.

13. **Benian, G.M., T.L. Tinley, X. Tang, and M. Borodovsky**. 1996. The *Caenorhabditis elegans* gene *unc-89*, required for muscle M-line assembly, encodes a giant modular protein composed of Ig and signal transduction domains. J. of Cell Biology **132**:835-848.

14. **Bolten, S. L., P. Powell-Abel, D. A. Fischhoff, and R. H. Waterston**. 1984. The *sup-7(st5) X* gene of *C. elegans* encodes a transfer RNA-Trp-UAG amber suppressor. Proc. Natl. Acad. Sci. , USA **81**:6784-6788.

15. **Branden, C. and J. Tooze**. 1991. Prediction, engineering, and design of protein structures, p. 247-268. In AnonymousIntroduction to Protein Structure. Garland Publishing, Inc. New York.

16. **Braun, T., G. Buschhausen-Denker, E. Bober, E. Tannich, and H. H. Arnold**. 1989. A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. The EMBO Journal **8**:701-709.

17. **Brenner, S.** 1974. The Genetics of *Caenorhabditis elegans*. Genetics **77**:71-94.

18. **Campanelli, J. T., G. G. Gayer, and R. H. Scheller**. 1996. Alternative RNA splicing that determines agrin activity regulates binding to heparin and α-dystroglycan. Development **122**:1663-1672.

19. **Chen, L., M. Krause, B. Draper, H. Weintraub, and A. Fire**. 1992. Zygotic expression of the *C. elegans* MyoD homologue *hlh-1* is not essential for body wall myogenesis. Science **256**:240-243.

20. **Chen, L., M. Krause, M. A. Sepanski, and A. Fire**. 1994. The *Caenorhabditis elegans* MyoD homologue *hlh-1* is essential for proper muscle function and complete morphogenesis. Development **120**:1631-1641.

21. **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.

22. **Church, G. M. and W. Gilbert**. 1984. Genomic Sequencing. Proc. Natl. Acad. Sci. , USA **81**:1991-1995.

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

24. **Coulson, A., R. H. Waterston, J. Kiff, J. Sulston, and Y. Kohara**. 1988. Genome linking with yeast artificial chromosomes. Nature **335**:184-186.

25. **Cousin, X., T. Hotelier, P. Lievin, J. Toutant, and A. Chatonnet**. 1996. A cholinesterase genes server (ESTER): a database of cholinesterase-related sequences for multiple alignments, phylogenetic relationships, mutations and structural data retrieval. NAR 132-136.

26. **de la Escalera, S., E.-O. Bockamp, F. Moya, M. Piovant, and F. Jimenez**. 1990. Characterization and gene cloning of neurotactin, a *Drosophila* transmembrane protein related to cholinesterases. The EMBO Journal **9**:3593-3601.

27. **Donoghue, M., H. Ernst, B. Wentworth, B. Nadal-Ginard, and N. Rosenthal**. 1988. A muscle-specific enhancer is located at the 3' end of the myosin light-chain 1/3 gene locus. Genes & Development **2**:1779-1790.

28. **Durfee, T., K. Becherer, P. Chen, S. Yeh, Y. Yang, A. E. Kilburn, W. Lee, and S. J. Elledge**. 1993. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Genes & Development **7**:555-569.

29. **Edmondson, D. G. and E. N. Olson**. 1989. A gene with homology to the *myc* similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes & Development **3**:628-640.

30. **Emmons, S. W.** 1988. The Genome, p. 47-79. In W. B. Wood (ed.), The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, USA.

31. **Ervasti, J. M., K. Ohlendieck, S. D. Kahl, M. G. Gaver, and K. P. Campbell**. 1990. Deficiency of a glycoprotein component of the dystrophin complex in dystrophic muscle. Nature **345**:315-319.

32. **Epstein, H.F. and Bernstein, S.I**. 1992. Genetic approaches to understanding muscle development. Developmental Biology **154**:231-244.

33. **Epstein, H.F. and Fischman, D.A**. 1991. Molecular analysis of protein assembly in muscle development. Science **251**:1039-1044.

34. **Fire, A.** 1986. Integrative transformation of *Caenorhabditis elegans*. EMBO J. **5**:2673-2680.

35. **Fire, A. and R. H. Waterston**. 1989. Proper expression of myosin genes in transgenic nematodes. EMBO J. **8**:3419-3428.

36. **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. Journal of Cell Biology **101**:1532-1549.

37. **Francis, R. and R. H. Waterston**. 1991. Muscle Cell Attachment in *Caenorhabditis elegans*. The Journal of Cell Biology **114**:465-479.

38. **Freiburg, A. and M. Gautel**. 1996. A molecular map of the interactions between titin and myosin-binding protein C. Implications for sarcomeric assembly in familial hypertrophic cardiomyopathy. European Journal of Biochemistry **235**:317-323.

39. **Furst, D. O., M. Osborn, R. Nave, and K. Weber**. 1988. The Organization of Titin Filaments in the Half-Sarcomere Revealed by Monoclonal Antibodies in Immunoelectron Microscopy: A Map of Ten Nonrepetitive Epitopes Starting at the Z Line Extends Close to the M Line. The Journal of Cell Biology **106**:1563-1572.

40. **Furst, D. O., M. Osborn, and K. Weber**. 1989. Myogenesis in the Mouse Embryo: Differential Onset of Expression of Myogenic Proteins and the Involvement of Titin in Myofibril Assembly. The Journal of Cell Biology **109**:517-527.

41. **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. of Cell Biology **129**:1127-1141.

42. **Glass, D. J., D. C. Bowen, T. N. Stitt, C. Radziejewski, J. Bruno, T. E. Ryan, D. R. Gies, S. Shah, K. Mattsson, S. J. Burden, P. S. DiStefano, D. M. Valenzuela, T. M. DeChiara, and G. D. Yancopoulos**. 1996. Agrin acts via a MuSK receptor complex. Cell **85**:513-523.

43. **Gossett, L. A., R. M. Hecht, and H. F. Epstein**. 1982. Muscle Differentiation in Normal and Cleavage-Arrested Mutant Embryos of *Caenorhabditis elegans*. Cell **30**:193-204.

44. **Greenwald, I. S. and H. R. Horvitz**. 1980. *unc-93(e1500)*: a behavioral mutant of *Caenorhabditis elegans* that defines a gene with a Wild-type null phenotype. Genetics **96**:147-164.

45. **Hall, Z. W. and J. R. Sanes**. 1993. Synaptic structure and development: the neuromuscular junction. Cell/Neuron **71/10**:99-121.

46. **Hanzlik, T. N., Y. A. I. Abdel-Aal, L. G. Harshman, and B. D. Hammock**. 1989. Isolation and Sequencing of cDNA Clones Coding for Juvenile Hormone Esterase from *Heliothis virescens*. The Journal of Biological Chemistry **264**:12419-12425.

47. **Hedgecock, E. M., J. G. Culotti, and D. H. Hall**. 1990. The *unc-5*, *unc-6*, and *unc-40* Genes Guide Circumferential Migrations of Poineer Axons and Mesodermal Cells on the Epidermis in *C. elegans*. Neuron **2**:61-85.

48. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351-359.

49. **Herman, R. K.** 1988. Genetics, p. 17-45. In W. R. Wood (ed.), The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

50. **Higgins, D. G., A. J. Bleasby, and R. Fuchs**. 1992. CLUSTAL V: improved software for multiple sequence alignment. CABIOS **8**:189-191.

51. **Higgins, D. G. and P. M. Sharp**. 1988. CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene **73**:237-244.

52. **Hodgkin, J., K. Kondo, and R. H. Waterston**. 1987. Suppression in the nematode *Caenorhabditis elegans*. TIGS **3**:325-329.

53. **Hodgkin, J., A. Papp, R. Pulak, V. Ambros, and P. Anderson**. 1989. A new kind of informational suppression in the nematode *Caenorhabditis elegans*. Genetics **123**:301-313.

54. **Hodgkin, J., R. H. A. Plasterk, and R. H. Waterston**. 1995. The nematode *Caenorhabditis elegans* and its genome. Science **270**:410-414.

55. **Horvitz, H. R., S. Brenner, J. Hodgkin, and R. K. Herman**. 1979. A uniform genetic nomenclature for the nematode *Caenorhabditis elegans*. Mol. Gen. Genet. **175**:129-133.

56. **Hresko, M. C., B. D. Williams, and R. H. Waterston**. 1994. Assembly of Body Wall Muscle and Muscle Cell Attachment Structures in *Caenorhabditis elegans*. The Journal of Cell Biology **124**:491-506.

57. **Hubbard, S. C. and R. J. Ivatt**. 1981. Synthesis and processing of asparagine-linked oligosaccharides. Annual Review of Biochemistry **50**:555-583.

58. **Jarvik, J. and D. Botstein**. 1973. A genetic method for determining the order of events in a biological pathway. Proc. Natl. Acad. Sci. , USA **70**:2046-2050.

59. **Kagawa, H. and K. Gengyo**. 1989. Paramyosin gene (*unc-15*) of *Caenorhabditis elegans* molecular cloning, nucleotide sequence and models for thick filament structure. Journal of Molecular Biology **207**:311-333.

60. **Kennedy, B. P., E. J. Aamodt, F. L. Allen, M. A. Chung, M. F. P. Heschl, and J. D. McGhee**. 1993. The gut esterase gene (*ges-1*) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Journal of Molecular Biology **229**:890-908.

61. **Kennelly, P. J. and E. G. Krebs**. 1991. Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. Journal of Biological Chemistry **266**:15555-15558.

62. **Kimble, J. and D. Hirsh**. 1979. The post-embryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. Dev. Biol. **70**:396-417.

63. **Kleiman, R. J. and L. F. Reichardt**. 1996. Testing the Agrin Hypothesis. Cell **85**:461-464.

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

65. **Krejci, E., N. Duval, A. Chatonnet, P. Vincens, and J. Massoulie**. 1991. Cholinesterase-like domains in enzymes and structural proteins: functional and evolutionary relationships and identification of a catalytically essential aspartic acid. Proc. Natl. Acad. Sci. , USA **88**:6647-6651.

66. **Kruger, M., J. Wright, and K. Wang**. 1991. Nebulin as a length regulator of thin filaments of vertebrate skeletal muscles: correlation of thin filament length, nebulin size, and epitope profile. J. of Cell Biology **115**:97-107.

67. **Labeit, S., D. P. Barlow, M. Gautel, T. Gibson, J. Holt, C.-L. Hsieh, U. Franke, K. Leonard, J. Wardale, A. Whiting, and J. Trinick**. 1990. A regular pattern of two types of 100-residue motif in the sequence of titin. Nature **345**:273-276.

68. **Labeit, S., T. Gibson, A. Lakey, K. Leonard, M. Zeviani, P. Knight, J. Wardale, and J. Trinick**. 1991. Evidence that nebulin is a protein-ruler in muscle thin filaments. FEBS **282**:313-316.

69. **Labeit, S. and B. Kolmerer**. 1995. Titins: giant proteins in charge of muscle ultrastructure and elasticity. Science **270**:293-296.

70. **Landel, C. P., M. Krause, R. H. Waterston, and D. Hirsh**. 1984. DNA rearrangements of the actin gene cluster in *Caenorhabditis elegans* accompany reversion of three muscle mutants. J. Mol. Biol. **180**:497-513.

71. **Levin, J.Z. and H.R. Horvitz**. 1992.The *Caenorhabditis elegans unc-93* gene encodes a putative transmembrane protein that regulates muscle contraction. J.of Cell Biology **117**(1):143-155.

72. **Long, R. M., H. Satoh, B. M. Marin, S. Kimura, F. J. Gonzalez, and L. R. Pohl**. 1988. Rat liver carboxylesterase: cDNA cloning, sequencing, and evidence for a multigene family. Biochem. Biophys. Res. Commun. **156**:866-873.

73. **Lu, M., D. DiLullo, T. Schultheiss, S. Holtzer, J. M. Murray, J. Choi, D. A. Fischman, and H. Holtzer**. 1992. The vinculin/ sarcomeric α-actinin/ actin nexus in cultured cardiac myocytes. J. Cell Biology **117**:1007-1022.

74. **MacKrell, A.J., Blumberg, B., Haynes, S.R., and Fessler, J.H**. 1988. The *lethal myospheroid* gene of *Drosophila* encodes a membranhe protein homologous to vertebrate integrin β subunits. Proc.Natl.Acad.Sci., USA **85**:2633-2637.

75. **Malthiery, Y. and S. Lissitzky**. 1987. Primary structure of a human thyroglobulin deduced from the sequence of its 8448-base complementary DNA. Eur. J. Biochem. **165**:491-498.

76. **Maniatis, T., E. F. Fritsch, and J. Sambrook**. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

77. **Maruyama, I. N., D. M. Miller, and S. Brenner**. 1989. Myosin heavy chain gene amplification as a suppressor mutation in *Caenorhabditis elegans*. Mol. Gen. Genet. **219**:113-118.

78. **McGhee, J. D.** 1987. Purification and characterization of a carboxylesterase from the intestine of the nematode *Caenorhabditis elegans*. Biochemistry **26**:4101-4107.

79. **McKenna, N. M., C. S. Johnson, and Y. Wang**. 1986. Formation and Alignment of Z-lines in Living Chick Myotubes Microinjected with Rhodamine-labeled Alpha-Actinin. The Journal of Cell Biology **103**:2163-2171.

80. **McKim, K. S., C. Matheson, M. A. Marra, M. F. Wakarchuk, and D. L. Baillie**. 1994. The *Caenorhabditis elegans unc-60* gene encodes proteins homologous to a family of actin-binding proteins. Mol. Gen. Genet. **242**:346-357.

81. **McMahan, U. J. and B. G. Wallace**. 1989. Molecules in Basal Lamina that Direct the Formation of Synaptic Specializations at Neuromuscular Junctions. Dev Neurosci **11**:227-247.

82. **Mello, C. C., J. M. Kramer, D. Stinchcomb, and V. Ambros**. 1991. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J. **10**:3959-3970.

83. **Moerman, D. G. and D. L. Baillie**. 1979. Genetic organization in *Caenorhabditis elegans*: fine-structure analysis of the *unc-22* gene. Genetics **91**:95-104.

84. **Moerman, D. G., H. Hutter, G. P. Mullen, and R. Schnabel**. 1996. Cell autonomous expression of perlecan and plasticity of cell shape in embryonic muscle of *Caenorhabditis elegans*. Developmental Biology **173**:

85. **Moerman, D. G., S. Plurad, R. H. Waterston, and D. L. Baillie**. 1982. Mutations in the *unc-54* myosin heavy chain gene of *C. elegans* that alter contractility but not muscle structure. Cell **29**:773-781.

86. **Moritz, C. and D. M. Hillis**. 1996. Molecular Systematics: Context and Controversies, p. 1-13. In D. M. Hillis, C. Moritz, and B. K. Mable (ed.), Molecular Systematics. Sinauer Associates, Inc. Sunderland, Mass. USA.

87. **Newman, S.M. and Wright, T.R.F**. 1981. A histological and ultrastructural analysis of developmental defects produced by the mutation, *lethal(1)myospheroid*, in *Drosophila melanogaster*. Developmental Biology **86**:393-402.

88. **Obermann, W. M., M. Gautel, F. Steiner, P.F. van der Ven, K. Weber, and D. O. Furst**. 1996. The structure of the sarcomeric M band: localization of defined domains of myomesin, M-protein, and the 250-kD carboxy-terminal region of titin by immunoelectron microscopy. Journal of Cell Biology **134**:1441-1453.

89. **Olson, P. F., L. I. Fessler, R. E. Nelson, R. E. Sterne, A. G. Campbell, and J. H. Fessler**. 1990. Glutactin, a novel *Drosophila* basement membrane-related glycoprotein with sequence similarity to serine esterases. The EMBO Journal **9**:1219-1227.

90. **Osinska, H. E. and L. F. Lemanski**. 1993. Immunofluorescent studies on Z-line-associated protein in cultured cardiomyocytes from neonatal hamsters. Cell Tissue Res. **271**:59-67.

91. **Park, E. and H. R. Horvitz**. 1986. *C. elegans unc-105* mutations affect muscle and are suppressed by other mutations that affect muscle. Genetics **113**:853-867.

92. **Patel, N. H. and C. S. Goodman**. 1992. Non-radioactive Labeling and Detection of Biomolecules. Springer-Verlag, Berlin.

93. **Pen, J., G. J. Bolks, M. L. L. Hoeksema-Du Pui, and J. J. Beintema**. 1990. Serine esterases: structural conservation during animal evolution and variability in enzymatic properties in the genus Drosophila. Genetica **81**:125-131.

94. **Piette, J., J. L. Bessereau, M. Huchet, and J. P. Changeux**. 1990. Two adjacent MyoD1-binding sites regulate expression of the acetylcholine receptor α-subunit gene. Nature **345**:353-355.

95. **Rachinsky, T. L., S. Camp, Y. Li, T. J. Ekstroem, M. Newton, and P. Taylor**. 1990. Molecular cloning of mouse acetylcholinesterase: tissue distribution of alternatively spliced mRNA species. Neuron **5**:317-327.

96. **Rees, D. J. G., S. E. Ades, S. J. Singer, and R. O. Hynes**. 1990. Sequence and domain structure of talin. Nature **347**:685-689.

97. **Rogalski, T. M., E. J. Gilchrist, G. P. Mullen, and D. G. Moerman**. 1995. Mutations in the *unc-52* Gene Responsible for Body Wall Muscle Defects in Adult *Caenorhabditis elegans* are Located in Alternatively Spliced Exons. Genetics **139**:159-169.

98. **Rogalski, T. M., B. D. Williams, G. P. Mullen, and D. G. Moerman**. 1993. Products of the *unc-52* gene in *Caenorhabditis elegans* are homologous to the core protein of the mammalian basement membrane heparan sulfate proteoglycan. Genes & Development **7**:1471-1484.

99. **Schumacher, M., S. Camp, Y. Maulet, M. Newton, K. MacPhee-Quigley, S. S. Taylor, T. Friedmann, and P. Taylor**. 1986. Primary structure of *Torpedo californica* acetylcholinesterase deduced from its cDNA sequence. Nature **319**:407-409.

100. **Smith, T. H., A. M. Kachinsky, and J. B. Miller**. 1994. Somite Subdomains, Muscle Cell Origins, and the Four Muscle Regulatory Factor Proteins. The Journal of Cell Biology **127**:95-105.

101. **Sonnhammer, E. L. L. and R. Durbin**. 1996. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene **167**:1-10.

102. **Spieth, J., G. Brooke, S. Kuersten, K. Lea, and T. Blumenthal**. 1993. Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. Cell **73**:521-532.

103. **Squire, J. M.** 1986. Muscle: Design, Diversity, and Disease. Benjamin/Cummings Publishing Co., Inc. Menlo Park, CA.

104. **Sternberg, P. W. and H. R. Horvitz**. 1991. Signal transduction during *C. elegans* vulval induction. TIGS **7**:366-371.

105. **Sulston, J., Z. Du, K. Thomas, R. Wilson, L. Hillier, R. Staden, N. Halloran, P. Green, J. Thierry-Mieg, L. Qiu, S. Dear, A. Coulson, M. Craxton, R. Durbin, M. Berks, M. Metzstein, T. Hawkins, R. Ainscough, and R. H. Waterston**. 1992. The *C. elegans* genome sequencing project: a beginning. Nature **356**:37-41.

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

107. **Sulston, J. E. and H. R. Horvitz**. 1977. Post-embryonic cell lineages of the nematode *Caenorhabditis elegans*. Dev. Biol. **56**:110-156.

108. **Sulston, J. E., E. Schierenberg, J. G. White, and J. N. Thomson**. 1983. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev. Biol. **100**:64-119.

109. **Swofford, D. L.** 1993. PAUP 3.1. **PAUP**:(Abstract)

110. **Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis**. 1996. Phylogenetic Inference, p. 407-482. In D. M. Hillis, C. Moritz, and B. K. Mable (ed.), Molecular Systematics. Sinauer Associates, Inc. Sunderland, Mass. USA.

111. **Talesa, V., E. Culetto, N. Schirru, H. Bernardi, Y. Fedon, J. Toutant, and M. Arpagaus**. 1995. Characterization of a null mutation in *ace-1*, the gene encoding class A acetylcholinesterase in the nematode *Caenorhabditis elegans*. FEBS **357**:265-268.

112. **Taylor, P.** 1991. The Cholinesterases. The Journal of Biological Chemistry **266**:4025-4028.

113. **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. Journal of Cell Science **106**:319-330.

114. **Wang, K.** 1985. Sarcomere-associated cytoskeletal lattices in striated muscle. Cell Muscle Motil. **6**:315-369.

115. **Wang, K. and J. Wright**. 1988. Architecture of the sarcomere matrix of skeletal muscle: immunoelectron microscopic evidence that suggests a set of parallel inextensible nebulin filaments anchored at the Z-line. The Journal of Cell Biology **107**:2199-2212.

116. **Waterston, R. H.** 1988. Muscle, p. 281-335. In W. B. Wood (ed.), The Nematode *Caenorhabditis Elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

117. **Waterston, R. H.** 1989. The minor myosin heavy chain, mhcA, of *Caenorhabditis elegans* is necessary for the initiation of thick filament assembly. The EMBO Journal **8**:3429-3436.

118. **Waterston, R. H. and S. Brenner**. 1978. A suppressor mutation in the nematode acting on specific alleles of many genes. Nature **275**:715-719.

119. **Waterston, R. H. and G. R. Francis**. 1985. Genetic analysis of muscle development in *Caenorhabditis elegans*. Trends in NeuroScience 270-276.

120. **Waterston, R. H., D. Hirsh, and T. R. Lane**. 1984. Dominant mutations affecting muscle structure in *Caenorhabditis elegans* that map near the actin gene cluster. J. Mol. Biol. **180**:473-496.

121. **White, J. G., E. Southgate, J. N. Thomson, and S. Brenner**. 1986. The structure of the nervous system of *Caenorhabditis elegans*. Philos. Trans. R. Soc. London **314**:1-340.

122. **Wilkins, J. A., M. A. Risinger, and S. Lin**. 1986. Studies on Proteins That Co-purify with Smooth Muscle Vinculin: Identification of Immunologically Related Species in Focal Adhesions of Nonmuscle and Z-lines of Muscle Cells. The Journal of Cell Biology **103**:1483-1494.

123. **Williams, B. D. and R. H. Waterston**. 1994. Genes critical for muscle development and function in *Caenorhabditis elegans* identified through lethal mutations. The Journal of Cell Biology **124**:475-490.

124. **Wills, N., R. F. Gesteland, J. Karn, L. Barnett, S. L. Bolten, and R. H. Waterston**. 1983. The genes *sup-7 X* and *sup-5 III* of *C. elegans* suppress amber nonsense mutations via altered transfer RNA. Cell **33**:575-583.

125. **Wood, W. B.** 1988. Embryology, p. 215-241. In W. B. Wood (ed.), The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

126. **Wood, W. B.** 1988. Introduction to *C. elegans* biology, p. 1-16. In W. B. Wood (ed.), The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

127. **Yost, H. J.** 1992. Regulation of vertebrate left-right asymmetries by extracellular matrix. Nature **357**:158-161.

128. **Zengel, J. M. and H. F. Epstein**. 1980. Identification of genetic elements associated with muscle structure in the nematode *C. elegans*. Cell Motility **1**:73-97.

Appendix A

Characterization of a clustered cholinesterase-like gene family

1. Introduction

This section describes the identification and characterization of a cDNA, pUNC112-1, and the genomic region from which it is derived. The cDNA was isolated as described in Chapter 3 as a candidate *unc-112*-derived clone. Subsequently, it was demonstrated that pUNC112-1 did not encode *unc-112*. However, the cDNA and the genomic region from which it is derived are of interest independent of *unc-112*. The cDNA encodes a potentially membrane-bound protein with a short carboxy-terminal domain and a large amino-terminal domain similar to the subfamily of cholinesterases. The cDNA is derived from one member, ZC376.2, of a tandem five gene cluster of approximately 70% similarity in amino acid sequence.

The extent of similarity with cDNA or amino acid sequences from other species is a property both of the evolutionary history of the sequences and of the function of the protein within cells; generally, highly similar protein sequences function similarly (15). The evolutionary history of homologous sequences can be reconstructed using maximum parsimony algorithms (86,110). The derived evolutionary history can be used to distinguish sequences that are merely similar from those that are true homologs. There are several types of homology, and ascertaining the type to which the selected sequences belong is essential for obtaining correct inferences about gene phylogeny and species phylogeny (86). Similar sequences that are derived by gene duplication are paralogous and may be used only to infer gene phylogenies and not species phylogenies. Similar sequences that are derived from one ancestral gene are orthologous and may be used to infer both gene and species phylogenies. Exhaustive maximum parsimony analysis determines the minimal number of evolutionary steps required to explain a given set of data, in this case amino acid sequences, by evaluating every possible tree (110). Parsimony analysis is particularly useful in evaluating the relationships of several sequences have the same percent similarity to the sequence of

interest. For example, two protein sequences may be 40% similar to a third protein sequence in that all contain immunoglobulin (Ig) repeats and therefore belong to the Ig superfamily. However, the first protein sequence may be a *bona fide* immunoglobulin and the other two sequences may be cell-cell or cell-substrate adhesion proteins containing Ig domains. Determining an evolutionary tree reveals these differences because two of the sequences tend to cluster on one branch and separate from the first.

Phylogenetic analysis of this paralogous, cholinesterase-like gene family indicate that they belong to a unique and ancient branch of the cholinesterase subfamily. The biological function of all five members are unknown at this time.

2. Materials and Methods

A. Sequence determination

The protocol for labeling and sequencing the double-stranded DNA clones with [α-^{35}S] dATP (500 Ci/mmole, Amersham) and sequence-specific oligonucleotide primers (oligos) was a modification of the Sequenase kit instructions (USB): 2-4 μg DNA, 2.5 pmole priming oligo, and 1 μl DMSO in a final volume of 9 μl was denatured by heating to 95 °C for 5 minutes and chilled on ice. The modified labeling cocktail contained 0.5 μl DMSO in addition to the manufacturer-specified components. All incubation temperatures, times, and solutions were according to the Sequenase kit instructions. All oligos used in sequencing were synthesized on a PCR Mate (model 391 DNA Sequencer, Perkin-Elmer). The sequences for all oligos have been listed in Table 5. The sequencing reaction products were electrophoresed in standard 6% polyacrylamide-8 M urea sequencing gels. The gel was transferred, dried under vacuum, and exposed to Kodak XAR film overnight. Each cDNA was sequenced on both strands a minimum of three times each. The assembly of cDNA sequences and analysis was accomplished with the Staden DNA Analysis program, version 1 (Amersham).

Another strategy was employed in the sequence analysis of p2A2. A series of deletions in p2A2 were generated using the Exo III nuclease method of Henikoff (48). In summary, DNA was digested with

TABLE 5

Oligonucleotide sequences and designations for pUNC112-1

OLIGO NAME	SEQUENCE	SEQUENCE LOCATION	STRAND LOCATION
MG2	ACG GAG CTA TTC AAC TGG	3528-3545	SENSE
MG3	GTT CAT CTG GAA GGC ATC	3093-3110	ANTISENSE
MG4	TGG AGG CGA AGG AGT ATA	2529-2546	SENSE
MG5	GAT TGG AAC TAC AAC TGC TG	3319-3338	SENSE
MG6	CGA GTC CAT CTC ACA ACT	2872-2889	ANTISENSE
MG7	TTG GAG GAT CAG GAA TGG	3776-3793	SENSE
MG8	CAT TCT GCA GGA GCA GTT	2942-2969	SENSE
MG9 *	CTT TTC TCC GTC AGT CGA	3934-3951	SENSE
MG10	GGT ACC AAA AGC ATA GGC	3665-3682	ANTISENSE
MG11	GAT TTC GCA TCC TCT CCA	3484-3501	ANTISENSE
MG12	GAC CCT GTC CTA TGT AGT	3225-3242	ANTISENSE
MG13 ✠	ATG ACG AAT CGT CCA CCA	2689-2706	ANTISENSE
MG14	CGG TGG TAT TCT CTA CGT	4240-4257	SENSE
MG16	TTG GAA CTT CGA CCT CTC	4120-4137	ANTISENSE
MG19 ✠	TGG TGG ACG ATT CGT CAT	2689-2706	SENSE
MG20 *	TCG ACT GAC GGA GAA AAG	3934-3951	ANTISENSE
MG21	GGC TTA CGG AAT CTC AGA	2482-2499	ANTISENSE

* Oligos MG9 and MG20 are complementary.

✠ Oligos MG13 and MG19 are complementary.

Sst II, which leaves a protective 3' overhang, and BamHI which leaves an Exo III-digestible 5' overhang. Approximately 300 units of Exonuclease III (Exo III) was added to 3 μg of p2A2 (Sst II-BamHI digested) and incubated at 37 °C. An aliquot of Exo III-digested DNA was removed every 20 seconds for a total of 16 samples. Single-stranded DNA ends were removed by S1 nuclease digestion at 23 °C for 30 minutes. The ends of the DNA were blunt-ended with a mixture of deoxynucleotides and DNA polymerase Klenow fragment at 37 °C for 30 minutes. The DNA was recircularized with T4 DNA ligase at 15 °C overnight. Each DNA sample was used to transform aliquots of DH5α *E. coli* using the method of Hanahan (REF) and grown on LB plates containing 100 μg/ml ampicillin. The desired deletion subclones were selected by sizing individual DNA samples digested with EcoRI on 0.8% agarose gels. Thirteen subclones were selected ranging in size from 5.6 kb (full length insert plus vector) to 3 kb (vector alone). The deletion subclones were sequenced from the deleted end using the "reverse" universal priming oligonucleotide as described above.

B. Determination of the genomic structure of the T02B5 to ZC376 region

The complete genomic sequence and partial annotation of cosmid ZC376 and sequence with no annotation of T02B5 were obtained prior to submission to ACEDB and Genbank courtesy of Dr. Steve Jones, Sanger Center, Hinxton, England. The DNA sequence of pUNC112-1 was compared to the 40 kb region spanned by both cosmid sequences using the Dotter 2.0 program (101) with a window size of 25 and minimum identity of 65% to identify the location and distribution of coding sequences. The genomic sequence was broken into 5 overlapping segments corresponding to 5 tandem genes homologous to the cDNA. Segment 1 extended from nucleotide 29200 to 18000, segment 2 from 18050 to 14000, segment 3 from 14500 to 10500, segment 4 from 11000 to 5000, and segment 5 from 6000 to 1. The locations of protein coding sequences were refined by searching for the sequences of oligonucleotides at a minimum of 70% identity in each segment using the SEARCH subprogram of the DNA analysis program written by Dr. Alan Goldin. The open reading frame of each exon was identified by comparing translations of the genomic sequence with the translation of pUNC112-1, and the probable splice junctions identified using the *C. elegans* 5' and 3' consensus sequences (30). Predicted cDNAs and translations were assembled

from the identified exons of all five genes and named ZC376.1 from segment 1, ZC376.2 from segment 2, ZC376.3 from segment 3, T02B5.a from segment 4, and T02B5.b from segment 5.

C. Sequence alignment and parsimony analysis of a clustered gene family

Amino acid sequences of *C. elegans* ZC376.1, ZC376.2, ZC376.3, T02B5.A,T02B5.B, *ges-1, ace-1*, C42D4.2, Drosophila gliotactin and neurotactin, *H. virescens* juvenile hormone esterase, *Torpedo californica* acetylcholinesterase, and mouse acetylcholinesterase were aligned using the CLUSTAL V program (Macintosh version) (50,51) and further refined by eye. One group containing only *C. elegans* ZC376.1, ZC376.2, ZC376.3, T02B5.A,T02B5.B, *ges-1* (60), *ace-1* (111), and C42D4.2 esterases and a second group containing *C. elegans* ZC376.1, ZC376.2, ZC376.3, T02B5.A,T02B5.B, Drosophila gliotactin (6) and neurotactin (26), *H. virescens* juvenile hormone esterase (46), *Torpedo californica* acetylcholinesterase (99), and mouse acetylcholinesterase (95) were subjected to exhaustive maximum parsimony analysis using PAUP 3.1 program (109). Each member of the two groups was used as the outgroup to derive phylogenetic trees.

3. RESULTS

A. Sequence determination of the cDNAs.

The sequence of p2A2 (from which pUNC112-1 was derived) was determined and used as a template to orient and overlap the other five cDNAs (see figure 25). Only the 18 bp region of p10A2 adjacent to the Reverse primer site of the vector was not homologous to the equivalent region in p2A2. Three of the clones (p2B1, p10A2, and p10C2) aligned with a poly-dT track immediately adjacent to an EcoRI linker sequence. This end also corresponded to the 3' end of the largest open reading frame in the 2 kb EcoRI fragment and did not extend into the 0.5 kb EcoRI fragment. The results indicated that the 0.5 kb EcoRI fragment in p2A2 was an artifact of cDNA library synthesis, and that the end of the native mRNA corresponded to the ends of p2B1, p10A2, and p10C2. Therefore the 0.5 kb fragment was deleted from

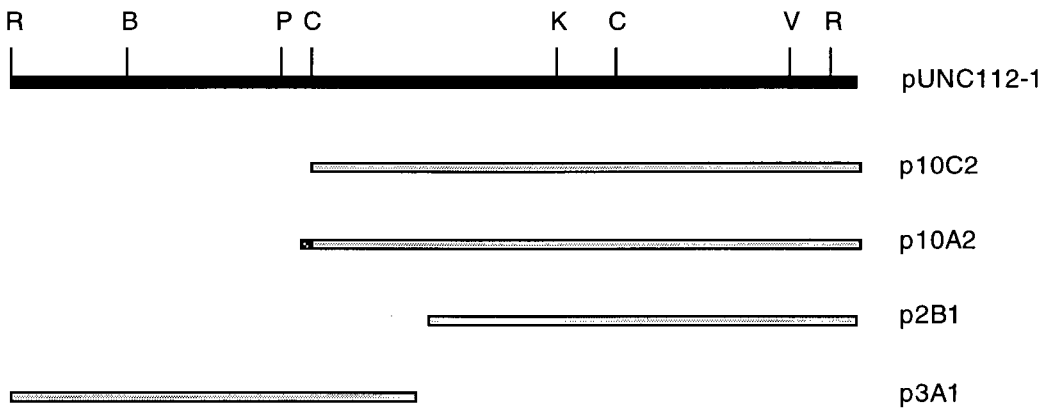


Figure 25: Alignment of four cDNAs with pUNC112-1.

The 2.2 kb pUNC112-1 clone (black bar) indicates the relative positions of key restriction sites including Bcl I (B), Cla I (C), EcoRI (R), EcoRV (V), Kpn I (K), and Pst I (P). The smaller clones (stippled bars) are aligned by size and sequence homology below pUNC112-1. The name of each clone is to the right of each bar. Clones p10C2, p10A2 and p2B1 terminate at the same 3' poly-dT run as pUNC112-1. Clone p3A1 terminates in a short poly-dT run within pUNC112-1. The 5' end of p3A1 contains 59 bp of non-overlapping sequence with pUNC112-1 which includes a polyadenylation signal. Likewise, the 5' end of p10A2 contains 18 bp of nonhomologous sequence with pUNC112-1 (gray bar). The scale is 0.5 cm = 100 bp.

p2A2 and the plasmid containing only the 2 kb fragment was renamed pUNC112-1 and further characterized.

The sequence of the 2 kb cDNA contains an open reading frame of 2037 bp beginning 25 nt in from the EcoRI cloning site (see figure 26). The predicted polypeptide is 672 amino acids in length and has a molecular mass of 76,353 Da. The sequence contains 10 cysteines, 9 of which are found on the N-terminal side of a putative transmembrane sequence. The amino acid sequence contains an 11 amino acid hydrophobic domain, a large domain of 623 amino acids with five possible glycosylation sites (57), followed by a sequence of 20 hydrophobic amino acids and a very short C-terminal domain of 16 amino acids containing two residues that are potential sites for phosphorylation (61) (see figure 27).

The pUNC112-1 sequence was initially thought to be a partial length cDNA since it did not contain a splice leader sequence type 1 (SL1) or type 2 (SL2) (albeit not all *C. elegans* genes have splice leaders), and the approximate mRNA size of 3 kb from Northern analysis differed from the 2 kb cDNA. In order to isolate the "missing" 1 kb sequence the λACT-RB1 and λACT-RB2 cDNA libraries were screened using a 1 kb Hpa I-Kpn I fragment from the 5' end of pUNC112-1. Five clones were obtained from the screening of the λACT-RB2 library (no clones were obtained from λACT-RB1) called pACT2-6, pACT2-7, pACT2-8a, pACT2-8b, and pACT2-9.

To ascertain the potential of the pACT2 cDNAs to contain sequence extending further 5' than the original 2 kb cDNA, oligonucleotides to sequences at the beginning, middle, and end of the Hpa I-Kpn I probe region, and 200 nt 5' of the Hpa I site, were hybridized to DNA spotted on filters from each cDNA clone. The results of the autoradiograph were compared with Xho 1 (the site used to construct the cDNA library in λACT; 28) digests of the cDNA inserts. The clones were predicted to contain between 500 nt and 1500 nt of additional DNA. The two largest cDNAs, pACT2-7and pACT2-8a, were sequenced using the MG21 oligo to overlap the sequence of pU112-1 with the new cDNAs and extend the open reading frame further 5'. Instead, both sequences terminated within 15 nt of the 5' end of pU112-1 and contained a single nucleotide difference that changed a TAG stop codon to a glutamine residue at position 179 in pUNC112-1. These results corrected and extended the open reading frame of pUNC112-1 an additional 42 amino acids to 670 and added a putative translation start methionine residue followed by 11

```
5/+1
AGACATGTTTTTCCGATTTTTCATTTTATGGGCAATATTGCAAAATGTGCGGTCTACCAAAGTTAAATTGTCAACTGGAACAATTG
     M  F  F  R  F  F  I  L  W  A  I  L  Q  N  V  R  S  T  K  V  K  L  S  T  G  T  I

87/28
AAGGGAGAACCCTCACCGCTTCTTATTCTCCACTCGGAAATTAGACCGGAATTGTGTTTTTCGGTGTTCCATATGTTGAACCTCCA
E  G  R  T  L  T  A  S  Y  S  P  L  G  N  */Q  T  G  I  V  F  F  G  V  P  Y  V  E  P  P

172/58
GTTGGAAATCTGAGATTCCGTAAGCCGCGTCCACCAAAACCCTGGGACGGAGTTTTGGAGGCGAAGGAGTATAAAGCGGCTTGTAT
 V  G  N  L  R  F  R  K  P  R  P  P  K  P  W  D  G  V  L  E  A  K  E  Y  K  A  A  C  M

258/87
GAGTGATCAGAAAAAAACCTACAAAAATGGTGTTGGAGGGCCTGTATCAGAAGACTGTCTTTATGCCAATGTGTTCACAAACCAAT
  S  D  Q  K  K  T  Y  K  N  G  V  G  G  P  V  S  E  D  C  L  Y  A  N  V  F  T  N  Q

344/115
ATTGTTTGGAGCATAAAAACTGCTCAGTAATGTTAACCATACATGGTGGACGATTCGTCATAGAATCAGCCTCAGCGTATGATCCT
Y  C  L  E  H  K  N  C  S  V  M  L  T  I  H  G  G  R  F  V  I  E  S  A  S  A  Y  D  P

430/144
GAAATTATAATTAATAATTTTGTGGGTCAAGGCAGAAATATTGTTGTAGTTACTTTCAACTATCGTCTAGGATTATTTGGACTGGG
 E  I  I  I  N  N  F  V  G  Q  G  R  N  I  V  V  V  T  F  N  Y  R  L  G  L  F  G  L  G

516/173
AATGCTCAATGGAGATAATCAAGATACAAATTTTGGATTATACGATATATTGGAAGTTGTGAGATGGACTCGAAAAGAAATCAAAA
  M  L  N  G  D  N  Q  D  T  N  F  G  L  Y  D  I  L  E  V  V  R  W  T  R  K  E  I  K

602/201
ATTTCGGAGGTGACAAGGATAGAATAACGATGCAGGGGCATTCTGCAGGAGCAGTTTTTACTGCCGCGTTCTCTACATCTCCGCTG
N  F  G  G  D  K  D  R  I  T  M  Q  G  H  S  A  G  A  V  F  T  A  A  F  S  T  S  P  L

922/230
TCAAAAGGTCTTATTCACCAACAAATTATAATGAGTGCATCGATGTCTAATCTTTCGAAAAAATCAAATGTGAAAGAGCTGACAGT
 S  K  G  L  I  H  Q  Q  I  I  M  S  A  S  M  S  N  L  S  K  K  S  N  V  K  E  L  T  V

1008/260
AGTTGCCAGAATAGTTGGATGCCTTCCAGATGAACATGGATTTCCAAAATTATCGAACGTAGAAGTTGAAAAAGCCTATACCTGCC
  V  A  R  I  V  G  C  L  P  D  E  H  G  F  P  K  L  S  N  V  E  V  E  K  A  Y  T  C

1094/288
TTCGTTCAAAATCTGCTCAACAAATTCTTGATGCTCAGCTGTATATGCTCCGTAACACTACATACTACATAGGACAGGGTCATATA
L  R  S  K  S  A  Q  Q  I  L  D  A  Q  L  Y  M  L  R  N  T  T  Y  Y  I  G  Q  G  H  I

1180/317
GATGGGATATATCAAGTTGATTATCCTGATAATCTATATGCAACAAATTCAATTTACCCGATTAATACTATGATTGGAACTACAAC
 D  G  I  Y  Q  V  D  Y  P  D  N  L  Y  A  T  N  S  I  Y  P  I  N  T  M  I  G  T  T  T

1266/346
TGCTGAATTGATGGATTCAAGATACATTGTTGATCCAAAAAATACTGGCAAAAAAGAAATACTACTTCAAAATCTATGCGAGCATA
  A  E  L  M  D  S  R  Y  I  V  D  P  K  N  T  G  K  K  E  I  L  L  Q  N  L  C  E  H

1352/374
TTGGTTATGAACTCTACAAAGAGCCTGAAGAGTTTTCACAAAAATGTCGCAACTTTTACAAAAATGGAGAGGATGCGAAATCATTG
I  G  Y  E  L  Y  K  E  P  E  E  F  S  Q  K  C  R  N  F  Y  K  N  G  E  D  A  K  S  L

1438/403
GGGGATGATATGGAATTTTATTACGGAGCTATTCAACTGGCGAATGTACATTCCAGCAAAGATACAAACGTTTTTGTGTACAGCTT
 G  D  D  M  E  F  Y  Y  G  A  I  Q  L  A  N  V  H  S  S  K  D  T  N  V  F  V  Y  S  F

1524/432
TGACTACAAGGGAGCTGGGAATGCGTACAAGAAATATGACAAATCTCTTTCACCCAAACACGGCGAAGACTTCGCCTATGCTTTTG
  D  Y  K  E  A  G  N  A  Y  K  K  Y  D  K  S  L  S  P  K  H  G  E  D  F  A  Y  A  F

1610/460
GTACCAATCGCGGAAATTTCTCGACAAAGGATTATGTTATCGAGTATATTTATAGTGGAATGTTTGCTGATTTTGTGAACTTTGGT
G  T  N  R  G  N  F  S  T  K  D  Y  V  I  E  Y  I  Y  S  G  M  F  A  D  F  V  N  F  G

1696/489
GATCCTTCACCTTTGGAGGATCAGGAATGGGCGCAGTATACCCCGGAAAAACGAGAATATTTTTTGATCGATTTTGATAAGAACTT
```

```
 D  P  S  P  L  E  D  Q  E  W  A  Q  Y  T  P  E  K  R  E  Y  F  L  I  D  F  D  K  N  F

1782/518
TACAATGCCTGGAATGAAAGATCACTATTACCCAAAGGCTTTGGAGTTTTGGAGTACATGTGGAAGCAAGTCTTTCAAGGAACACT
  T  M  P  G  M  K  D  H  Y  Y  P  K  A  L  E  F  W  S  T  C  G  S  K  S  F  K  E  H

1868/546
TTTCTCCGTCAGTCGATATATTTATAATCGGAAATCTCTTGAATCCAATTATGTCTCATTTGAATCATAATGAAACTGGCCCCGAT
F  S  P  S  V  D  I  F  I  I  G  N  L  L  N  P  I  M  S  H  L  N  H  N  E  T  G  P  D

1954/575
AAGACTTTTGAGCAATTTGATAAATTGTATAACGAACGGGAAGAGTTTTTGAAACTTCTAAAAGCTGTAAGGAAATTGGAAATTCA
 K  T  F  E  Q  F  D  K  L  Y  N  E  R  E  E  F  L  K  L  L  K  A  V  R  K  L  E  I  Q

2040/604
AAAGAAAATGTGGAGAGGTCGAAGTTCCAAAGATTTAATGCTGAAAGAGCTTGTGGATTTGGAGAAAATAAACGAAGAAGAGCAAG
  K  K  M  W  R  G  R  S  S  K  D  L  M  L  K  E  L  V  D  L  E  K  I  N  E  E  E  Q

2126/632
AAACTGGAGCTAATTTCCTTCTCATAATATTTGGTGGAACTCTTCTCGGTGGTATTCTCTACGTCTCCATTTCTCATTTCTGCCTC
E  T  G  A  N  F  L  L  I  I  F  G  G  T  L  L  G  G  I  L  Y  V  S  I  S  H  F  C  L

2212/661                          2245/672
CATCATCGTTCTCGTGATGGATATCAGCTCTTGAAATAATTATTCAAACTTATGTGTTTGTCTGTATTATGATTTTTATGTCTAAA
 H  H  R  S  R  D  G  Y  Q  L  L  K  *

2298
TAAATGTTCTCTTACAAAAAAACGGAATTCCGATTAATTTTTTATTACAATCTGTTCATAGAAAAGCCTAAAACTTGTGGCAGTGA

2384
GATTTAAGTAAAATGGAAAATATAGAAACAAGAAAAATATCGCTGAGAATGTGGTGTTGTGGGAAGAACTTTGCACTGTACCCGTT

2470
GTCCAGTTTCCTTCTTAGGCTCTGTACCCAGCGTCTTGAGCTGGGGCCTCTGGAGCTGTCTCTTCTTGGGCAACTGGCTCGGCGGC

2556
TTGCTCTACTGGTTGTTCGGCAGCGGGGCAGACTCTTCAACAGCTGGAGCAGCAGCGCTTCATCTCCATAAGAGTTTTGGGCTTGT

2642
CTCTTAGATCTGTATCCGGCGTCTTGAGCTGGGGCCTCTGGGGCGGCCTCTTCTTGAGCAACTGGCTCGGCGGCTTGCTCTACTGG

2728
TTGTTCGGCAGCGGGGGCAGACTCTTCAACAGCTGGAGCAGCAGCGGCTTCATCTCCGTAAGAGTTTTGGGCTTGTCTTTTAGATC

2814                                                  2858
TGTATCCGGCGTCTTGAGCTGGGGCCTCTGGGGCAGCCTCTTCTTG
```

Figure 26: The coding sequence and translation of pUNC112-1.

The single letter amino acid code is shown below and the base number/amino acid number above each line of DNA sequence. The cholinesterase amino acid motifs are in **bold italics**. All cysteines are in **bold** type, with those likely to form cholinesterase loops A, B and C double underlined. The predicted transmembrane motif is underlined.

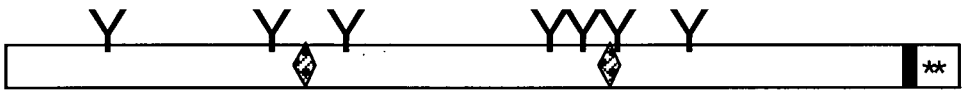


Figure 27: Domain structure of the predicted protein sequence of pUNC112-1.

The open reading frame of the cDNA (open bar) is 2.2 kb (0.5 cm = 100 bp). The relative positions of seven potential glycosylation sites (Y), cholinesterase active site motifs (diamonds), one potential phosphorylation site (**) and putative transmembrane domain (black box) are indicated.

hydrophobic residues. Furthermore, the termination of pACT2-7and pACT2-8a sequences very close to the termination of pUNC112-1 indicates that pUNC112-1 is very likely to be a full length cDNA.

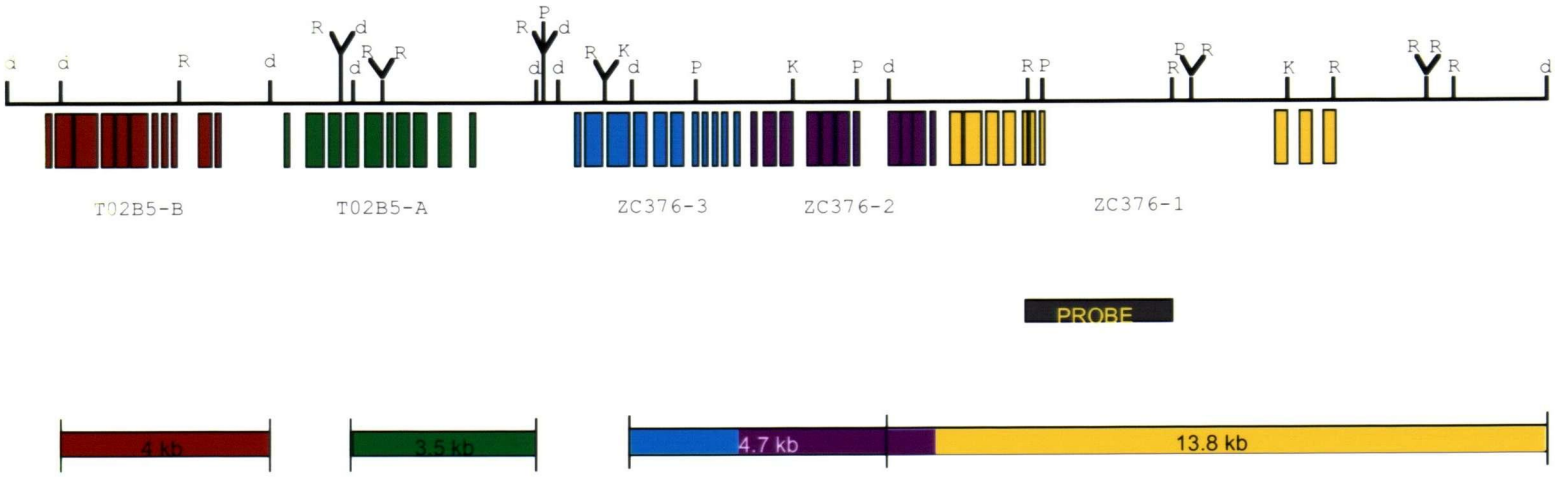
B. Genomic structure of the region encoding pUNC112-1

The *C. elegans* genome sequencing groups in St. Louis, USA and Hinxton, UK have recently completed sequencing the region containing cosmids T02B5 and ZC376 and forwarded the information to me prior to submission to ACEDB, Genbank and EMBL databases. The sequence of pUNC112-1 was compared to both cosmid sequences using Dotter 2.0 (101) to identify similar regions. Suprisingly, five tandem copies, one 100% identical to the cDNA and four with greater than 70% similarity to the cDNA, were identified in the genomic sequence (see figure 28). The five genes have been named ZC376.1, ZC376.2, ZC376.3, T02B5.A and T02B5.B according to the nomenclature of the *C. elegans* genome sequencing project. The first three genes are spaced closely together with gaps of approximately 300 bp. The last two genes are separated from ZC376.3 by 2 kb, and from each other by 1.5 kb. The putative intron and exon boundaries were determined for each of the five genes via computer analysis. ZC376.2, ZC376.3, T02B5.A and T02B5.B are composed of 11 exons whereas ZC376.1 is composed of 10 exons (exon 2 of ZC376.1 is divided into exons 2 and 3 in the other four genes). Furthermore, ZC376.1 contains an unusually large (4.3 kb) intron between exons 3 and 4 which is predicted to contain another gene, ZC376.4, in the antisense orientation to the five tandem genes.

Model cDNAs were assembled from the exons of each of the five tandem genes and used to examine the similarities of the predicted gene products. The predicted protein sequences of all five model cDNAs indicate that they have the same organization as the predicted protein from pUNC112-1: a 12-19 amino acid (AA) hydrophobic sequence at the N-terminus, a 618-715 AA hydrophilic domain, a 20 AA hydrophobic sequence, and 14-19 AA hydrophilic C-terminus. The five proteins are on average 40% identical, with up to 55% identity in the amino acids from 101-201 and down to 0.5% identity in the first 20 amino acids of the N-terminus and 40 amino acids just before the predicted transmembrane domain (see figure 29).

Figure 28: Genomic organization of a tandem five gene cluster identified by hybridization with pUNC112-1.

The cluster spans cosmids ZC376 and T02B5 over approximately 21 kb. Transcription of all five genes is from right to left. The exons for ZC376.1 are magenta, ZC376.2 are in purple, ZC376.3 are blue, T02B5.A are in green, and T02B5.B are red. A sixth gene, ZC376.4, is predicted to lie within the 3/4 intron of ZC376.1 and is transcribed left to right (not shown). Notice the very short distances between the first and last exons of adjacent genes, particularly between ZC376.1, Zc376.2 and ZC376.3. The relative positions of EcoRI [R], HindIII [d], Kpn I [K] and Pst I [P] sites are indicated on the line above the exons. The location of the estimated 2.5 kb EcoRI fragment used in cDNA cloning (gray box) and the HindIII fragments that hybridize with the probe (open boxes) are indicated below the exons. The scale is 1 cm = 1 kb.


T02B5-B
T02B5-A
ZC376-3
ZC376-2
ZC376-1
PROBE
4 kb
3.5 kb
4.7 kb
13.8 kb

```
ZC376.1         MILQYLYF-AVLIHLSISTRIQLSTGTIEGKILNATYSPFGNQTATVFLG
ZC376.2         MFFRF-FILWAILQNVRSTKVKLSTGTIEGRTLTASYSPLGNQTGIVFFG
ZC376.3         MVLLFLFLLSTLLHISISTRVQLSTGTIEGKILTSTYSPLGNHTGIGFLG
T02B5.A         MLKPATVLLLI--QMYVCKRVQLSSGTIEGKILNISYSPLGNQSATVFLG
T02B5.B         MIFKFVVLLWFSAQFAISTKVQLST-IIEGKQLNSTYSPFGNQSAIAFLG
                *      .     .      ...**.  ***. *  .***.**...  *.*

ZC376.1         IPYVEPPIGELRFRKPRLLKSWEGVLETKDYKPACMSYWRKTFKNGFVGE
ZC376.2         VPYVEPPVGNLRFRKPRPPKPWDGVLEAKEYKAACMSDQKKTYKNGVGGP
ZC376.3         IPFVEPPAGNLRFRKPRPPIPWDGILETKEYKPACMSDAKKTYKNGVGGP
T02B5.A         IPFVEPPIGDLRYRKPRPPKSWEGVLVTNEYKSACMSNATKTYKNKFGGP
T02B5.B         IPYVEPPIGDLRFRKPRPQKTWSGVLETKKFKPACMSNSTKTYKNGIGGP
                .*.**** *.**.****    * *.* .. .*.****   **.**   *

ZC376.1         ISEDCLYANVFTNQYCLQNKNCSVMIVVHGGRLVCESGSAFKPEIFINNF
ZC376.2         VSEDCLYANVFTNQYCLEHKNCSVMLTIHGGRFVIESASAYDPEIIINNF
ZC376.3         ISEDCLYANVFTNQYCMKRKNCSVMIVVHGGRILTESASAFNPEILVNNF
T02B5.A         ISEDCLYLNVLTNEYCLENKNCSVMMIVHGGGYLTESASTFNPEILINNF
T02B5.B         ISEDCLYLNVLTNQYCLENKNCSVVIIVHGGGYLFESAATFNHEILVNNF
                .****** **.**.**.  *****.. .***  . **....  **..***

ZC376.1         VGQDRNIVVVTFNYRLGVFGFGVFNGETGDTNVAMYDMLEAVKWVRKEID
ZC376.2         VGQGRNIVVVTFNYRLGLFGLGMLNGDNQDTNFGLYDILEVVRWTRKEIK
ZC376.3         VGQGRNIVVVTFNYRLGTFGFGVLNGEKGDSNVGMFDMLEAVKWTRKEVH
T02B5.A         VGQGRNIVVVTFNYRLGLFSFGQFNGDRGDKNFGLYDMIESVNWVRREIE
T02B5.B         VGQDRNIVVVTINYRLGIFGFGHLNGFEGDKNVGIFDTLEAVKWVRREIK
                ***.*******.***** *..* .**   * * ...* .* * * *.*.

ZC376.1         QFGGNKDRITMAGHSAGAGLIVDFTSSTLSKGLLHQQIVMSAPLQDISKS
ZC376.2         NFGGDKDRITMQGHSAGAVFTAAFSTSPLSKGLIHQQIIMSASMSNLSKK
ZC376.3         NFGGNKDKLTMVGHSAGGSLVGAFTSSPLSKGMLNQQIIMSGALYQMGKL
T02B5.A         NFGGNKNRITLAGHSAGASMIVAFTSSPLTKGLVHQQIIMSAPMTNMSKK
T02B5.B         QFGGNKDRITLAGHSAGAGVVAAFTISPLTKGLIHQQIVMSGPMTNMSKK
                .***.*...*. *****.     *. * *.**...***.**... ...*

ZC376.1         ANFKGMTIVAQNVGCIPKEYGFRKLSKTQINKTYLCLQNKSAQDLLHAQL
ZC376.2         SNVKELTVVARIVGCLPDEHGFPKLSNVEVEKAYTCLRSKSAQQILDAQL
ZC376.3         ANFKGMTAMAQKAGCLPEVFGFRKLSQVRIDKTYSCLRNISAQEVLDAQL
T02B5.A         SNFKGMTVMAQMVGCLSEEIGFNKLSEEQVENTYSCLRKKSAQQILDAQL
T02B5.B         SNFKGMTAMAERVGCHSKKYGFNKLPEHEVEQVYSCLRKKPAQELLDAQL
                .* * .* .*   ** .   ** **.   ... * **.  .**..*.***

ZC376.1         SMQQNSTFYFGSPRVDGDFITDYPDNLFNFNTIYPINTFIGTTTGELRDS
ZC376.2         YMLRNTTYYIGQGHIDGIYQVDYPDNLYATNSIYPINTMIGTTTAELMDS
ZC376.3         WVLQNTTYYISVPHIDGEFLLNYSDEILASGTIHPINSMIGTTTAELRDP
T02B5.A         WLLQNSTYFLGAPPIDEHFLTDYPENLYASKSIYPINTLIGTTTLEVEES
T02B5.B         WLLQNSTFYFGAPPVDGEFITDYPDNLYSEKSIFSINTLLGTTTRELQDT
                 . .*.*....   .*  .  .*....    .* .**. .**** *. .

ZC376.1         --LYITDPKND--RIKEQLLKNMCEHVGYELFEKPEEFTKKCGNYYKNGT
ZC376.2         --RYIVDPKNT--GKKEILLQNLCEHIGYELYKEPEEFSQKCRNFYKNGE
ZC376.3         --IYINDLKNA--DKKEELLKNLCEHIGYELYTEPEEFSRKCQKFYGNGD
T02B5.A         S--YIIDPAFA--DKKVELLENLCDHIGYVLYEEPETFSKKCQKQYMNGN
T02B5.B         GQRYISNSNYSKFETKETFLKNLCDQIGIEIYKESEDFSSKCQKWYINGS
                   ** .        *  .* *.*...*  ..  .* *. ** . * **
```

```
ZC376.1         DAQFLSDDMEFYSGAIKVANAHKRANSKVFMYSYDYKGAGTAFHKYLEAP
ZC376.2         DAKSLGDDMEFYYGAIQLANVHSSKDTNVFVYSFDYKEAGNAYKKYDKSL
ZC376.3         DAQFLADDMEFYDGSIKVANAHASKNTKVFMYSYDYKDAGPAFKKYAQAP
T02B5.A         SSMNLSNEMEFYTPAIDFADSHTSGNTKVFLYVYDYRGAGPAYDRYLEVR
T02B5.B         DSKSLSDDMEFYTQAINLADAHARKGKKVYMYSYAYSGAGPAFKKYMKVK
                 .  *...****  .*  *. *     .*..* . *  ** *. .*

ZC376.1         SPHHSEDLIYTFGTSRGPFVAKDYVIERIYSGMLANFINFEDPSPSKTQQ
ZC376.2         SPKHGEDFAYAFGTNRGNFSTKDYVIEYIYSGMFADFVNFGDPSPLEDQE
ZC376.3         PPHHSEDLIYVFGTSRGNFTEKDYVIEQIYSGMFADFVNFGNPSPSKEQK
T02B5.A         SPHHSEDLIYVFGTHRGIFAPKDYIIEKIYSGMFADFVNFENPLPSGDQK
T02B5.B         SPHHSE-LIYTFGTHRGVFAPKDYVIEYIYSGMFADFINFEDPSPSTDQP
                .* *.*  . * *** ** *  ***.** *****.*.*.** .* *   *

ZC376.1         WRQYTQEKREYFLIDFDKNFTMPGMKDHYYTRALDFWSTAGKKSFSERFS
ZC376.2         WAQYTPEKREYFLIDFDKNFTMPGMKDHYYPKALEFWSTCGSKSFKEHFS
ZC376.3         WKEYSPEKREYFLIDFDKNFTMPGTRDGYYSRALEFWSTAGTKSFSEHYS
T02B5.A         WNQYTKENREHFLINFDKNFITPGMRDNYYTEAYEFWSTVGKKSFKEEWS
T02B5.B         WLQYTKENKEYFLVDFDQNLTMPGMKKNYYAKAYEFWSTAGQKSFKEEWS
                * .*. *..* **..**.*.  ** .  **  * .**** * *** *  *

ZC376.1         PSLDNFVMSILTDPIVSHLKGVASGPDKSIEQFEKVFNERESFLEELKLR
ZC376.2         PSVDIFIIGNLLNPIMSHLNHNETGPDKTFEQFDKLYNEREEFLKLLKAV
ZC376.3         PSLDSFTTGILIDPIVSHMKGVATGPDKTFEQFEKMLLEREMFLKTLKSE
T02B5.A         PSLDTFTCALVISPLVSHMKQTTTAFDKTFEQTELLYKEEVNFLKREKLE
T02B5.B         PSLDTFIITNLVAPIITHMDQITFDVDKTVEQMEILKIERENFLKSEKLK
                **.* *    .  *...*.       **. ** . .  *   **   *

ZC376.1         RKLKLEKKIGVQMRNS---VVEFE--------KDADQGTGGVLSEFQYYT
zC376-2         RKLEIQKKMW---RGRSSKDLMLK-ELVDLEKINEEEQETGA--NF----
ZC376.3         RKLELLKEKWRKIRNGGGRIVERA-GNAPMEMRDGKKGSRGGVS-----L
T02B5.A         RTQELKMETKR--RDKALRIQNRKNGLANKEITEGDEEDESS----KLDI
T02B5.B         RKSELEMSRWI--KKGRKRVINKRMGFMPVEIIENEESESDG---GFSI
                *   .        .                    .

ZC376.1         LPIILGVTVFGSIIYASISHFCIPNKSRTGYQLLE
ZC376.2         LLIIFGGTLLGGILYVSISHFCLHHRSRDGYQLLK
ZC376.3         LLIIFGGTLLGGILYVTISHFCLHHKSREGYQLLK
T02B5.A         LLIISAGTLFGGILYVTLPNVILQKRARDGYELLS
T02B5.B         GLLLFGGTLLGGIVYV--SSYQFFRRSRNGYQLIN
                  .. . *..*.*.*   .   .  ..* **.*.
```

Figure 29: Protein comparison of clustered five gene family.

Single amino acid codes for each of the predicted protein products of ZC376.1, ZC376.2, ZC376.3, T02B5.A and T02B5.B. The active site serine and histidine domains are double underlined, the predicted transmembrane domains are single underlined, similar amino acids are identified by (.) and identical amino acids by (*) and **bold** type. Notice that two regions, one at the N-terminus and a second just before the transmembrane domain, have very little identity between the five proteins.

C. Comparative sequence analysis

The predicted pUNC112-1 cDNA sequence and the model cDNA sequences from the five tandem genes were compared with the peptide and nucleotide databases at the National Center for Biotechnology Information (NCBI) GENINFO ® BLAST network service. The results of the database search showed that all of the predicted protein sequences are related to the cholinesterase subfamily of serine proteases. The predicted protein sequences contained all the hallmarks of a cholinesterase (112): 1) proteolytic sequences shared between cholinesterases and serine hydrolases flanking the "active site" S-200 and H-440; 2) the order of the serine and histidine residues in the primary sequence characteristic of cholinesterases [S-200, D-397, H-440 for Torpedo ACE (65)] and not of the trypsin or subtilisin subfamilies (which are H-57, D-101, S-195 for trypsin and D-32, H-64, S-221 for subtilisin); 3) conservation of disulfide loops A, B, and C (in the larger family members) formed by six cysteines; and 4) greater than 50% overall similarity of amino acid residues. It was surprising to note that the *C. elegans* general esterase, *ges-1* (60,78), had less similarity to the five predicted protein sequences than the cholinesterases from other organisms. Some notable similarities from the BLAST search were: acetylcholinesterase from *Torpedo* (99), carboxylesterase from rat (72), glutactin (89) and neurotactin (9) from *D. melanogaster*, thyroglobulin from human (75), and juvenile hormone esterase from the moth *H. virescens* (46) (see figure 30). Recently, another gene called gliotactin from *D. melanogaster* was found to have sequence and structural similarity to the predicted pUNC112-1 protein (6). Significantly, the five predicted proteins from ZC376 and T02B5 contain a serine (S-172) shown in other cholinesterase family members to be essential for enzyme activity. This serine residue is replaced in thyroglobulin, glutactin, neurotactin and gliotactin either by glycine, arginine, or alanine. These proteins are believed to be non-functional esterases that perform other functions such as cell-substrate adhesion. Both juvenile hormone esterase and acetylcholinesterase contain this serine residue in the same relative position as the five predicted proteins and are functional esterases. Therefore, it is possible that the five predicted proteins from ZC376 and T02B5 function as membrane-bound esterases.

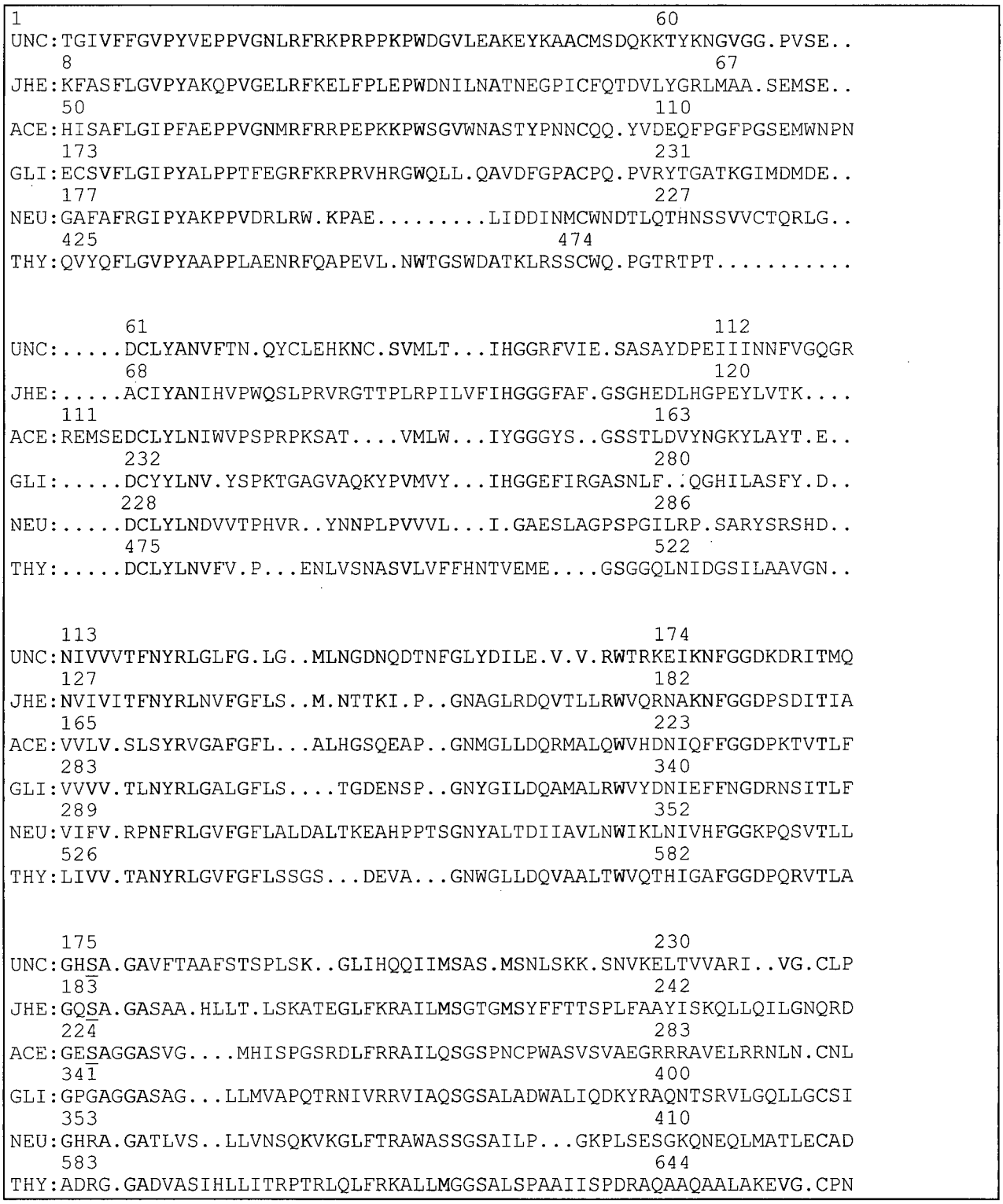


Figure 30: Protein sequence alignment of pUNC112-1 with highly similar proteins.

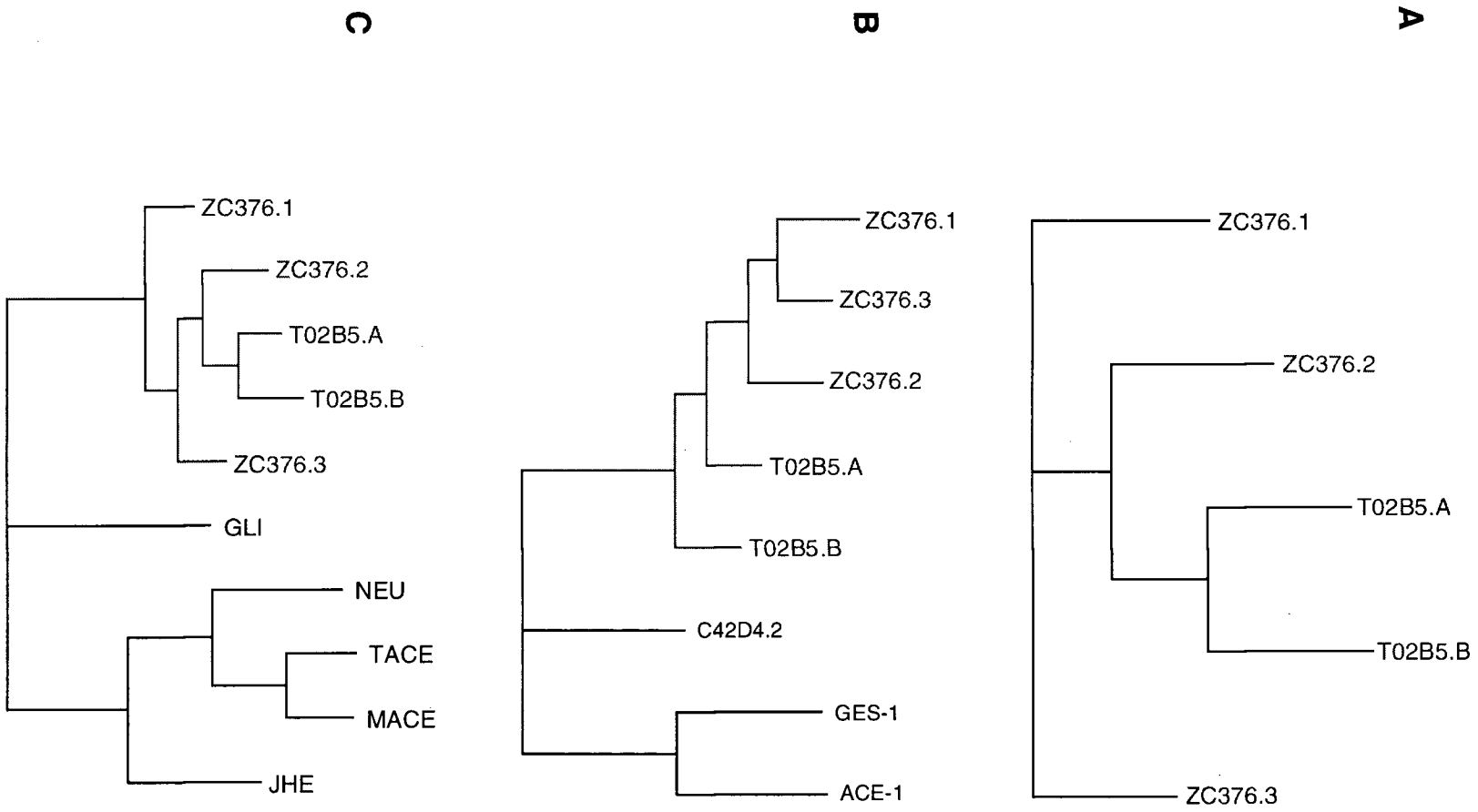
Single amino acid codes of putative UNC112 protein (UNC), H. virescens juvenile hormone esterase (JHE) (46), Torpedo acetylcholinesterase (ACE) (99), Drosophila gliotactin (GLI) (6), Drosophila neurotactin (NEU) (9), and bovine thyroglobulin (THY) (75) are compared. Identical amino acids are in boldface, corrective gaps are represented by dots. The essential serine residue of the esterase active site is underlined.

D. Parsimony analysis of five clustered cholinesterase family members

The predicted protein sequences of all five model cDNAs were compared to each other, other *C. elegans* esterases, and cholinesterases from a variety of organisms using maximum parsimony analysis in order to examine the phylogenetic relationships among the genes. A tree consisting only of the clustered, five gene, paralogous family and rooted using ZC376.1, ZC376.3, or ZC376.2 indicates that T02B5.A and T02B5.B are most closely related (see figure 31a). When other *C. elegans* esterases were then added for comparison and individually used to root the trees, the gene family remained grouped together and undisrupted by the general gut esterase *ges-1* (60), acetylcholinesterase *ace-1* (111), or the unknown esterase C42D4.2 identified by the *C. elegans* genome sequencing project (see figure 31b). This result suggests that the clustered gene family has evolved independently of other *C. elegans* esterase genes. The paralogous gene family from *C. elegans* was then compared with esterases from Drosophila, Torpedo, mouse and moth (see figure 31c). The *C. elegans* gene family remained grouped together on one of the three branches of the tree and undisrupted by mouse or Torpedo acetylcholinesterases, Drosophila gliotactin or neurotactin, or *H. virescens* juvenile hormone esterase regardless which of these esterases was used to root the trees. This result indicates that the phylogeny of the clustered gene family is distinct from gliotactin, acetylcholinesterases, neurotactin, and juvenile hormone esterase.

Figure 31: Phylogenetic trees predicted by maximum parsimony analysis for the clustered five gene family of cholinesterase-like proteins.

Panel **A** is a cladogram rooted using ZC376.1or ZC376.3 to examine the relationships of ZC376.1, ZC376.2, ZC376.3, T02B5.A and T02B5.B. T02B5.A and T02B5.B are most closely related because they branch from the same node. Panel **B** is a cladogram using C42D4.2 to root the tree and comparing the five gene family with other *C. elegans* esterases: *ace-1* (111), *ges-1* (60) and C42D4.2 (identified by the genome sequencing group). Notice that the five gene family branches off from the other esterase genes at the earliest point of the tree indicating that they are not closely related phylogenetically to these other esterases. CI=0.920. Panel **C** is a cladogram using gliotactin to root the tree and comparing the five gene family of *C. elegans* with other cholinesterase family members from Torpedo acetylcholinesterase (TACE) (99), mouse (MACE) (95), Drosophila gliotactin (GLI) (6) and neurotactin (NEU) (9), and *H. virescens* juvenile hormone esterase (JHE) (46). Notice that again the five gene family branches off from the other esterase genes at the earliest point of the tree. This suggests that these genes may form a unique subfamily within the cholinesterases with a function different from cell signal modification or cell adhesion. CI=0.934.


C
ZC376.1
ZC376.2
T02B5.A
T02B5.B
ZC376.3
GLI
NEU
TACE
MACE
JHE
B
ZC376.1
ZC376.3
ZC376.2
T02B5.A
T02B5.B
C42D4.2
GES-1
ACE-1
A
ZC376.1
ZC376.2
T02B5.A
T02B5.B
ZC376.3

4. Discussion

A. Sequence analysis of pUNC112-1

The sequence of pUNC112-1 contains an open reading frame of 2037 bp encoding a predicted polypeptide of 670 amino acids in length with a molecular mass of 76,353 Da. The 11 N-terminal amino acids are hydrophobic and may serve as a signal to transport the molecule to the cell surface (1). A large domain of 623 amino acids follows the short hydrophobic domain and contains 9 cysteines and five possible glycosylation sites (57), indicating that it may be an extracellular domain. These cysteines may be used for intra- and inter-molecular disulfide bonding to stabilize the conformation of the domain and to anchor the molecule to the basement membrane. This domain is followed by a sequence of 20 hydrophobic amino acids capable of spanning the cell membrane once, and a very short C-terminal domain of 16 amino acids containing the tenth cysteine and two residues that are potential sites for phosphorylation (61) and therefore is likely to be located intracellularly (see figure 27).

The protein predicted from the sequence of pUNC112-1 appears to be a member of the cholinesterase subfamily of serine esterases based on a comparison with the NCBI database that shows greater than 50% identity to this class of proteases. Members of the cholinesterase subfamily possess two conserved regions containing amino acid residues essential for esterase activity and six cysteines forming three loops. Some members of the cholinesterase subfamily such as acetylcholinesterase and insect juvenile hormone esterase (46,99) inactivate signaling molecules, whereas other members lacking the active site serine may be capable of protein-protein interactions for cell adhesion or recognition (6,9,26). The deduced amino acid sequence of pUNC112-1 is an average of 42% identical to juvenile hormone esterase and acetylcholinesterase and an average of 38% identical to gliotactin and neurotactin (see figure 21). Six of the nine cysteines in the esterase domain are located in positions consistent with forming the three disulfide loops characteristic of cholinesterases. The other three cysteines may be involved in disulfide bonds linking this domain with other extracellular proteins such as collagen type IV.

In addition, pUNC112-1 protein sequence possesses serine and histidine residues in the same relative positions as the catalytic serine and histidine residues in acetylcholinesterases. The higher percentage of identity and the potential for catalytic activity together indicate that pUNC112-1 protein may bind a signaling molecule and modify it by hydrolysis of an ester bond. The potential for pUNC112-1 to possess esterase activity could be tested with a colorimetric in-gel assay (93).

In addition to encoding a potentially-functional esterase domain, the pUNC112-1 exhibits a hydrophobic segment capable of spanning the muscle cell membrane and a very short cytoplasmic tail which may be phosphorylated. Therefore, the pUNC112-1 protein may be membrane-bound in order to modify an unknown signaling molecule and then signal this activity to the inside of the cell. If this protein is located in muscle cell membranes, a strategic and intimate association with basement membrane, dense body and M-line proteins may be possible. However, no information has been obtained concerning the type of cell expressing pUNC112-1 protein or its distribution. These issues can be addressed by the development of specific antibodies to a unique portion of the protein. Recently several antibodies have been raised to a portion of sequence flanking the active-site histidine fused to GST. Once the appropriate conditions for use of these antibodies have been determined, the in situ distribution of pUNC112-1 protein can be examined by indirect immunofluorescence microscopy and compared with the distribution of pUNC112-1 mRNA labeled with digoxigenin (92). The results of these experiments will indicate the type of cells expressing the mRNA and the site of distribution of the protein. From the primary protein structure, determination of esterase activity, and the cell type and distribution results it may be possible to formulate and test hypotheses for the functional role of this novel protein in *C. elegans*.

B. Sequence analysis of the genomic structure.

Sequencing of cosmids T02B5 and ZC376 has recently been completed by the *C. elegans* genome sequencing groups in St. Louis, USA and Hinxton, UK. A comparison of pUNC112-1 with the cosmid sequence revealed five tandem esterase-like genes: ZC376.1, ZC376.2, ZC376.3, T02B5.A and T02B5.B. The cDNA corresponded to ZC376.2 and was approximately 70% identical to the other four genes at the DNA level. The cDNA was used to determine the exon-intron structures of all five genes. Interestingly,

ZC376.1 differs from all the others in two ways which may have implications for its phylogeny (see below): 1) intron 2/3 is removed, fusing exons 2 and 3 into a larger exon 2; and 2) intron 3/4 is unusually large and is predicted to contain another gene in the antisense orientation, ZC376.4. Furthermore, the first three genes are very closely spaced, with approximately 300 bp separating the 3' end of one gene from the 5' end of the next gene. In contrast, the next two genes are separated from the first three by 2 kb and each other by 1.5 kb. This suggests that the regulatory regions for at least the first three genes, and possibly all five genes, are located primarily 5' of ZC376.1 and may be transcribed as a polycistronic message (102) or include an untranslated segment located in an "exon A" 5' of ZC376.1 which is spliced to each gene product by "skipping over" the exons for the intervening genes in a mechanism analagous to the mechanism described for *cha-1/unc-17* (3).

All of the gene products from the five genes have the same primary structure: a short hydrophobic sequence at the N-terminus that is not long enough to span the membrane but may serve as a signal to transport the protein to the cell surface, a large hydrophilic domain with significant similarity to cholinesterases and likely to be glycosylated and disulfide-linked to other proteins in addition to forming at least three intramolecular disulfide bonds, another hydrophobic domain capable of spanning the cell membrane, and a short hydrophilic domain at the C-terminus with potential sites for phosphorylation. The members of this gene family therefore may function by very similar mechanisms but either in different cell types or at different times in development. In the future, the spatial or temporal distribution of each of these proteins might be determined using antibodies unique to each protein with indirect immunofluorescent staining of animals at all developmental stages.

C. Parsimony analysis of five clustered cholinesterase family members.

The similarity of the sequences from the members of clustered five gene family with cholinesterases from other *C. elegans* genes and from other species is an indication of the function of these proteins in development or maintenance of cells and the evolutionary history of these genes. Parsimony analysis was used to evaluate the similarity of the members of the gene family with each other, other *C. elegans* cholinesterases, and cholinesterases from other species in order to predict the phylogeny of this

paralogous gene family. In all analyses, T02B5.A and T02B5.B are most closely related, ZC376.2 is more closely related to T02B5.A and T02B5.B than to ZC376.1 or ZC376.3. The relationships of ZC376.1 and ZC376.3 appear to be less closely related to the other three members but their positions on the tree branch is uncertain. The loss of intron 2/3 and insertion of an antisense gene in intron 3/4 of ZC376.1 suggests that this gene has diverged the most from the other four genes evolutionarily. In order to determine whether this gene is more ancient or more recent than the others it could be evaluated at the DNA level with the other four sequences by including the deleted intron and the altered intron sequence in the parsimony analysis.

The results of exhaustive searches for the evolutionary tree with the fewest number of changes between proteins consistently grouped the paralogous genes together, and the other cholinesterases were grouped on other branches of the tree. Significantly, the general esterase of *C. elegans*, *ges-1*, is more closely related to the acetylcholinesterase encoded by *ace-1* than either gene is to the five gene family. Likewise, Drosophila neurotactin and *H. virescens* juvenile hormone esterase are more closely related to mammalian and marine ray acetylcholinesterases than all of these genes are to the five gene family. An examination of the ESTHER representation of evolutionary relationships of a wide range of esterases and lipases (25) indicates that non-specific esterases and neurotactin are three nodes removed from acetylcholinesterases. Therefore, these results indicate that the members of the five gene family are of very ancient origin, and may be at least four nodes removed from acetylcholinesterases on ESTHER, and probably form a subfamily of esterases with a unique enzymatic function.

Appendix B

Characterization of an unequal crossover event within the clustered cholinesterase-like gene family

1. Introduction

This section describes the identification and characterization of the site of unequal recombination within the clustered cholinesterase-like genomic region from the nematode strain DM1208. Unequal recombination was suggested from figure 21 which contains all wildtype Hind III bands plus and additional 2.8 kb band when hybridized with a 2.5 kb Kpn I-EcoRI fragment from cosmid T02B5. The most conserved exons from all 5 genes were tested in various combinations for the ability to generate the Hind III fragment pattern seen in DM1208. Only a recombination event between exons 2 from ZC376.2 and T02B5.A generated the observed pattern. This result was tested by PCR amplification of a fragment that carries both ZC376.2 and T02B5.A exon 2 sequences.

2. Materials and Methods

Seven oligonucleotides were synthesized as described in section A of Appendix A. MG22 is an antisense 20-mer in the intron between exons 4 and 5 of ZC376.2 with the sequence: 5'-ACTTTTCAGACCTGCCTCGA-3'. MG23 is a sense 20-mer in exon 1 of T02B5.A with the sequence: 5'-AATGCTCAAACCTGCYACCG-3'. MG24 is a sense 20-mer in exon 2 of T02B5.A with the sequence: 5'-CCACTGGGAAATCAATCTGC-3'. MG25 is a sense 20-mer in the intron between exons 4 and 5 of ZC376.2 with the sequence: 5'-GCTACTCCTTGTCTGAACGA-3'. MG26 is an anti-sense 20-mer in exon 4 of T02B5.A with the sequence: 5'-TTCTTGTCTCCTCGATCACC-3'. MG27 is an anti-sense 19-mer in

the intron between exons 4 and 5 of ZC376.2 with the sequence: 5'-CGAGCCAAAATTGACACAC-3'. MG28 is a sense 20-mer in the intron between exons 1 and 2 of T02B5.A with the sequence: 5'-CCTTCGCAACGTAAAACTAG-3'. All have a T_m of 60 °C except MG27 which has a T_m of 58 °C. MG22 was paired with either MG23, MG24, or MG25 and MG26 was paired with either MG23 or MG24. These oligo pairs were used in PCR to amplify fragments from N2 and *unc-112(r367ra202)* genomic DNA. The conditions used were as follows: 200 ng of genomic DNA, 20 nmole/μl of each dATP, dCTP, dGTP, and dTTP, 25 pmole/ μl of sense oligonucleotide, 25 pmole/ μl of anti-sense oligonucleotide, 1mM $MgCl_2$, 2.5 units Taq DNA polymerase (courtesy of J. McGhee) in 1X Taq buffer (Promega). Incubations were performed in a Perkin-Elmer Thermocycler (model) at 95 °C for 30 seconds, 53 °C for 1 minute, and 72 °C for 90 seconds for 35 cycles with a final incubation at 72 °C for 5 minutes. The PCR products were separated on a 2% agarose gel in 1X TBE electrophoresis buffer at 95 V, stained with EtBr, and photographed. The gel was soaked in 0.25 M HCl for 30 minutes, then in 1.5 M NaCl and 0.5 M NaOH for 40 minutes, followed by 1 M NH_4OAc for 1 hour. DNA was transferred to Nytran applied to both sides of the gel overnight. DNA was crosslinked to the Nytran by UV exposure for 3 minutes.

MG27 and MG28 were end-labeled using T4 Polynucleotide kinase and [γ–^{32}P]ATP of 5000 Ci/mmole (Amersham) according to Gibco-BRL instructions. The Nytran filters were prehybridized at 37 °C for 1 hour in 0.5 M NaPO4 pH 7.2, 7% SDS (22). Each probe was individually hybridized to a filter in fresh hybridization solution at 37 °C overnight. Filters were washed twice in 1X SSPE at 40 °C for 5 minutes each. The filters were then exposed to XAR film (Kodak) with an intensifying screen at -80 °C for 6 hours.

3. Results

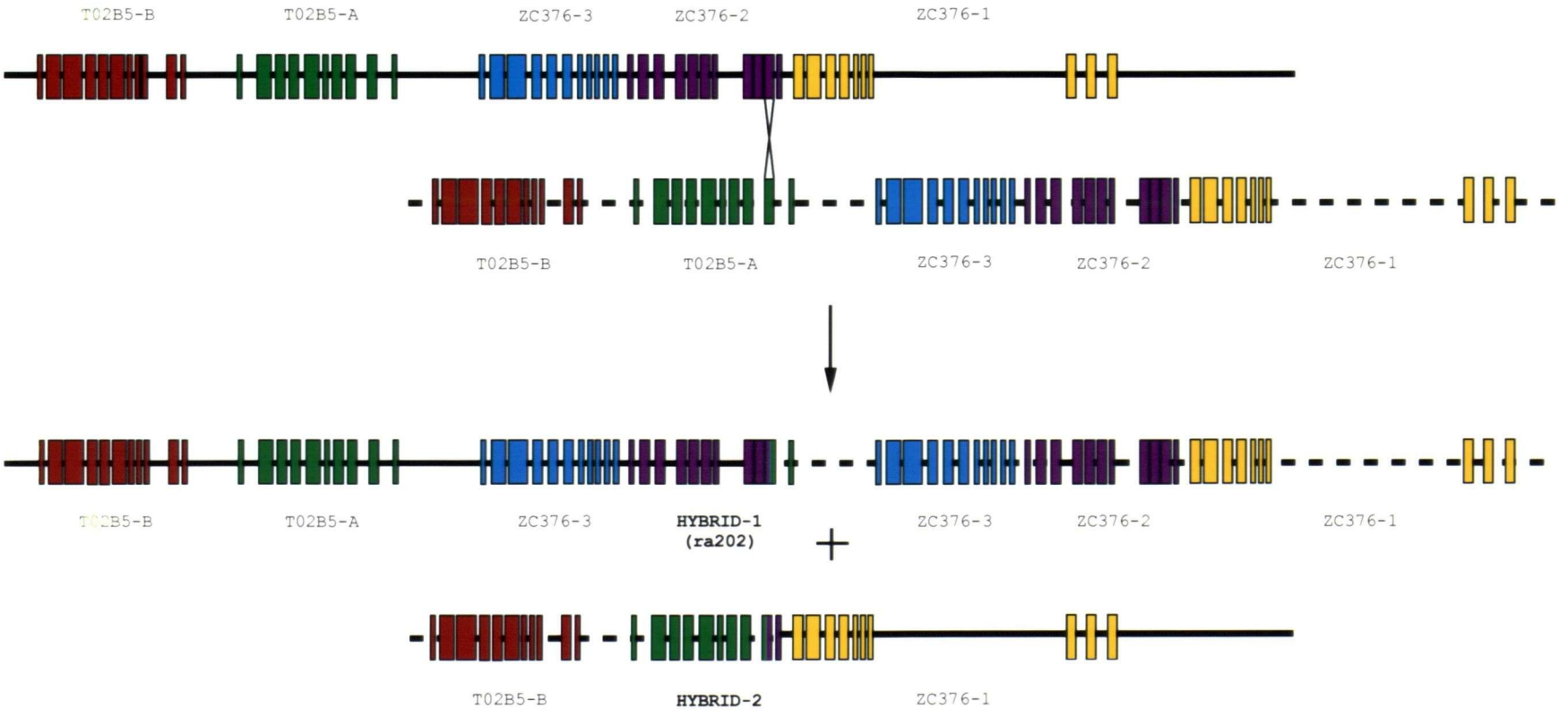
The genomic origin of the 2.5 kb T02B5 probe used to identify the polymorphism in the strain containing *unc-112(r367ra202)* was determined by correlating the EcoRI, HindIII and Kpn I fragment sizes with the predicted intron-exon structures of the five tandem genes (see figure 28). The 9.5 kb Kpn I fragment overlaps eight exons of ZC376.1 and nine exons of ZC376.2 including the large intron containing

another gene in the opposite orientation (ZC376.4 not shown). Within the 9.5 kb Kpn I fragment is a 2.8 kb EcoRI fragment that extends from a site halfway through the large 3/4 intron to a site within exon 6 of ZC376.1. If this is the region from which the probe is derived, then it must also be able to cross-hybridize with regions of similar sequence that generate HindIII bands of 3.5 kb, 4 kb, 4.7 kb, and 13.8 kb (see figure 21). The ZC376.1 intron appears to be unique, but the segment containing exons 4, 5 and part of 6 in ZC376.1 is approximately 73% identical to the analogous exons in the other four genes. This degree of identity would be sufficient to hybridize to HindIII bands of 3.5 kb, 4 kb, 4.7 kb and 13.8 kb containing the analogous exons (compare figure 21 with figure 28).

The polymorphism identified in the strain containing *unc-112(r367ra202)* using the cosmid subfragment generated a new HindIII fragment of approximately 2.5 kb in addition to all of the Wild type bands. The simplest explanation for this result is a gene duplication resulting from unequal recombination. The best sites for pairing and recombination are probably exons with a high percentage of sequence identity such as exons 2, 5 and 8 in all five genes. This was tested with a scale model of two sister chromatids variously aligned to create all combinations of exon 2 pairings and determine the sizes and numbers of HindIII fragments produced for both DNA strands of each pairing. The same was done to generate all possible products of exon 5 pairings and exon 8 pairings. Only an unequal recombination event of exon 2 between ZC376.2 and T02B5.A generated all Wild type HindIII band sizes plus a 2.4 kb band (see figure 32). This event yielded 7 genes, including a second copy of T02B5.A and a hybrid gene consisting of exon 1 and some portion of exon 2 of T02B5.A and the balance of exon 2 and exons 3-11 of ZC376.2.

Figure 32: Model of the unequal crossover event generating the polymorphism associated with the unc-112(r367ra202) mutation.

The sister chromatids are shown preparing to exchange strands at the X. One chromatid is shown with a solid bar connecting the exons of the five gene cluster and the second chromatid is shown with a dashed bar connecting the exons. The exchange of strands occurs at a point within exon 2 of ZC376.2 and exon 2 of T02B5.A generating the two hybrid chromatids shown below the arrow. Hybrid 1 contains a wildtype duplication of ZC376.3 and a duplication consisting of exon 1 and a portion of exon 2 from T02B5.A and exons 3-11 of ZC376.2 resulting in a total of seven genes in the tandem cluster. Hybrid 2 has lost wildtype copies of ZC376.2, ZC376.3 and T02B5.A. These genes have been replaced by a fusion gene consisting of exon 1 and a portion of exon 2 from ZC376.2 and exons 3-11 of T02B5.A. Only ZC376.1 and T02B5.B are unchanged.

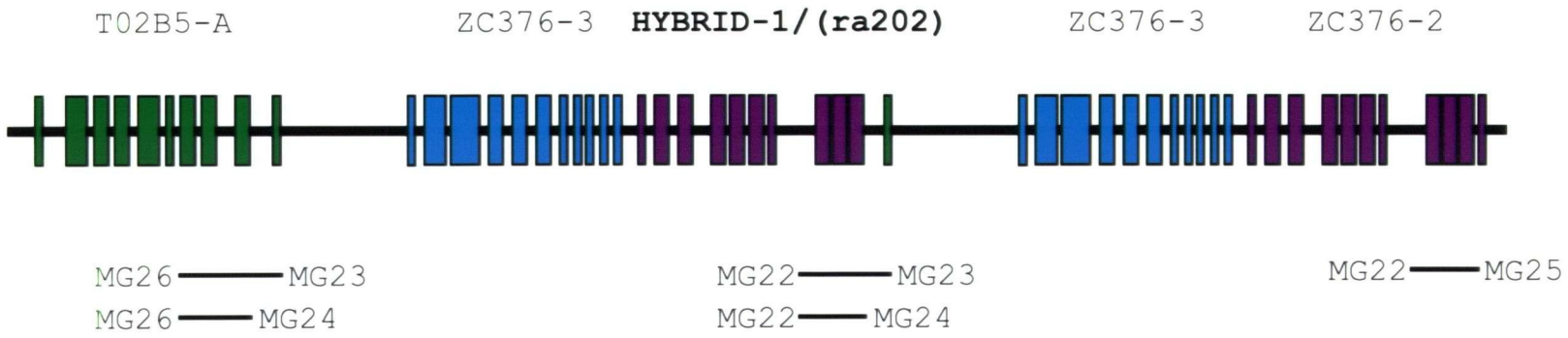

T02B5-B
T02B5-A
ZC376-3
ZC376-2
ZC376-1
T02B5-B
T02B5-A
ZC376-3
ZC376-2
ZC376-1
T02B5-B
T02B5-A
ZC376-3
HYBRID-1
(ra202)
ZC376-3
ZC376-2
ZC376-1
T02B5-B
HYBRID-2
ZC376-1

Oligonucleotides were synthesized to DNA flanking exons 2-4 in both ZC376.2 and T02B5.A to test the result of the thought experiment described above. Only oligos MG22, MG23 and MG24 should be capable of generating a PCR fragment if the unequal recombination event had occurred in the vicinity of exon 2 in both genes (see figure 33). Suprisingly, MG22 and MG24 did not generate a PCR product, whereas MG22 and MG23 gave a PCR product of 850 bp. This suggests that the recombination event occurred between MG23 and MG24 which replaced the T02B5.A-derived MG23-containing sequence with the corresponding region in ZC376.2.

Figure 33: Identification of the unequal crossover event generating the polymorphism associated with the unc-112(r367ra202) mutation.

Panel **A** shows the predicted organization of the fused ZC376.2 and T02B5.A gene and the flanking genes. Below the sites and names of the oligonucleotides used to amplify various fragments from the region. The oligo pair MG25 and MG22 produce a band of 700 bp, MG23 and MG26 produce a band of 1300 bp, and MG24 and MG26 produce a band of 800 bp as PCR controls for the normal ZC376.2 and T02B5.A genes. Only the oligo pairs of MG23 and MG22 or MG24 and MG22 are predicted to produce PCR products. Panel **B** shows the resulting PCR products on an ethidium bromide-stained agarose gel. The molecular weight markers are in 100 bp increments and the sizes are given on the left side. Control lanes containing reaction mixes with no DNA are indicated by "0" above the lanes, N2 DNA reactions are indicated by "+", and *unc-112(r367ra202)* containing DNA reactions are indicated by "M." Group one samples included the oligos MG25 and MG22, group 2 samples included MG23 and MG26, group 3 samples included MG24 and MG26, group 4 samples included MG23 and MG22, and group 5 samples included MG24 and MG22. Only the mutant DNA sample produced a PCR product with oligos MG23 and MG22 of 850 bp. The absence of a product in the mutant DNA sample in group 5 suggests that the region from ZC376.2 containing the MG24 binding site is either deleted or replaced by T02B5.A DNA.

A


T02B5-A
ZC376-3
HYBRID-1/(ra202)
ZC376-3
ZC376-2
MG26——MG23
MG26——MG24
MG22——MG23
MG22——MG24
MG22——MG25

B


1
2
3
4
5
0 + M
0 + M
0 + M
0 + M
0 + M
2072
1100
600
100

4. Discussion

The polymorphisms identified in *unc-112(r367ra202)* and *unc-112(r367ra207)* that enabled the isolation of the cDNA encoded by ZC376.2 were predicted to be due to unequal crossing over between exons with the greatest identity in all five genes based on the results (see figure 21). The result of testing all possible combinations of crossing over between pairs of exon 2s in all five genes indicates that this is very likely to be the site of recombination between ZC376.2 and T02B5.A. The event duplicates T02B5.A and generates a fusion gene of the first portion of ZC376.2 to the fusion site in exon 2 and the downstream portion of T02B5.A, thereby increasing the number of tandem genes from five to seven. The prediction of a fusion gene between ZC376.2 and T02B5.A from this thought experiment was tested using oligonucleotides flanking exon 2 in both genes. The 850 bp PCR product generated by an antisense oligo from T02B5.A and a sense oligo from ZC376.2 confirms the result of the thought experiment. The inability of a second ZC376.2 sense oligo to generate a PCR product in combination with the T02B5.A antisense oligo indicates that this site is missing in the fusion gene. This result may be due to a small deletion of this portion of the gene during recombination, or that T02B5.A sequence, which cannot hybridize with this oligo, extends to this site.

In summary, the polymorphism identified in *unc-112(r367ra202)* that enabled the isolation of the cDNA encoded by ZC376.2 was shown by PCR analysis to be due to unequal crossing over between the second exons of ZC376.2 and T02B5.A. The event duplicates T02B5.A, generates a fusion gene and increases the number of tandem genes from five to seven. The nature of the gene fusion might be determined in the future by sequencing the PCR product and comparing it with the genomic sequences.