AN ANALYSIS OF UNC-52/PERLECAN DOMAIN IV IMMUNOGLOBULIN REPEATS IN MYOFILAMENT ASSEMBLY OF *Caenorhabditis elegans*.

by

JASON ALLAN BUSH

B.Sc., Univeristy of British Columbia, 1992

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS OF THE DEGREE OF
MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Zoology)

We accept this thesis as conforming to the
required standard

THE UNIVERISTY OF BRITISH COLUMBIA

October 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 Oct 15/97

DE-6 (2/88)

ABSTRACT

The *unc-52(II)* gene encodes the nematode homologue of perlecan, a mammalian basement membrane (BM) proteoglycan. UNC-52 is essential for proper myofilament assembly and muscle attachment in *C. elegans*. The longest predicted UNC-52 protein has five structural domains, including a LDL-receptor-like domain (domain II), a laminin-like domain (domain III), an NCAM-like domain (domain IV) and a globular laminin-like domain (domain V). The *unc-52* gene consists of 37 exons which, through alternative splicing, generate a number of protein isoforms. These isoforms show both tissue and temporal specificity. The domain structure of *unc-52* lends itself to regulation via alternative splicing. In particular, domain IV contains multiple copies of an immunoglobulin-like repeat (Ig-repeat) and many of the repeats are encoded by individual exons. Alternative splicing of exons 16, 17 and 18 alters the number of Ig-repeats within this domain.

Null mutations in *unc-52* lead to a lethal Pat (paralyzed arrested elongation at two-fold) phenotype (e.g. *st549*), while mutations in exons 16-18 lead to a viable Unc phenotype. Through the sequence analysis of two new *unc-52* mutations, *st560* and *ra112*, I established that domain IV-containing isoforms of *unc-52* are critical for myofilament assembly during early development. These two mutations both lead to a Pat phenotype and result in premature truncation of UNC-52 isoforms with domain IV; *st560* is a stop codon in exon 13; and *ra112* is an deletion removing several Ig-encoding exons within domain IV.

Using a Tc1 excision and double-strand break repair scheme to vary the number of Ig repeats within an isoform, I addressed the functional significance of domain IV in establishing proper muscle assembly. I sought to determine the minimal number of Ig repeats necessary within domain IV to allow proper myofilament assembly and to determine the role of combinatorial Ig repeats in assembly. I characterized eight in-frame deletion alleles created by Tc1 excision which eliminate from 1 to 4 Ig repeats. Animals homozygous for these deletions appear wild-type in movement and muscle structure.

The *mec-8* gene encodes a putative RNA-binding protein that is required for some of the alternative splicing of *unc-52* (Lundquist *et al.*, 1996). In the absence of MEC-8 function, otherwise viable mutations in the alternatively-spliced region of *unc-52* are lethal because the affected exons are no longer spliced out of embryonic *unc-52* pre-mRNA (Lundquist *et al.*, 1996). I constructed several *mec-8; unc-52(deletion)* double mutants to ask whether MEC-8 function was still required in the absence of its splicing target. I found that *mec-8(null); unc-52(viable deletion)* combinations are phenotypically wild-type, suggesting a complete independence from *mec-8*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

BM	basement membrane
bp	basepairs
dpy	dumpy
ECM	extracellular matrix
kb	kilobase
LDL	low-density lipoprotein
mec	mechanosensory
mL	millilitre
mM	millimolar
mut	mutator
pat	paralyzed, arrested elongation at two-fold
PCR	polymerase chain reaction
ρmol	picamole
µL	microliter
unc	uncoordinated

ACKNOWLEDGEMENT

I would like to graciously thank my supervisor, Don Moerman, for his guidance, his financial support, and his constant intellectual stimulation. I thank Greg Mullen, Ken Norman and Poupak Rahmani for their scientific and literary assistance, being my friends, and partaking in necessary 'distractions'. I also thank Teresa Rogalski for her guidance and invaluable suggestions.

INTRODUCTION

I. *Extracellular Matrices*

Basement membranes (BM) are defined as specialized regions of the extracellular matrix (ECM)(Yurchenco & Schittny, 1990). These membranes form a dynamic network, acting as both a barrier and at times a zone for information transfer between two distinct tissue types. All extracellular matrices have the same basic components including collagens (type IV), laminins, nidogen and heparan sulfate proteoglycans (perlecan)(Timpl, 1993; Timpl and Brown, 1996). There is tremendous structural diversity within each of these components which leads to a qualitatively diverse array of matrices (Timpl and Brown, 1996). This diversity arises from the presence of multigene families for each of these components (van den Brule *et al*, 1995; Heikkila and Soininen, 1996) and/or alternative splicing of these ECM encoding genes (Noonan *et al*, 1991; Timpl, 1993; Lundquist *et al*, 1996). The BM is important as a structural substrate for the attachment of cells and also as the mediator of chemical gradients for the migration of cells. It appears to perform this latter function by acting as a molecular sieve for the diffusion of materials between tissue types (Inoue, 1994). The structural components of the matrix can also act as ligands via interactions with various transmembrane receptors (usually integrins) on cell surfaces (reviewed in Calderwood *et al*, 1997). In some cases, these interactions remain static and thus act as an anchor for stabilizing the attachment of cells, but in many cases the interaction may be more dynamic thus leading to signaling between adjacent tissues. Cellular responses to these extracellular signals

alters gene expression which can have profound effects on cellular behaviour including migratory properties, morphology, growth, and differentiation.

II. *Perlecan/UNC-52*

At present we know a great deal about the structural components of the ECM but much less is known about their functional roles. This situation is especially acute for proteoglycans. Proteoglycans are glycosylated proteins which have covalently attached highly anionic glycosaminoglycans (reviewed in Timpl, 1993). The major biological function of proteoglycans derives from the physicochemical characteristics of the glycosaminoglycan component of the molecule, which provides hydration and swelling pressure to the respective tissue enabling it to withstand compressional forces (Yanagishita, 1993). Other biological functions include control of filtration through BMs and binding or storage of growth factors and protease inhibitors (Timpl, 1993). The concerted action of proteases that degrade the protein core and heparanases that remove the heparan sulfate may modulate the bioavailability of bound growth factors. There are other proteoglycans found in various vertebrate tissues which are usually defined through the binding of different sugar side chains such decorin (Yamaguchi and Ruoslahti, 1988), aggrecan (Doege *et al*, 1990) and bamacan (Couchman *et al*, 1996).

Perlecan is recognized as the most abundant heparan sulfate proteoglycan in vertebrates. The gene for perlecan was first cloned and sequenced in mouse (Noonan *et al*, 1991) and later, a human (Murdoch *et al*, 1992; Kallunki & Tryggvason, 1992) and a *Caenorhabditis elegans* (Rogalski *et al*, 1993) homologue were characterized. All versions share an identical five-domain modular

structure with highest overall homology to well-known polypeptides. The first domain (domain I) is unique to this polypeptide and is the only region where the *C. elegans* homologue significantly differs from the other two; the second (domain II) is similar to the receptor domain of the LDL class of receptors; domain III shares homology with the short chains of laminin A; domain IV is composed of contiguous units (21 repeats in humans, 14 in mouse and *C. elegans*) most like the neural cell adhesion molecule family of immunoglobulin (Ig) repeats (Rogalski *et al*, 1993); and domain V is composed of laminin globular regions separated by short EGF-like repeats (Noonan *et al*, 1991; Murdoch *et al*, 1992; Kallunki & Tryggvason, 1992; Rogalski *et al*, 1993; T.M. Rogalski, G.P. Mullen and D.G. Moerman, unpub. results). In *C. elegans*, the gene that encodes perlecan is the *unc-52* locus. It is a large gene spanning over 25 kb, and is composed of 37 exons that potentially encodes an open reading frame of almost 3500 amino acid residues (Rogalski *et al*, 1993, 1995; T.M. Rogalski, G.P. Mullen and D.G. Moerman, unpub. results).

To better understand the role of perlecan in BM function, we have chosen the nematode *C. elegans* as our model system. By using a simple eukaryote, we can utilize genetics as a tool to manipulate cellular and extracellular components. Results gained through studying the molecular nature and functional roles of individual, or groups of components, in this simpler system may be applicable to more complex organisms. The great number of mutants in *C. elegans* and its well-characterized genome make it an ideal organism in which to study many biological processes (reviewed in Wood, 1988, Riddle, 1997). Research in our laboratory is aimed at determining the interactions between the BM and the precursors of the body wall

muscle that are necessary for the formation of a functional muscle quadrant in *C. elegans* (see Figure 1 for details). My specific project has been to further characterize the functional role of nematode perlecan in this developmental process (see below).

III. *Muscle Structure in C. elegans*

There are 95 striated body-wall muscle cells in the adult *C. elegans*. Only 81 cells are present at birth. The remaining cells are added early on during postembryonic development (Sulston and Horvitz, 1977; Sulston *et al*, 1983). These are arranged into four quadrants, two dorsal and two ventral. Each quadrant runs the length of the longitudinal axis and consists of a double row of spindle-shaped cells (Figure 2). The basic repeat unit within muscle that is responsible for contraction is the sarcomere (Figure 1; reviewed in Waterston, 1988; Moerman and Fire, 1997). There are 7-10 sarcomeres across the width of an adult muscle cell and these are easily visualized as A- or I-bands using polarized light microscopy. A sarcomere is composed of bundles of myosin-containing thick-filaments interdigitated with actin-containing thin filaments. Myosin is anchored by M-lines while actin is anchored by dense bodies. The myofilament lattice is anchored to the cell surface through a series of lateral attachments (focal adhesions) which tightly adheres the cell to the underlying BM and hypodermis as well as adjacent muscle cells (Francis and Waterston, 1985, 1991).

IV. *Focal Contacts*

In general, any disruptions in focal adhesion destroys the structural link between the cytoskeleton, ECM and substrate. These links are crucial since they are regions where signal transduction

Figure 1.

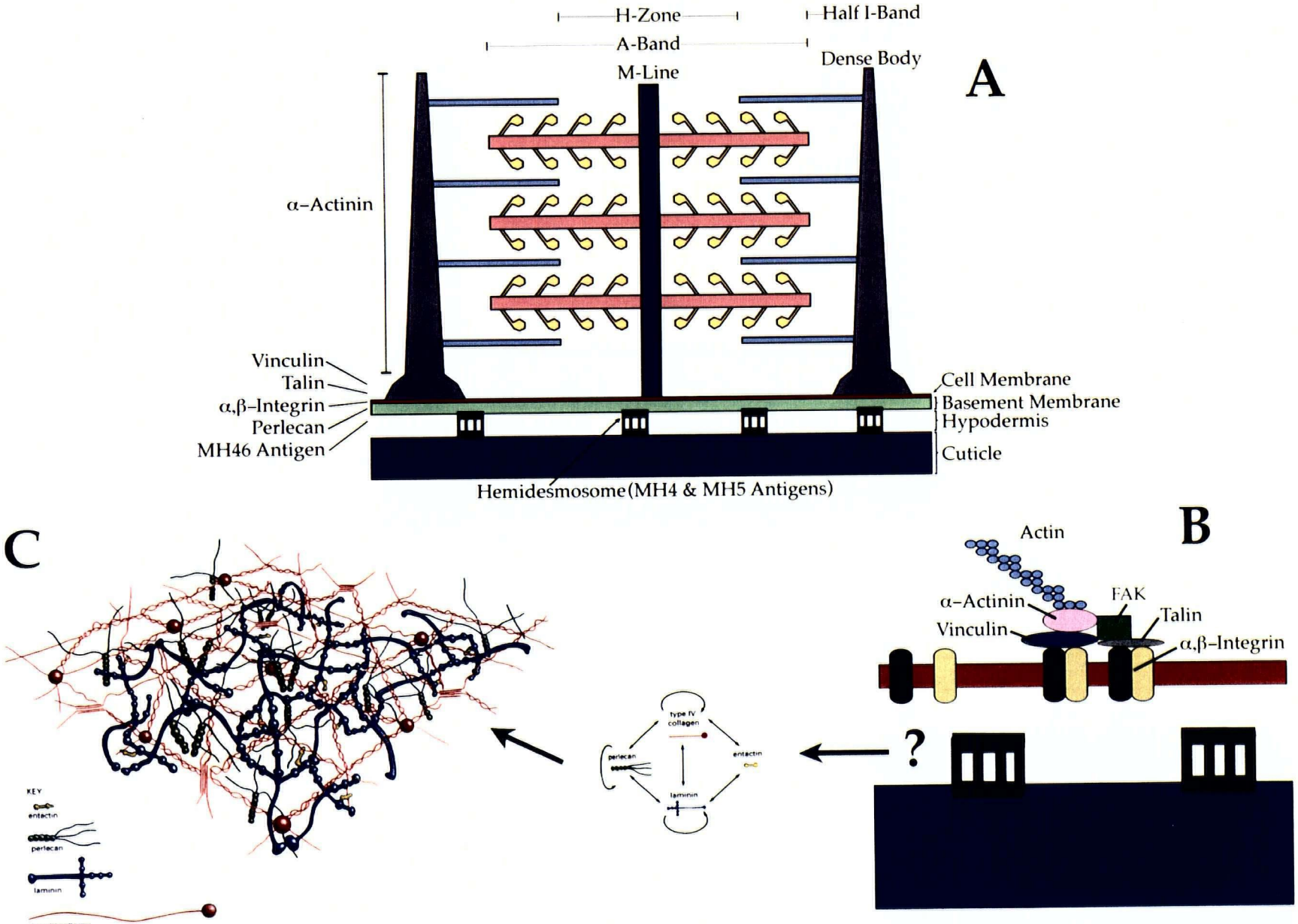
Structure of the sarcomere and basement membrane

Part A. A schematic representation of a sarcomere, the basic contractile unit of *C. elegans* muscle. The major structural features are indicated such as the actin-containing thin filaments (*blue*) which are anchored by dense bodies (*mauve*, analogs of vertebrate Z lines) and myosin-containing thick filaments (*light orange*) whose alignment is maintained by M-line components (*purple*). Some of the extracellular components are also shown. The *unc-52* gene products (*light green*), are found in the basement membranes underlying muscle cells and play an essential role in anchoring the dense bodies to the cuticle for lateral transmission of the forces generated within the sarcomere.

Part B. An enlargement of a portion of the basal surface of the muscle cell-BM-hypodermal junction. Many of the structural components of the focal adhesion complex at the membrane itself have been determined and their interactions documented (Burridge and Chrzanowska-Wodnicka, 1996) but the functional roles of the interacting proteins is not well-characterized. This is particularly obvious in the BM. As the transition diagram demonstrates, *in vitro* data show the possible interactions that could be occuring between the BM components but the nature of their assembly is unclear.

Part C. Diagrammatic representation of a possible BM architecture showing the putative layering of components (Adapted from Yurchenco and Schittny, 1990).

Figure 1. Structure of the sarcomere and basement membrane



relates to the assembly of cellular components. Through aggregation of integrin receptors, the constituents on the muscle cell membrane side of the BM which include vinculin, talin, α-actinin, and focal adhesion kinase (FAK) stimulate signal cascades which influences the formation of contractile proteins (Jockusch *et al*, 1995; Burridge and Chrzanowska-Wodnicka, 1996; Moerman and Fire, 1997). The BM network acts as a mediator of adhesion to the opposing side as well. Mechanical stability here is conferred through molecularly distinct junctions called hemidesmosomes which are on the hypodermal side of the BM (see Figure 1). The exoskeleton of *C. elegans* therefore has a direct linkage to the muscle, mediated by the ECM.

V. *Muscle Development*

The development of *C. elegans* body-wall muscle is well characterized (Sulston *et al*, 1983; reviewed in Moerman & Fire, 1997). Muscle cells are not simply clonally derived in most cases. Several of the early blastomeres give rise to muscle and other specialized cell types which implies a rather elaborate pattern of commitment and differentiation (Sulston *et al*, 1983). Regardless of blastomere derivation, all body wall muscle cells are eventually oriented within the developing embryo along the lateral sides of the embryo adjacent to the hypodermis at around 290 minutes after first cleavage (Figure 3). This is the earliest time at which structural proteins of muscle can be detected (Epstein *et al*, 1993; Hresko *et al*, 1994; Moerman *et al*, 1996). Most cells are postmitotic at this stage and the muscle components are only diffusely distributed within the cell (Hresko *et al*, 1994). They then migrate along the hypodermis to their final dorsal or ventral position and form

Figure 2.

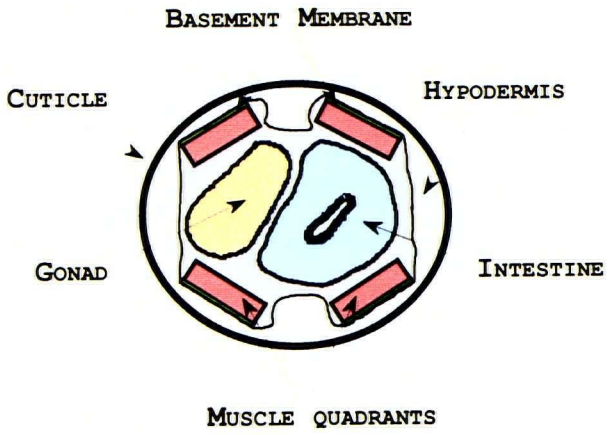
Anatomical structure of the worm

Part A. Diagram of a posterior cross-section through the adult hermaphrodite showing major anatomical structures. Note that there are four muscle quadrants (red) with an adjacent basement membrane (BM, green) juxtaposed to the hypodermis.

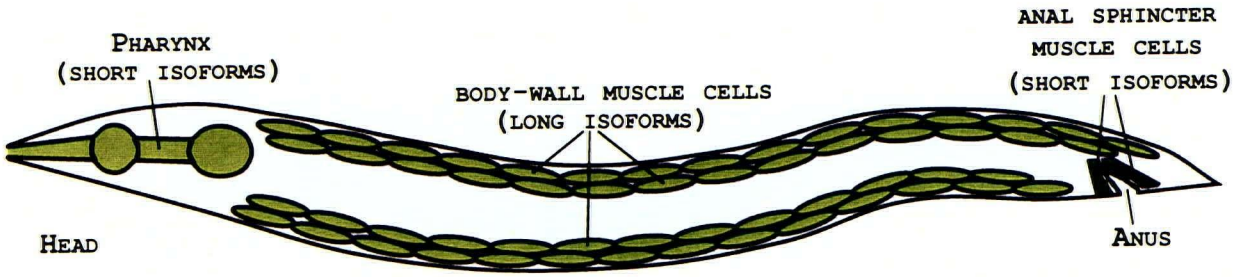
Part B. Lateral view of a canonical worm showing only two quadrants composed of a double row of muscle cells. Cells are indicated in green to represent UNC-52/Perlecan in the adjacent BM. Pharynx and anal sphincters contain short isoforms of UNC-52 while the body wall muscles contain long isoforms.

Figure 2.
Anatomical structure of the worm

A



B



quadrants (Figure 3; Hresko, 1994; Moerman *et al*, 1996; Schnabel *et al*, 1997). Myofilament components then become localized to the cell membranes, presumably through signal cascades and subsequent organellar arrangement, where they form focal contacts with the underlying components of the BM and hypodermis which are also localized to these regions. At this stage (approximately 350 minutes), we see the secretion from muscle of the proteoglycan, perlecan, that is encoded by the *unc-52* locus (Figure 3; Moerman *et al*, 1996).

Genetic analysis reveals that nematode perlecan is essential for the proper assembly of the highly ordered myofilament lattice within developing body wall muscle cells (Rogalski *et al*, 1993; Hresko *et al*, 1994; Williams and Waterston, 1994). After being localized outside the muscle cell, UNC-52/perlecan functions as part of an adhesion complex to anchor muscle (Moerman *et al*, 1996). It is hypothesized that it may interact with receptor molecules, possibly the integrin complex, on the surface of muscle cells. Support for this model comes from several lines of evidence: 1) the co-localization of immunostaining of both integrin and perlecan over the dense bodies and M-lines of muscle cells (Francis and Waterston, 1985, 1991; G.P. Mullen and D.G. Moerman, unpub. results; 2) the genetic hierarchy (epistasis) established through the construction of double mutants of integrin and perlecan genes (Williams and Waterston, 1994; G.P. Mullen and D.G. Moerman, unpub. results); 3) the *in vitro* studies done on mammalian homologs demonstrating that RGD sequences of perlecan bind specific integrin subunits (Hayashi *et al*, 1992; Chakravarti *et al*, 1995; Ruoslahti, 1996). This extracellular attachment is essential to transmit the muscular contractions generated by the body-wall muscles to the hypodermis

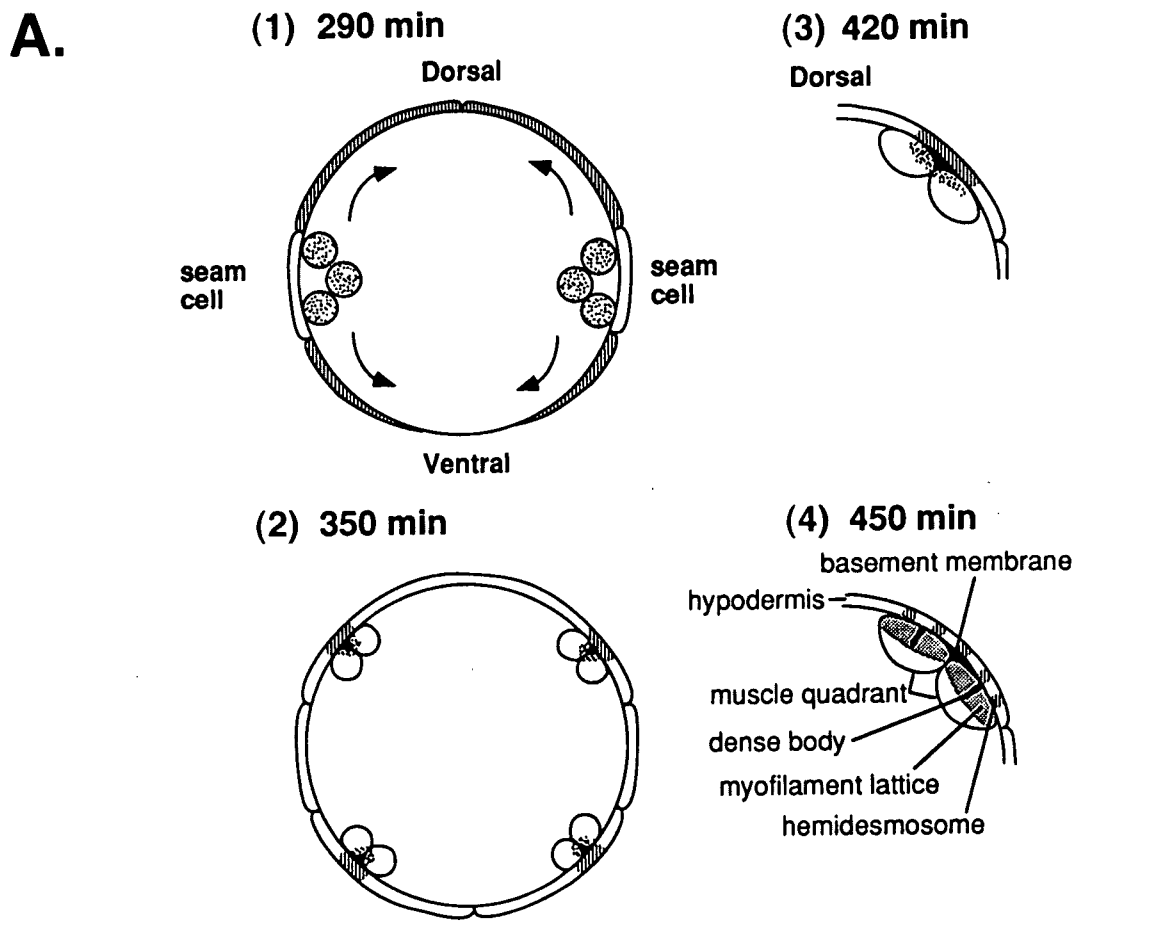
Figure 3.

Development of *C. elegans* and its muscle

Part A. Panels *1-4* summarize the process of myofilament assembly in *C. elegans*, depicting cross sections of embryos at various stages of development. The diagram is adapted from data presented in Hresko et al.(1994) and reviewed in Moerman (1997). (*1*) An embryo at 290 min. after first cleavage. Muscle cells (*circles*) have begun to accumulate myofilament components (*dots*). Muscle cells lie adjacent to the lateral hypodermis at this stage but are beginning their migration to final positions beneath the dorsal or ventral hypodermis. The hypodermis is a thin layer of cells covering most of the embryo. Components destined for hemidesmosomes are present in the hypodermis (represented by *hatched regions*). (*2*) 350 min. embryo. Muscle cells are asymmetric because myofibrillar constituents (*dots*) are localized to membrane regions adjacent to other muscle cells and the hypodermis. Both basement membrane and hemidesmosomal components have become concentrated at these focal regions. (*3*) Dorsal muscle quadrant of 420 min. embryo. Muscle cells (now *oval*) have become flattened and myofibrils (*dots*), basement membrane (*black*) and hemidesmosome components (*hatched region*) are coextensive. (*4*) Dorsal muscle quadrant of 450 min. embryo shows the complete organization of the myofilament lattice and its extracellular anchorage.

Part B. Embryogenesis in wild-type and Pat mutants (1) Movement generated by contractions of body-wall muscle occur when embryos reach the 1.5-fold length (2) In Pat mutants, 1.5-fold embryos fail to start moving and remain paralyzed. Elongation continues only until 2-fold stage then arrests. Embryos often hatch and die as misshapen larvae.

Figure 3.
Development of C. elegans and its muscle



and cuticle in order to facilitate movement. Since UNC-52 is produced by all the body-wall muscle cells (Francis and Waterston, 1991; Rogalski *et al*, 1993; Moerman et al, 1996), without a functional gene product, muscle structure is disorganized to the extent that embryonic death occurs.

VI. *Mutants and Isoform Diversity*

Genetic studies of *unc-52* have defined two broad classes of mutant alleles based on their phenotypes and patterns of complementation (Brenner, 1974, Mackenzie *et al*, 1978; Waterston *et al*, 1980, Gilchrist and Moerman, 1992, Rogalski *et al*, 1993, Williams and Waterston, 1994). One class contains viable uncoordinated (Unc) animals while the other contains embryonic lethal animals. The viable class of mutants are characterized by their paralyzed phenotype which becomes apparent in later larval stages (Mackenzie *et al*, 1978; Gilchrist & Moerman, 1992). As adults, the mutants are completely paralyzed except for some movement in the head region. All mutants of this class have point mutations within a small interval containing three adjacent, alternatively spliced exons (exons 16, 17, 18; Rogalski *et al*, 1993, 1995). Each of these exons encodes a single Ig repeat (Rogalski *et al*, 1993). The effect of these molecular lesions is to eliminate some of the long UNC-52 isoforms produced in growing and adult animals.

The second class of *unc-52* alleles consists of embryonic lethals, all of which share a Pat phenotype (paralyzed arrested elongation at two-fold; Williams & Waterston, 1994). Prior to the results described in this thesis, only two Pat alleles were

identified molecularly, *unc-52(st549)* and *unc-52(ut111)*. The *ut111* alteration results from the insertion of an endogenous transposable element (called Tc1) into exon 2 and *st549* is a nonsense mutation leading to a termination codon in exon 7 (Figure 4; Rogalski *et al*, 1995). Since these exons are expressed in all UNC-52 isoforms, these mutations lead to little or no expression of the *unc-52* gene product. In *ut111* homozygous animals, excision of Tc1 can occur at either the DNA or RNA level and leads to weak expression of UNC-52 and a somewhat later lethal phenotype than what is observed in *st549* animals (Rogalski *et al*, 1995). The Pat phenotype of *st549* animals defines the null phenotype of this locus since *st549* mutants lack immunoreactivity to antibodies that recognize all isoforms of UNC-52 (Rogalski *et al*, 1993; G.P. Mullen, unpub. results). In addition to *st549* and *ut111*, six other Pat alleles of this locus have been isolated (*ra112*, *ra401*, *st546*, *st560*, *st572*, *st578*--see Results). Identifying the molecular lesions in these *unc-52* Pat alleles has been a primary goal of this thesis.

Determining the location within *unc-52* and thus the biochemical basis for these mutant phenotypes has been challenging because of the complex patterns of alternative splicing that gives rise to a number of protein isoforms of *unc-52* (Rogalski *et al*, 1993, 1995; Lundquist *et al*, 1996). One of the tools at our disposal to define the region where these alterations may lie is the various UNC-52-specific antisera. Immunostaining has revealed that the different UNC-52 isoforms that are produced have distinct spatio-temporal locations (G.P. Mullen and D.G. Moerman, pers. comm.). There are three groups of *unc-52* products based on the presence or absence of domain IV/V (Rogalski *et al*, 1993; T.R. Rogalski, G.P. Mullen and D.G. Moerman, unpub. results). These isoforms are 1) a domain I,

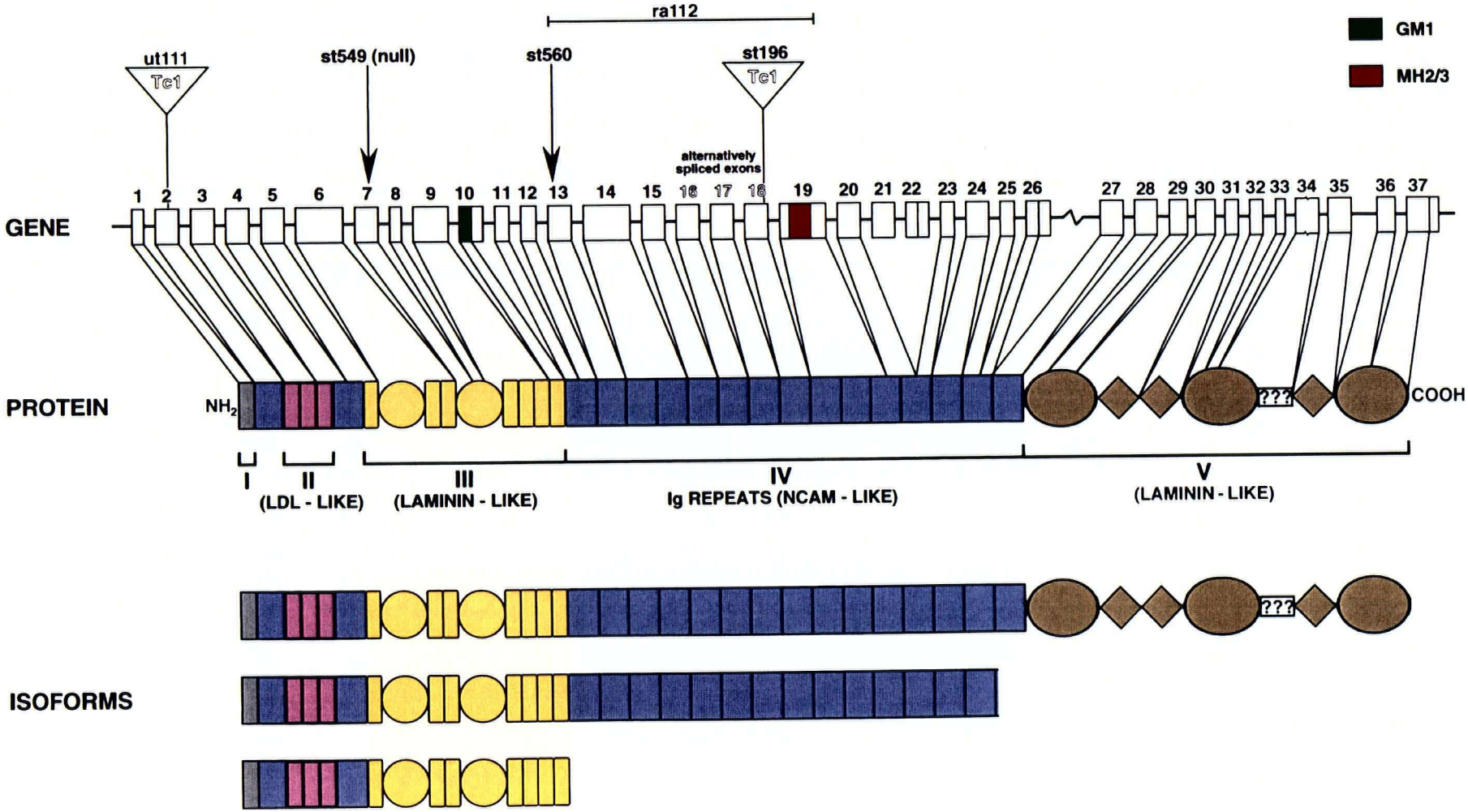
Figure 4.

unc-52 gene schematic

Representation of the complete unc-52 gene showing the exon (*boxes*) and intron (*line*) structure as well as the domain structure and 3 possible protein products. The gene contains 37 exons which span over 25 kb. There are at least 4 alternative poly(A)-addition sites (colored boxes). Several exons are alternatively-spliced. To date, only data demonstrating that exons 6, 16, 17 and 18 are involved in splicing exists. Sequenced Pat alleles including the null st549, and the domain IV knockouts, st560 and ra112 are shown. The longest ORF encodes a protein of approximately 3500 residues that can be divided into 5 domains (Domains I-V) based on homologies (protein motifs are as described in the text). The epitopes for antibodies GM1 (green) and MH2/3 (red) are indicated and correspond to sequences in exon 10 and exon 19, respectively. The GM1 epitope is common to all UNC-52 isoforms while MH2/3 recognizes only the long isoforms containing Domain IV/V. Representatives of both the short and long isoforms are indicated.

Figure 4.

unc-52 gene schematic



II and III form; 2) a domain I, II, III and IV form; 3) a domain I, II, III, IV and V form. Because the domain V-containing isoform has only recently been discovered, it is not included in this study. Rabbit polyclonal antisera was generated to recognize specific epitopes of UNC-52 in order to distinguish the domain I, II, III isoforms from the domain I, II, III, IV isoforms. We refer to these as the short and long isoforms, respectively. As indicated, the short isoform lacks domain IV and includes only the first three domains: in embryos this isoform is localized to the BMs underlying the pharyngeal and anal sphincter/depressor muscles (G.P. Mullen and D.G. Moerman, unpub. results). In contrast, domain IV-containing long isoforms are not present in the pharyngeal or the anal muscles at any stage of embryogenesis. Instead, these long variants are restricted to the BMs underlying the body wall muscle cells during embryonic development (G.P. Mullen and D.G. Moerman, unpub. results; see Figure 5 for details), only later are they co-expressed in tissues along with short isoforms (i.e. pharynx and anal muscles). In larvae and adults, additional tissues accumulate long isoforms of UNC-52. These include the uterine and vulval muscles in the hermaphrodite; the myoepithelial sheath of the gonad; and the male-specific muscles in the male tail (Francis and Waterston, 1991; G.P. Mullen, T.M. Rogalski and D.G. Moerman, unpub. results). Recently, the *mec-8* gene has been shown to encode a putative RNA-binding protein that regulates some of these splicing events at *unc-52* (Lundquist *et al*, 1996). The genetic analysis of *unc-52* and its interaction with genes such as *mec-8* has provided insight into the importance of this alternative-splicing for generating isoform diversity within a developmental context.

Figure 5.

Immunofluorescence staining of wild-type and mutant embryos with antibodies to UNC-52 and Myosin Heavy Chain A

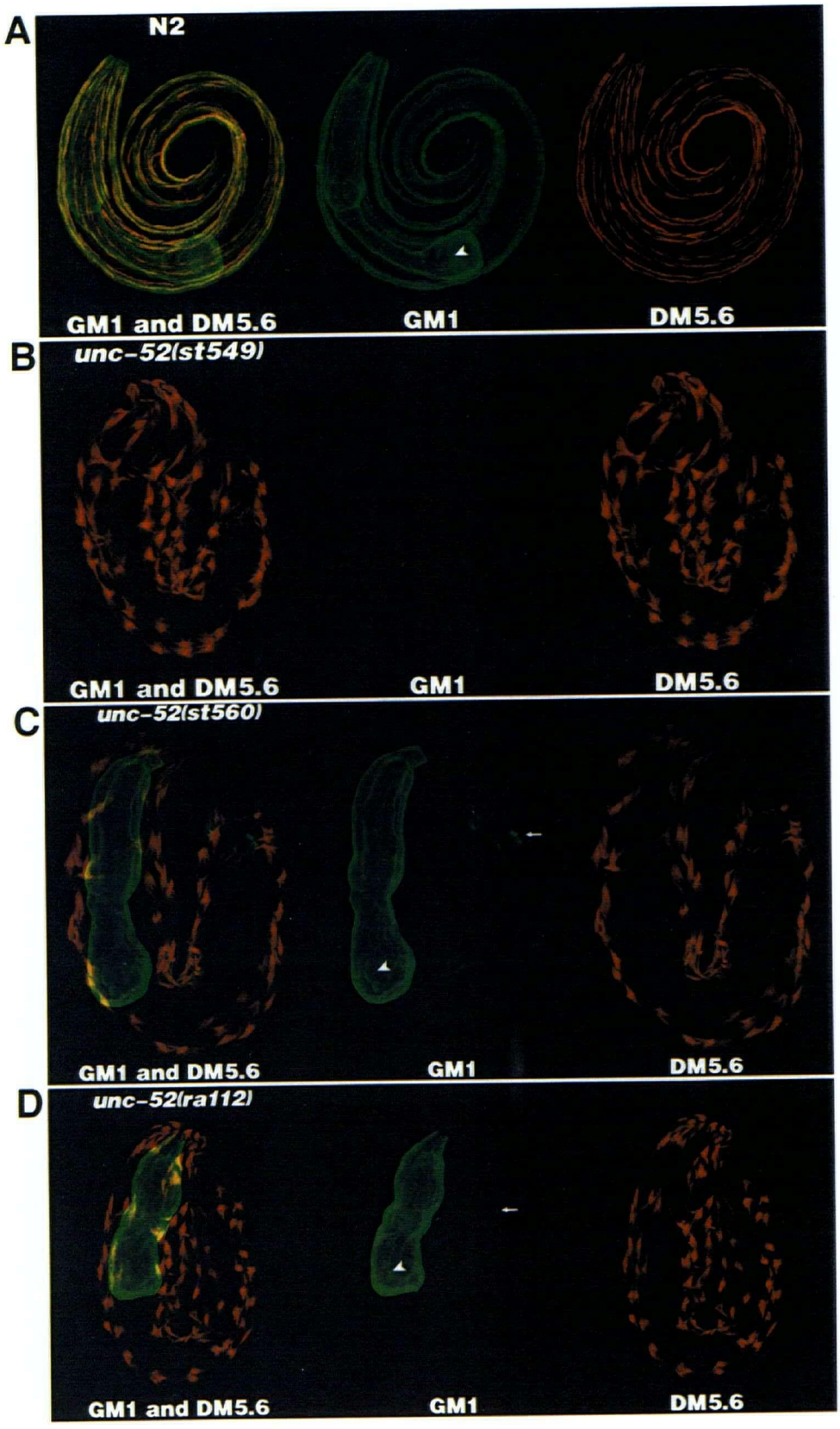
Panel A. Wild-type embryos double-labeled with GM1, which recognizes a region common to all *unc-52* gene products, and DM5.6, which recognizes the minor body-wall muscle myosin (MHC A). GM1 is shown in green (FITC) and DM5.6 is shown in red (TRSC). The left panel shows both antibodies simultaneously, while the middle and right panels show GM1 and DM5.6 alone. GM1 stains the basement membranes underlying the body-wall muscles, as well as the pharynx and anal sphincter/depressor muscles. White arrowhead marks the posterior bulb of the pharynx.

Panel B. *st549*, the putative null allele, eliminates all GM1 staining and has severely disorganized myosin.

Panel C. *st560*, a Domain IV-specific lethal allele, is a premature stop transition (C --> T)in exon 13. White arrowhead marks the posterior bulb of the pharynx. Small white arrow points to anal sphincter muscles.

Panel D. *ra112*, a lethal deletion eliminating 3283 bp of Domain IV. White arrowhead marks the posterior bulb of the pharynx. Small white arrow points to anal sphincter muscles.

Figure 5. Immunofluorescence staining of wild-type and mutants



VII. *Model and Strategies*

We speculate that extensive isoform diveristy may reflect a requirement for particular isoforms needed at different times or in different tissues. As described above, antibodies to UNC-52 demonstrate that isoforms are differentially distributed in space and time. Why this is required is not clear, but the different muscles involved do experience different mechanical stress. Perhaps different isoforms of UNC-52 have different adhering properties or biomechanical strength. Our phenotypic characterization of the lethal class of mutants (see Results) has led us to specualte that domain IV-containing isoforms are essential for myofilament formation and stability in body wall muscles. Sequence analysis of two of the mutants confirms that the primary defect in these animals is the loss of long UNC-52 isoforms while retaining the short isoforms (see Results for further explanation of domain V). This corroborates the importance of these long isoforms in muscle development. Domain IV consists of 14 Ig repeats spread over 16 exons. Analyses of alternative splicing of *unc-52* messages has shown that much of the diversity in splicing centers on a limited region of domain IV (Rogalski *et al*, 1995). To attempt to determine the role of this splicing, we explored the following questions: Are individual Igs within this group of particular importance during muscle formation, or do the Igs simply act as structural spacers to determine the length of the protein, or is there some other reason for this large repeat domain? Strategies were developed to answer these questions.

A key observation that bears on the relative importance of Ig repeats in domain IV was made by Erin Gilchrist in her studies of intragenic revertants of *unc-52* viable alleles (Gilchrist and

Moerman, 1992). It was found that exon-skipping in the domain IV region of exons 15 through 19 could lead to suppression of mutations within this region (Gilchrist and Moerman, 1992; Rogalski *et al*, 1995). Also it was determined that an intragenic revertant, *unc-52(ra38)*, has a small deletion removing the equivalent of one complete exon and is virtually wild-type (Gilchrist & Moerman, 1992; Rogalski *et al*, 1995). These results demonstrate that individual Ig units from this region can be removed without adversely affecting muscle. We extended these experiments to determine whether intact muscle could be maintained after the removal of multiple Ig units.

A Tc1 excision strategy was utilized to isolate several *unc-52* deletion revertants and then a subset of the revertants with altered domain IV regions were used to examine the importance of Ig copy number and specific Ig elements for UNC-52 function. The *unc-52(st196::Tc1)* allele has a Tc1 insertion in exon 18 (Figure 4) and reverts at a high frequency in a mutator background. Upon excision, Tc1 leaves a double-strand break that must be repaired. These free ends can be immediately repaired by the cell's DNA repair machinery or they may be the target of exonuclease activity (Engels *et al*, 1990; Gloor *et al*, 1991; Moerman *et al*, 1991). A free end may be involved in strand invasion which will provide a DNA template for repair. Depending on whether repair is completed or interrupted at some point the site can be returned to wild type or may contain a rearrangement in the form of a small duplication or deficiency (Engels *et al*, 1990; Gloor *et al*, 1991; Moerman *et al*, 1991; this process is diagrammed in Figure 6). From an analysis of several *unc-52* deletion revertants, it was determined that as many as four consecutive Ig repeats can be removed without any adverse effect on muscle development in this organism. This suggests that after a

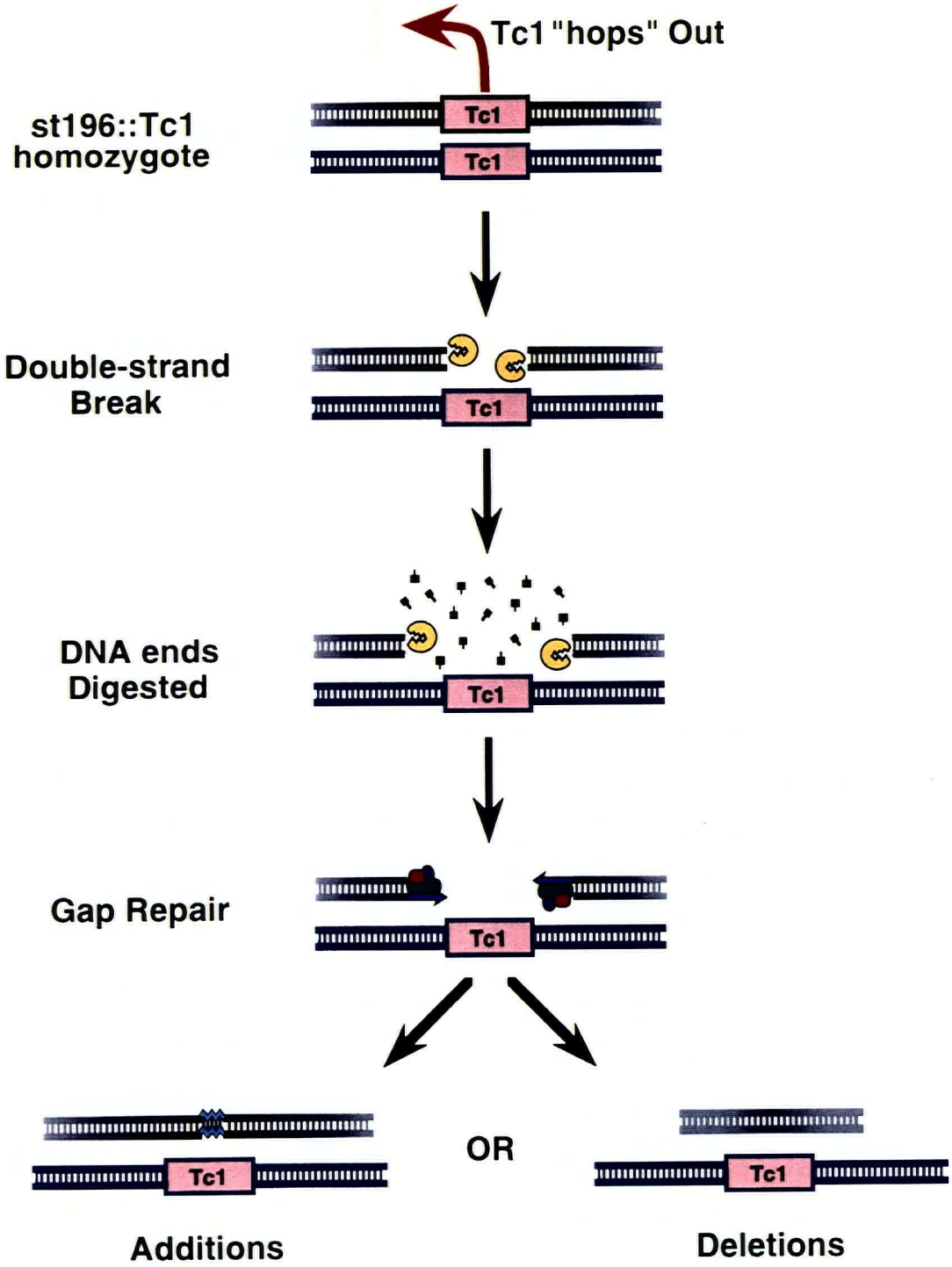
Figure 6.

Tc1 excision can create novel insertion/deletion mutations in a gene

In a mutator background, the Tc1 element in *unc-52(st196::Tc1)* can excise from its insertion site, at a relatively high frequency, creating double-strand gaps. During the repair process, additions or deletions of DNA can occur.

Figure 6.

Tc1 excision scheme



certain minimal number, Ig copy number is not of primary importance and that Ig repeats within the exon 16-18 interval are non-essential.

VIII. *Goals and Summary*

The goal of this thesis has been to examine in more detail the role of domain IV of perlecan in myofilament assembly and muscle attachment. This has been pursued through a combination of molecular and genetic techniques. Specific questions to be addressed include, 1) Is domain IV critical for body wall muscle development? 2) Are certain Ig units more critical than others? and 3) Is overall length of domain IV a crucial factor? These questions are addressed by the analysis of several mutations within the *unc-52* locus that have specific alterations. By identifying two Pat mutations within domain IV, the essential importance of long isoforms of this protein for body wall muscle function has been demonstrated. The various domain IV deletion revertants that have been characterized, illustrate the inherent structural plasticity of nematode perlecan and may reflect the redundancy inherent in all large ECM proteins. In summary, this study of domain IV alterations gives further insight into the molecular function of UNC-52 and tests the hypothesis that domain IV is essential during embryonic muscle development.

MATERIALS & METHODS

I. *Nematode strains, maintenance, genetic constructs*

We have followed standard *C. elegans* genetic nomenclature (Horvitz *et al.*, 1979). Standard laboratory techniques were employed for the handling of nematode stocks (Brenner, 1974). Unless otherwise noted, genetic experiments were performed at 20°C. The stocks used were the wild-type strain N2; RW6011, *unc-52(st546)/mnDp34II*; RW6013, *unc-52(st560)/mnDp34II*; RW6014, *unc-52(st572)/mnDp34II*; DM5105, *unc-52(st578)/mnDp34II*; DM3102, *mut-4(st700)I/unc-52(st196::Tc1)II*; TU74, *mec-8(u74)I*; DM4301, *mut-4(st700)I; unc-52(st196::Tc1)II*, DM4411, *unc-52(ra511)II*; DM4412, *unc-52(ra512)II*; DM4413, *unc-52(ra513)II*; DM4414, *unc-52(ra514)II*; DM4415, *unc-52(ra515)II*; DM4416, *unc-52(ra516)II*; DM4417, *unc-52(ra517)II*; DM4418, *unc-52(ra518)II*; DM4419, *unc-52(ra519)II*. Some strains were provided courtesy of the CGC.

II. *PCR amplification of genomic DNA*

Standard PCR reactions were performed as described by Barstead *et al.* (1991) with the following modifications. Either 5-6 embryos (in the case of lethals) or 2 adult hermaphrodites were placed in 3 µL of lysis buffer (50 mM KCl, 10 mM Tris (pH 8.0), 2.5 mM $MgCl_2$, 0.45% Tween-20, 0.45% NP-40, 60 µg/ml Proteinase K) in the lid of a 0.5 ml eppendorf tube with 15 µL of mineral oil. These tubes were briefly spun in a microcentrifuge and then incubated in a thermocycler for 30 min. at 37°C, then 5 min. at 95°C to inactivate the Proteinase K. These lysates were then used directly for

standard PCR (as described in Rogalski *et al*, 1993, 1995) or long-range PCR which differs only slightly (Nielson *et al*, 1994). Each method includes per reaction 10 mM dNTPs, 25 ρmol forward primer, 25 ρmol reverse primer. Long range PCR used 2.5 μL low-salt 10X buffer (200 mM Tris-HCl, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1% Triton X-100, 1 mg/mL BSA, Stratagene), and the corresponding proprietary TaqPlus Polymerase (Stratagene) in the amount of 3-5 units (depending on the size of product being amplified) and dH_2O to 22 μL. The standard PCR method included 25 mM $MgCl_2$ and 2.5 μL standard 10X reaction buffer, and ~2.5 units Taq Polymerase (Gibco-BRL). All PCR mixtures were amplified in a Perkin-Elmer-Cetus 480 thermocycler using the following conditions of 30 sec at 95°C, 30 sec at annealing temperature (53-57 °C) and 1 to 5 min. at polymerization temperature of 72 °C for 30 cycles followed by 5 min. at 72 °C. We were successful in generating genomic fragments as large as 5 kb in size. The following primer sets were used (Table 1):

III. *PCR sequencing*

The PCR-amplified genomic DNA fragments were directly sequenced using *unc-52* primers and the BRL dsDNA Cycle Sequencing System as described in Rogalski *et al*. (1993, 1995). Briefly, sequencing reaction mixes consisted of 1-2 μL from PCR-amplified genomic DNA reaction, 4.5 μL 10X Taq Sequencing buffer (300 mM Tris-HCl, 50 mM $MgCl_2$, 300 mM KCl, 0.5% w/v W-1 buffer, BRL), 0.5 μL Taq

Table 1.

Primer sets used in amplification of genomic DNA, the approximate positions along the gene, the size of product amplified and the exons which are flanked

Forward primer	Position	Reverse primer	Position	Product size (base pairs)	Exons contained within product
GCGTGTGTCTCAATGTTTCC	7098 -7117	CCGCTTGAATTTTGGCTAGC	8396 -8377	1298	exon 11
TCCGTCAGACCGTTGATC	9337 -9354	AACGGTCGCTGGGGTTGAGT	10599-10580	1262	exon 14, 15, 16
ACCCAATTTGTGGTGTGTGC	7994 -8013	TCACAGCAATAACACCCACG	9161 -9142	1167	exon 12, 13
CGGTAATCGATCCACCACAT	8882 -8901	GGAAGTGGTCTTCCGTCCT	9733 -9715	851	exon 13, 14
TCCGTCAGACCGTTGATC	9337 -9354	ATGTACTCATCCTCGTCCGA	10131-10112	794	exon 14, 15
TCAGACTGGAAGTCACTGAG	10820-10837	CGAGGATAGACACAGTAGCA	12182-12163	738	exon 19
AAGTCGAATGGCTCCATGAC	11964-11983	GGTCGATTCCAACGTGCAGT	13268-13249	1304	exon 19, 20, 21
CGGAACATGTGAACCAGGAT	11327-11346	AGTGACTTCCAGTCTGACTG	11460-11441	133	insertion in exon 18 at Tc1 site
GACCGTTCATGTTACCAACG	10593-10612	AGTGACTTCCAGTCTGACTG	11460-11441	1867	insertion in exon 18 at Tc1 site
GTGCAGCTGAACATCAGAGA	9564 -9583	CGAGGATAGACACAGTAGCA	12182-12163	2619	insertion in exon 18 at Tc1 site

Polymerase (BRL), 5 μL ^{32}P-end-labeled primer (1 ρmol) and H_2O to 36 μL. Four 0.5 mL eppendorf tubes were labeled A, C, G, or T and 2.0 μL of the appropriate termination mix (2 mM each, BRL) containing the dideoxynucleotides (2 mM each, BRL) and 8 μL of the above sequencing reaction were added to each tube and overlaid with a drop of silicone oil. The sequencing reactions were performed in a Perkin-Elmer-Cetus DNA thermocycler with the following conditions: 30 sec at 95°C, 30 sec at 55 °C, 60 sec at 70 °C for 20 cycles and then 30 sec at 95 °C and 60 sec at 70 °C for 10 cycles. At the end of the 30 cycles 5 μL of stop solution (BRL) was added to each tube and these reactions were boiled for 5 min. and then placed on ice for 5 min. before being loaded onto a standard 6% polyacrylamide sequencing gel.

IV. *Isolation of deletion revertants*

To isolate Tc1 excision events from *unc-52(st196::Tc1)*, we first crossed this allele into a mutator background. Wild-type (N2) males were crossed to *dpy-5(e61); unc-52(st196)* hermaphrodites. Outcross male progeny (*dpy-5/+; unc-52/+*) were then mated to *mut-4(st700)* hermaphrodites. Wild-type hermaphrodite progeny were picked singly onto new plates and allowed to have self progeny. Plates were then scored for the presence of both Dpy and Unc progeny, indicating that the parental genotype was *dpy/mut-4; unc-52/+*. Unc non-Dpy progeny were then picked singly onto new plates and allowed to self. Unc animals that failed to segregate Dpy Unc-52 progeny were expected to have the genotype

mut-4(st700);unc-52(st196::Tc1). Several independent isolates were maintained and tested for mutator activity by screening for spontaneous reversion of the Unc phenotype. Several independent lines were established that reverted at a high frequency (~1 x 10^{-3}). These were maintained by picking single Unc animals to new plates and monitoring these for reversion of the Unc phenotype. Any plates that failed to segregate revertants were discarded.

Revertants were identified on the basis of their improved movement and larger body size compared with their Unc siblings. Revertants were picked singly to new plates and maintained several generations, until the revertant allele was homozygous (usually 4-6 generations before being assayed by PCR). We then used PCR to determine whether these revertants had detectable polymorphisms in the exon 18 region of *unc-52*. Candidates that appeared to have a polymorphism were maintained for further analysis, including sequencing.

V. *Lethal deletion revertants*

To isolate lethal Tc1 excision events, wild-type (N2) males were crossed to *unc-52(st196::Tc1); mut-4(st700)I* homozygotes and heterozygous hermaphrodite progeny were picked singly to new plates. Those animals that did not segregate paralyzed Unc offspring were brooded and these plates were scored for the presence of arrested embryos. Lethal alleles obtained in this manner were subsequently balanced with the free duplication, mnDp34 (Herman *et al.*, 1979).

VI. *Construction of mec-8; unc-52 doubles mutants*

To construct *mec-8; unc-52* double mutants, spontaneous males were isolated from revertant stock plates and used to establish male stocks. Homozygous *unc-52* revertant males were then crossed to *mec-8(u74)* hermaphrodites. Several of the resulting double heterozygous hermaphrodites were plated individually. From plates that segregated Mec animals, single Mec hermaphrodites were transferred and brooded. The parental hermaphrodite was then tested by PCR for the deletion (revertant) allele of *unc-52*. Strains carrying a deletion allele were retested using PCR for a wild-type *unc-52* allele. Strains that did not produce a wild-type PCR product were presumably homozygous for the deletion.

VII. *Dye-filling to stain amphid/phasmid neurons*

To confirm that putative *mec-8; unc-52(revertant)* animals were homozygous for *mec-8*, we stained these animals with the fluorescent dye 3,3'-dioctadecyloxacarbocyanine (DiO; Molecular Probes) to evaluate the dye-filling defects (Dyf phenotype) as described by Herman and Hedgecock (1990). Briefly, worms were bathed in a solution of 10 mg/ml in M9 buffer for 2-3 hours at room temperature and then transferred to a fresh plate for about 1 hour to remove excess dye before being mounted on slides for fluorescence microscopy (using an excitation wavelength of 488 nm).

VIII. *Immunoflourescence techniques and microscopy*

Immunoflourescence staining was performed on populations of embryos from N2, CB444, RW6013 and adults from N2 and DM4415. The specimens were fixed and stained using methods previously described

(Rogalski *et al.*, 1993, 1995). The antibodies used were a rabbit polyclonal antiserum to *unc-52*, GM1 (Moerman *et al.*, 1996); a mouse monoclonal antibody to *unc-52*, MH3 (Francis and Waterston, 1991; Rogalski *et al.*, 1993, 1995); and a mouse monoclonal antibody to myosin heavy chain A, DM5.6 (Miller *et al.*, 1983). The secondary antibodies, FITC-labeled donkey anti-rabbit $F(ab')_2$ and TRSC-labelled donkey anti-mouse $F(ab')_2$, were purchased from Jackson ImmunoResearch Laboratories, Inc.

Confocal images were collected using the MRC 600 system (Bio-Rad Microsciences Division) attached to a Nikon Optiphot-2 compound microscope. Optical sections were taken at 0.2 micron intervals and combined using the PROJECT ("maximum projection") function. For publication, image files were transferred to a Macintosh computer and arranged and annotated using Adobe Photoshop 3.0/4.0. The computer images were printed on a Codonics NP-1600 printer.

For polarized light microscopy, live worms were viewed following procedures described by Waterston *et al.* (1980).

RESULTS

I. *Domain IV of perlecan is essential for myofilament assembly in muscle and for nematode viability*

The UNC-52/Perlecan-encoding gene of *C. elegans* is a complex locus that encodes at least three different isoforms of the protein. These isoforms are 1) a domain I, II and III form; 2) a domain I, II, III and IV form; 3) a domain I, II, III, IV and V form. Since the domain V-containing isoform has only recently been discovered, it is not included in this study. Polyclonal antisera specific to certain regions of UNC-52 were generated to distinguish the domain I, II, III isoforms from the domain I, II, III, IV isoforms. We refer to these as the short and long isoforms, respectively. Despite the size and apparent complexity of the *unc-52* locus, it has only two broad classes of alleles, an Unc (uncoordinated) class and a lethal class. All mutations that lead to Unc animals lie within the alternatively-spliced exons 16, 17 and 18 in domain IV. Prior to this study, the locations of only two of the mutations that lead to lethality were known. The *ut111* alteration is the result of a Tc1 insert in exon 2 and *st549* results from a stop codon in exon 7 (Rogalski et al, 1993).

It was determined that *st549* represents the null state of the gene (Williams and Waterston, 1994; Rogalski et al, 1993). Staining of this mutant with, GM1, a polyclonal antibody that recognizes all isoforms of UNC-52 demonstrated that there is no detectable accumulation of UNC-52. The GM1 antisera recognizes epitopes encoded by a portion of exons 9 and 10 and thus detects both short and long isoforms of UNC-52 (Moerman *et al*, 1996). Of the remaining alleles, *ra401* has an identical phenotype and staining

pattern with UNC-52 antibodies as *st549*. Another allele, *st572*, is somewhat similar to *ut111* (G.P. Mullen, K. Norman and J. Bush unpub. results). Body wall muscle in animals homozygous for these mutations stains weakly with antisera to UNC-52 and it appears that the protein is unable to be transported from muscles to the BM. The remaining four alleles, *st546*, *st560*, *st578* and *ra112*, all share the Pat phenotype: homozygous progeny from these strains hatch and die at the two-fold stage of development (Figure 3; Table 2). This terminal phenotype is identical to *st549* animals, however, these mutants exhibit a different immunoreactive staining pattern with antibodies to UNC-52. We observed that body wall muscle immunoreactivity was absent in these mutant embryos, but that other tissues including the pharynx and anal sphincter/retractor muscles exhibit a wild-type staining pattern (G.P. Mullen, P. Rahmani, J. Bush, unpub. results). The presence of immunoreactive material in the pharynx and elsewhere but not in the body wall muscles of *st546*, *st560*, *st578* and *ra112* animals indicates that these mutants do not have an alteration in a region of the gene common to all UNC-52 isoforms (i.e. not in any of the first 3 domains). The four mutant strains were also examined with, MH3, a monoclonal antibody that recognizes an epitope encoded by exon 19 in domain IV (Francis and Waterston, 1991; Rogalski *et al*, 1993, 1995). This antibody detects all long isoforms of UNC-52. None of these four strains contained any reactive material to MH3 (G. P. Mullen, unpub. results). This obseravtion is consistent with the lack of staining in *st549* null embryos. A possible interpretation of these results is that the antibodies and mutants allow us to

Table 2.

The new domain IV-specific knockout mutations

Allele	Sequence Change	Mutation	Phenotype
st560	C -> T at 8989 (Gln -> ochre)	stop	Pat (arrested at two-fold)
ra112	8832 -> 12115 removed	deletion	Pat (arrested at two-fold)

discern the differential expression of the short and long isoforms of UNC-52 within the developing embryo; short forms are present in the pharynx and other non-body wall muscle, but only long isoforms are present in the body wall muscle.

In the above experiments, mutant embryos were co-stained with the mAb, DM5.6, which recognizes the minor body wall muscle myosin, MHC A (Miller et al, 1983; G.P. Mullen, pers. comm.). DM5.6 allows the visualization of the myofilaments in the body wall muscles. Within the body wall muscles of the four lethal mutants *st546, st560, st578* and *ra112*, myosin is not organized into ordered A-bands but instead forms large aggregates (see Figure 5, panels C and D). An important implication of these observations is that a long form of UNC-52 (containing domain IV) is necessary for myofilament lattice formation and/or maintenance in body wall muscle of *C. elegans*.

The above model suggested a differential distribution of short and long UNC-52 isoforms. The corollary that long isoforms are essential for myofilament organization would be largely confirmed if we could find a mutation within domain IV in any one of *st546, st560, st578* or *ra112*. The phenotypic characterization of these mutants using antibodies, as described by G.P. Mullen, led us to suspect that any sequence alterations in *unc-52* should lay downstream (3') of exon 10 because GM1 (exon 10) shows immunoreactive material in these mutants but MH2 (exon 19) does not. Genomic DNA was PCR-amplified from *st546, st560, st578* and *ra112* homozygous Pat embryos in the region downstream of exon 10 and upstream of exon 19 (see Table 1 for amplification products). While the mutations in *st546* or *st578* could not be found in this region (see Appendix for details), the nucleotide alterations in

st560 and *ra112* were determined. The *unc-52(st560)* mutation is a C to T transition at nucleotide 8989 in exon 13 changing a glutamine residue (CAA) to an ochre stop codon (TAA) (Table 2; Figure 4). At the protein level, this means that *unc-52* gene products are truncated in this mutant at the second immunoglobulin repeat of domain IV. The immunofluorescence observations using the antibody GM1 show that UNC-52 reactivity is present in the BM overlying the pharynx and anal sphincter/depressor muscles in this mutant, but only residual UNC-52 staining is observed over the body wall muscles (Figure 5, panel C). This residual staining could result from the antibody detecting a truncated product, or a low level of expression of the short UNC-52 isoforms from the body-wall muscles. While we favor the former possibility, we cannot distinguish between these two possibilities at this time. What is clear is that without a long isoform of UNC-52, body wall muscle is severely disorganized and lacks all myofilament organization (Figure 5, compare to wild-type).

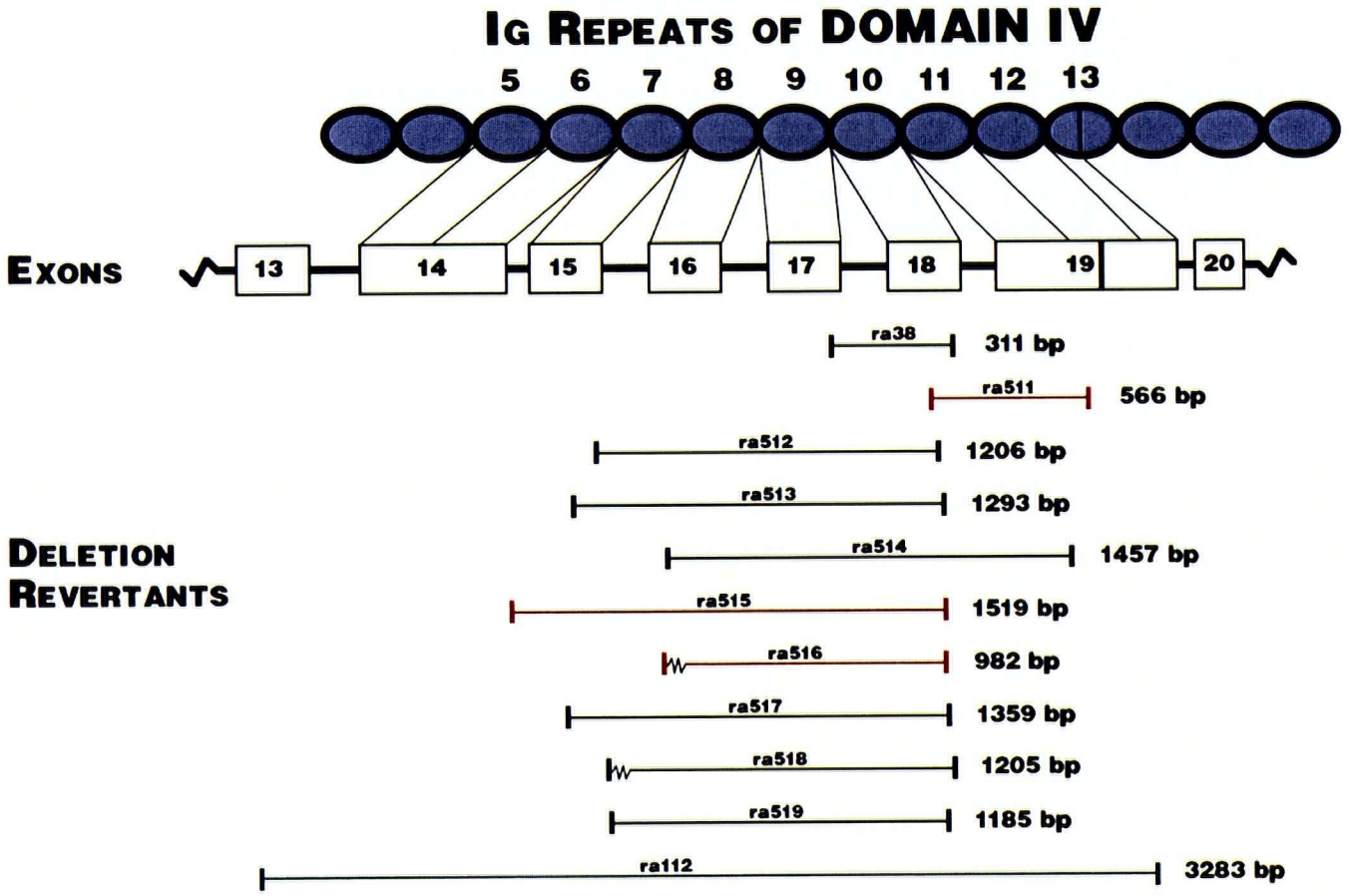
The *ra112* allele was initially isolated using a different approach than that used to identify the other Pat alleles. Transposons inserted within genes have previously been used successfully to isolate rearrangements of the genome in *C. elegans* (Kiff *et al*, 1988; Moerman *et al*, 1991). A small-scale reversion screen was performed with the intent of isolating new lethal *unc-52* mutations resulting after transposon excision and imprecise repair events in domain IV (Engels *et al*, 1990; Gloor *et al*, 1991; Moerman *et al*, 1991). We took advantage of the paralyzed Tc1 allele *unc-52(st196::Tc1)* This strain is ideal since the Tc1 mutation in *st196* is within exon 18 which is one of the alternatively spliced exons within domain IV. From approximately

Figure 7.

Deletions revertants in domain IV

Representation of the 10 deletion revertants in *unc-52* and the one lethal allele showing the relative extents of the breakpoints. Molecularly, the revertants range in size from 311 bp to 1519 bp while the Pat allele is an out-of-frame 3283 bp deletion. Even the largest viable revertant, ra515, which lacks 4 Ig repeats from the region, has a virtually wild-type phenotype. The red-coloured deletions are those used in the construction of double mutants with mec-*8* while the green-coloured deletion represents the Pat allele.

Figure 7.
Deletion revertants in domain IV



1200 heterozygous +/*unc-52(st196)* animals screened, a single Pat mutant was identified (P. Rahmani, J. Bush and G.P. Mullen, unpub. results). We examined PCR-amplified fragments from both wild type and mutant embryos in a 5 kb region of *unc-52* flanking the Tc1 insertion site in exon 18 to see if we could detect a molecular alteration within the locus. One of the amplified fragments from *ra112* homozygous embryos had a smaller fragment than what we observed in wild type animals (data not shown), and after sequencing across the region, it was determined that this alteration is a 3283 bp out-of-frame deletion (Table 2). This large deficiency is wholly contained within the domain IV encoding portion of *unc-52* (Figures 4 and 7). The immunostaining results demonstrate that the deficiency does not interfere with the expression of short isoforms of UNC-52, and to reiterate, there is no organized myosin. These results again confirm that a domain IV-containing long isoform of UNC-52 is required to permit myofilament assembly to occur at this time during embryonic development.

II. *Alternatively-spliced Domain IV immunoglobulin-encoding exons are dispensable*

Previous studies had suggested the hypothesis that individual exons within the region of alternative splicing in domain IV may be dispensable (Gilchrist and Moerman, 1992; Rogalski *et al*, 1995). In order to test this, a series of exons were removed from this region by again utilizing the transposon insertion allele, *unc-52(st196::Tc1)*. Instead of selecting for Pat animals, as in the screen that yielded *unc-52(ra112)*, we instead selected for intragenic revertants with viable phenotypes. We identified

approximately 120 independent revertants which exhibited improved movement compared with the paralyzed parental strain (J. Bush and G.P. Mullen, unpub. results). Using PCR primers that flank an approximately 2.6 kb region which encompasses the alternatively-spliced exons, fourteen wild-type looking strains showed clear DNA alterations (see Table 3). Upon sequencing, it was determined that four of the fourteen revertants have small insertions (from 3 to 30 bp), one has a small 9 bp in-frame deletion, while the remaining 9 revertants contain deletions removing up to several hundred basepairs (Table 3). These latter nine were of particular interest because these strains all exhibit nearly wild-type muscle structure and movement. The only observable phenotype appears to be a minor fragility of the muscle. Revertant worms do not appear to be as robust as wild-type animals when being transferred between plates and occasionally seem to be more susceptible to mechanical disruption of their body wall muscles. This is observed as individual worms being sluggish or uncoordinated in movement only after being transferred.

An analysis of the DNA breakpoints in the revertants shows some interesting features. The most prominent characteristic of the rearrangements is the close proximity of six of the deletion endpoints to the Tc1 insertion site (Figure 8). The alleles *ra512, ra513, ra515, ra516, ra517,* and *ra519* all have a right breakpoint within 12 bp of the original Tc1 insertion dinucleotide (TA), while the left breakpoints appear more random in their distribution. This may be a property of transposition or the DNA repair mechanism itself. Our selection for viable revertants may also impose some restraints on the deletion breakpoints. The sequences either side of the breakpoints in these deletions do not

Table 3.

Revertants obtained from Tc1 excision screen

Allele	Rearrangement
DM4305*	30 bp insertion
DM4310*	24 bp insertion
unc-52(ra507)	12 bp insertion
DM4313*	3bp insertion
unc-52(ra506)*	9 bp deletion
unc-52(ra511)	566 bp deletion
unc-52(ra512)	1206 bp deletion
unc-52(ra513)	1293 bp deletion
unc-52(ra514)	1457 bp deletion
unc-52(ra515)	1519 bp deletion
unc-52(ra516)	982 bp deletion
unc-52(ra517)	1359 bp deletion
unc-52(ra518)	1205 bp deletion
unc-52(ra519)	1185 bp deletion
unc-52(ra112)	3283 bp deletion

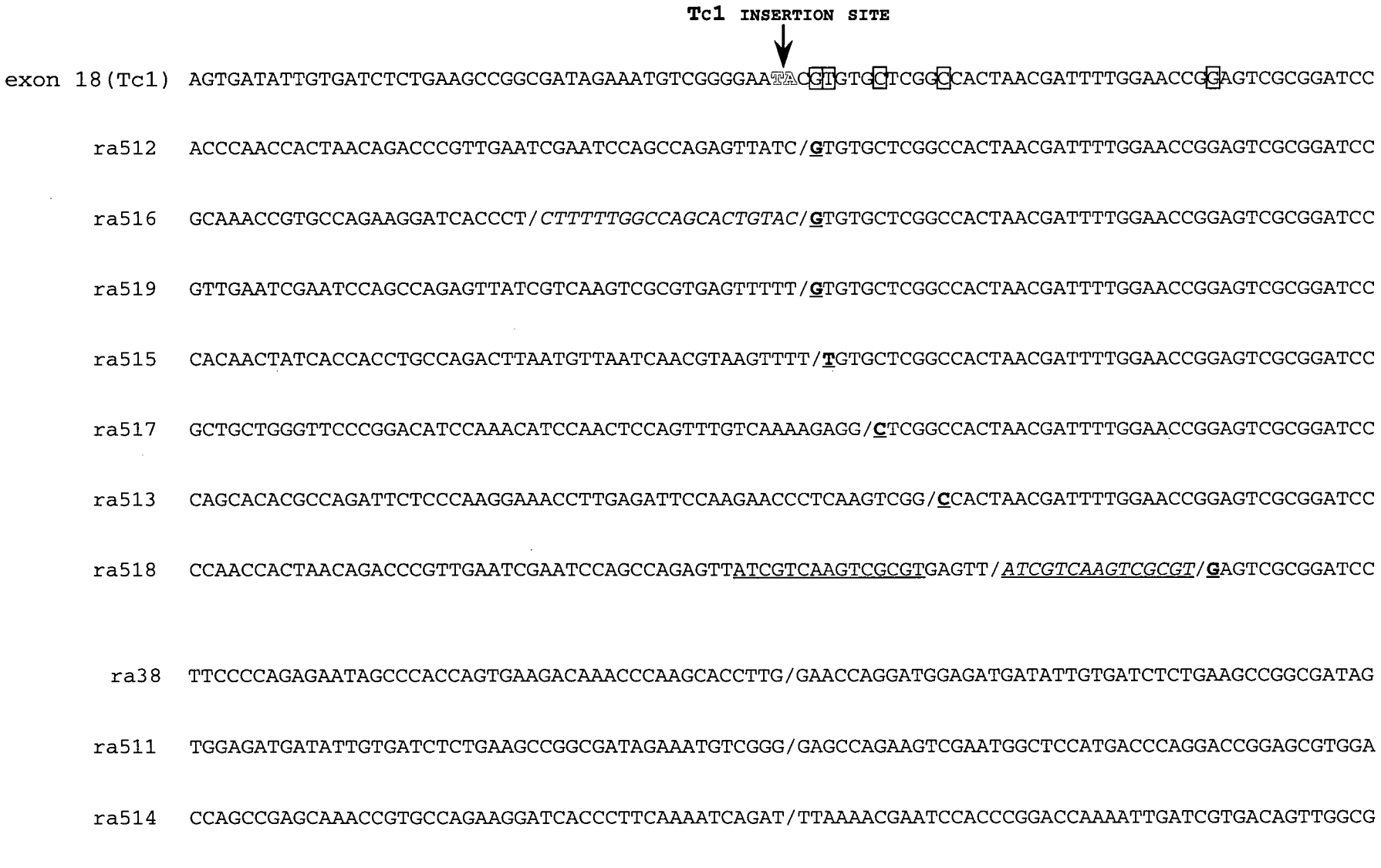
* denotes strains that were subsequently lost.

Figure 8.

Deletion breakpoints

(Top line) Partial exon 18 sequence indicating the TA dinucleotide (shown in outline) insertion site for Tc1 and some of its surrounding sequence. The alleles show their associated deletion breakpoints denoted as a slash(/). Notice the close proximity of the right-hand endpoints of ra512, ra513, ra515, ra516, ra517 and ra519 (boxed nucleotides) to the Tc1 site--all six mutations break within 12 nucleotides of TA. Italicized sequence represents the 'footprint' remaining after Tc1 excision followed by gap repair of the free DNA ends. Ra516 has a 20 nt addition with no similarities to surrounding sequence so it probably arose from the ends of Tc1 itself acting as template during strand invasion. Ra518 is clearly a duplication of one end possibly resulting through unequal cross-over. All other mutations have likely occurred through precise excision.

Figure 8.
Deletion breakpoints



share any obvious sequence similarities so they are unlikely to be the result of unequal crossing over, as was proposed for the intragenic revertant *unc-52(ra38)* (Rogalski *et al*, 1995). Two of the revertants (*ra516* and *ra518*) do contain small rearrangements in addition to a deletion, but these alterations still maintain the correct reading frame (see Figure 8 legend for further details). The other seven members of this set all contain perfect in-frame deletions (refer to Figure 7 for a schematic view of the genomic extent of each deficiency).

These seven domain IV deletions range from 982 bp to 1519 bp and remove between two and four Ig repeats (see Table 2 and Figure 7). In some instances (for example *ra512* and *ra514*) it appears that 'chimeric' exons may result in a hybrid Ig unit (Figure 9). These have the correct arrangement and spacing of conserved cysteines and surrounding amino acids believed to be important for tertiary folding of Igs (see Figure 10 for typical ribbon structure of an NCAM-like Ig). The chimeric exons of the other alleles are probably spliced out of the pre-mRNA as they appear to encode only partial Igs and these are likely non-functional (Figure 9). Each deletion revertant strain removes at least a couple of Ig repeats and yet this does not appear to have much of an effect on nematode growth, motility, nor on muscle development (Table 4). The strain, *unc-52(ra515)*, containing the longest deletion was stained with GM1, an antibody specific to UNC-52, and DM5.6, an antibody specific to a myosin heavy chain A (Miller *et al*, 1983), to determine the state of perlecan and the muscle integrity. This mutant exhibits a completely wild-type UNC-52 and myosin staining pattern (data not shown) despite lacking at least four immunoglobulin repeats from domain lV (exons 15, 16, 17, and

Table 4.

Characterization of noteworthy revertants alone and in double homozygous combination with a *mec-8* null

Strain	Muscle Structure	Movement	Amphid Staining	Phasmid Staining	Cold-Sensitivity
N2	wt	wt	12	4	no
unc-52(ra507)	wt	wt	12	4	no
unc-52(ra511)	wt	wt	12	4	no
unc-52(ra515)	wt	wt	12	4	no
unc-52(ra516)	wt	wt	12	4	no
mec-8(u74)	wt	variable	4-8	1-2	yes
mec-8(u74); ra507	mutant	pat	?	?	?
mec-8(u74); ra511	wt	sluggish	4-8	1-2	yes
mec-8(u74); ra515	wt	sluggish	4-8	1-2	yes
mec-8(u74); ra516	wt	sluggish	4-8	1-2	yes

Note: Not all data for *unc-52(ra507)* has been determined

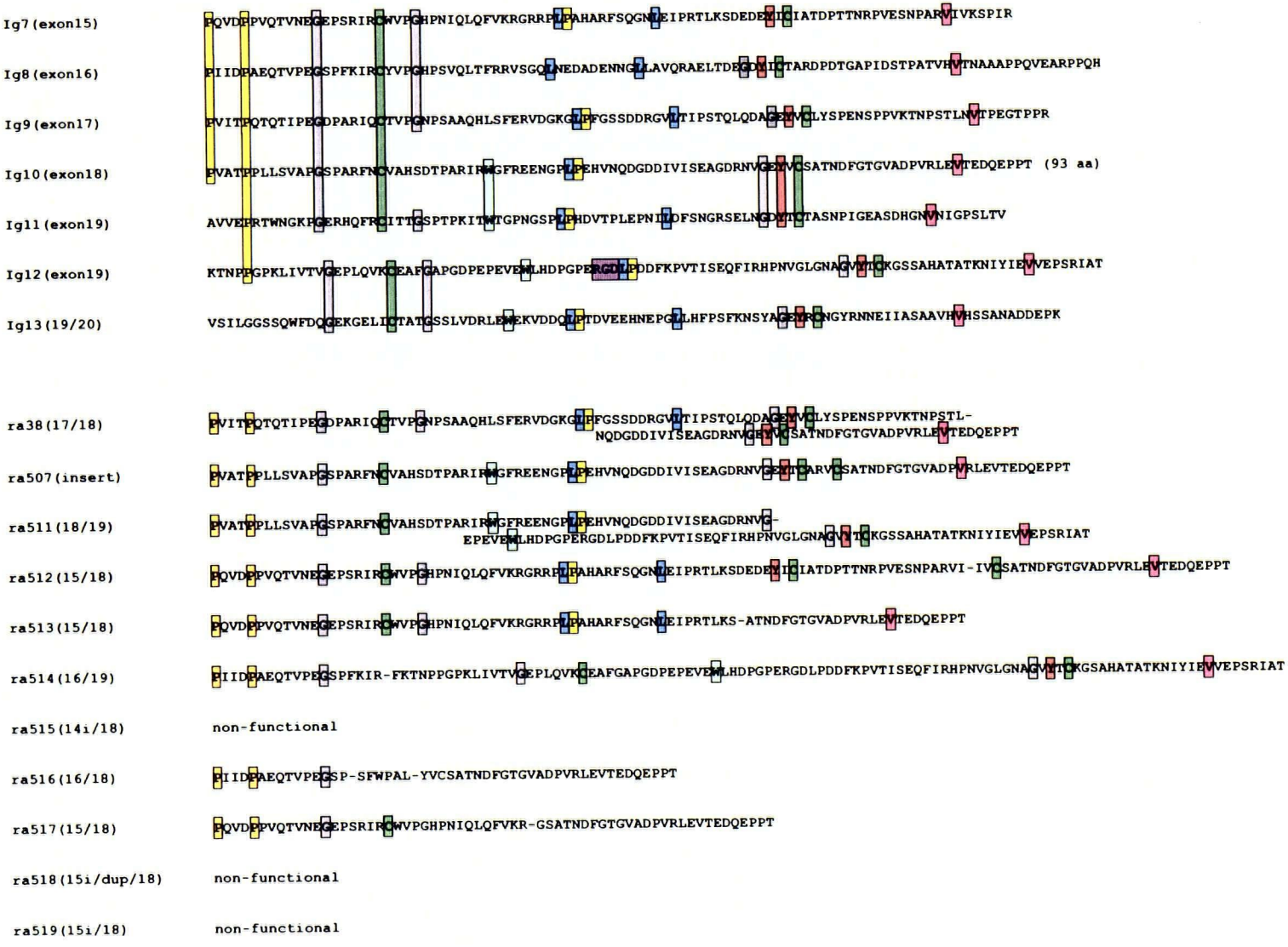
Figure 9.

Comparison of normal Igs and potential hybrids forced through chimeric exons

Amino acid sequence alignment of several Ig units from domain IV and the most conserved residues (coloured boxes). The two cysteines ([C], boxed in green) are absolutely conserved as they are required for the disulphide bridge essential to proper folding of Igs. Those revertants that may make a hybrid Ig unit are shown. The other alleles do not have enough sequence available to make anything resembling a functional Ig. Note the 3 cysteines present in ra38 and ra507 which may affect certain isoforms when MEC-8 is removed (see text). Also, note the location of the RGD motif (purple) in Ig12 encoded by exon 19.

Figure 9.

Comparison of normal and potentially hybrid Igs



18), and reducing the domain IV number of Ig repeats from fourteen to ten. The other deletion revertant alleles were examined using polarized light microscopy and they too had normal muscle (data not shown). These combined results indicate that many of the Ig repeats in this region of domain IV are indeed dispensable.

Since animals appear normal when homozygous for one of these deletions we attempted to see if we could vary the gene dosage and detect a phenotype. *unc-52(ra515)/unc-52(st560)* and *unc-52(ra511)/unc-52(st560)* heteroallelic combinations were constructed and examined. These animals were identical in phenotype to animals homozygous for the deletions indicating that a single copy of a revertant chromosome is sufficient for viability.

III. *Domain IV deletions eliminate the requirement for mec-8-mediated splicing of unc-52 during embryogenesis*

Alternative splicing within the domain IV encoding region of *unc-52* transcripts appears to be regulated (Rogalski *et al*, 1993, 1995; Lundquist *et al*, 1996). One component of this regulatory system is a putative RNA binding protein encoded by the gene *mec-8* (Lundquist *et al*, 1996). When viable *unc-52* mutations are crossed into a *mec-8(null)* background, the resulting double mutants exhibit a synthetic lethal phenotype reminiscent of Pat alleles of *unc-52* (Lundquist and Herman, 1994). Two of the splice choices within domain IV are *mec-8* dependent: the splicing of exon 15 to 19 and exon 16 to 19 (Lundquist *et al*, 1996). All other splice decisions occur through *mec-8* independent mechanisms (Figure 11). In *mec-8(null); unc-52(viable)* embryos splicing around exons containing translational stop mutations does not occur and translation is

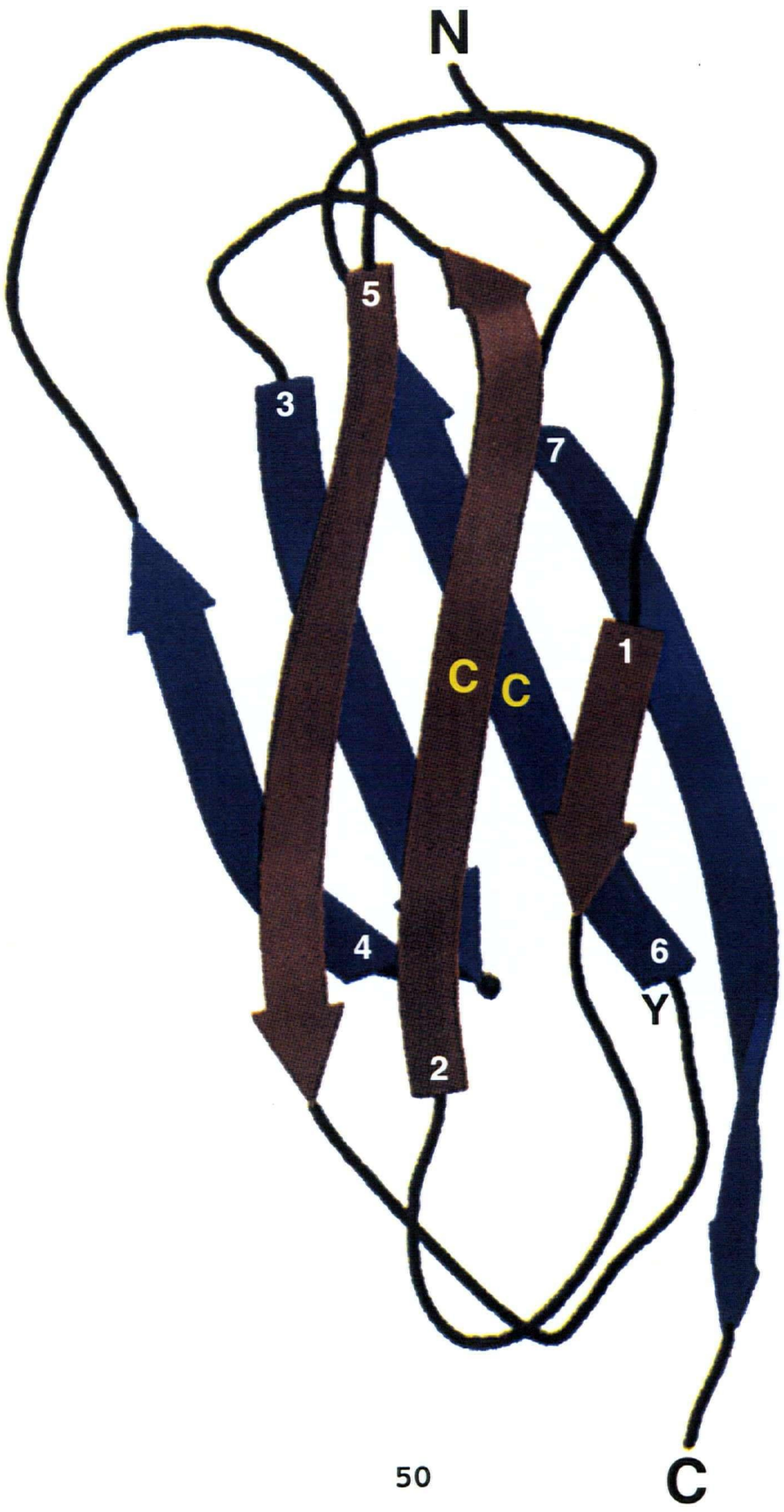
Figure 10.

Ribbon diagram of a C2-type immunoglobulin

β-sheet structure of an NCAM subtype of Ig showing the seven-stranded tertiary folding. The locations of the conserved cysteines are shown as well as the tyrosine (Y) which is the residue that includes the TA dinucleotide for Tc1 insertion. (Adapted from Vaughn and Bjorkman, 1996).

Figure 10.

Ribbon diagram of C2-type immunoglobulin



prematurely truncated during embryogenesis (Lundquist *et al*, 1996). Therefore the synthetic lethal phenotype of *mec-8(null);unc-52(viable)* double mutants results from the absence of functional UNC-52 isoforms with domain IV, as does the lethality of *unc-52(st560)* or *unc-52(ra112)* homozygotes.

We reasoned that if we could remove the target region from within the *unc-52* gene, perhaps it could become independent of any requirement for MEC-8. To test this hypothesis, the following double mutants were constructed with the *mec-8* null allele, *mec-8(u74)*: *mec-8(u74);unc-52(ra515)*, because this eliminates all of the alternatively-spliced region; *mec-8(u74);unc-52(ra511)*, since it removes the splice junctions for exons 18 and 19; and *mec-8(u74);unc-52(ra516)*, as it is the next smallest deletion removing exons 17 and 18. It was observed that these combinations are not synthetic lethal and are virtually wild-type in phenotype (Table 4). These worms are somewhat sluggish, but this is also observed with *mec-8* homozygotes. To confirm that these animals were homozygous for *mec-8*, they were examined for the dye-filling (Dyf) defects in the amphid/phasmid neurons which are observed in *mec-8* mutants (see Materials & Methods). This is simply a diagnostic tool indicative of the Mec phenotype. The reduced number of amphid and phasmid cell bodies that stained confirmed that these animals were homozygous for the *mec-8* mutation (Table 4). These observations demonstrate that by elimination of the *unc-52* substrate region (exons 15 to 18), we can remove any dependence on MEC-8 function at this locus.

Another phenotype associated with *mec-8* mutants is an embryonic and early larval cold-sensitive lethality. At 15°C, null mutants of *mec-8* exhibit an approximately 30% mortality rate

(Lundquist and Herman, 1994). Since this lethality may be due to a problem in muscle attachment, we were interested in examining the double mutant to see if *unc-52* might be implicated in this phenotype. If aborted *unc-52* splicing caused the cold sensitive lethality, then a *mec-8(null);unc-52(deletion-revertant)* double mutant should no longer have any cold-sensitive lethality. However, the double mutants are all cold sensitive (Table 4) which suggests that MEC-8 has yet another target besides *unc-52* that is important for muscle attachment.

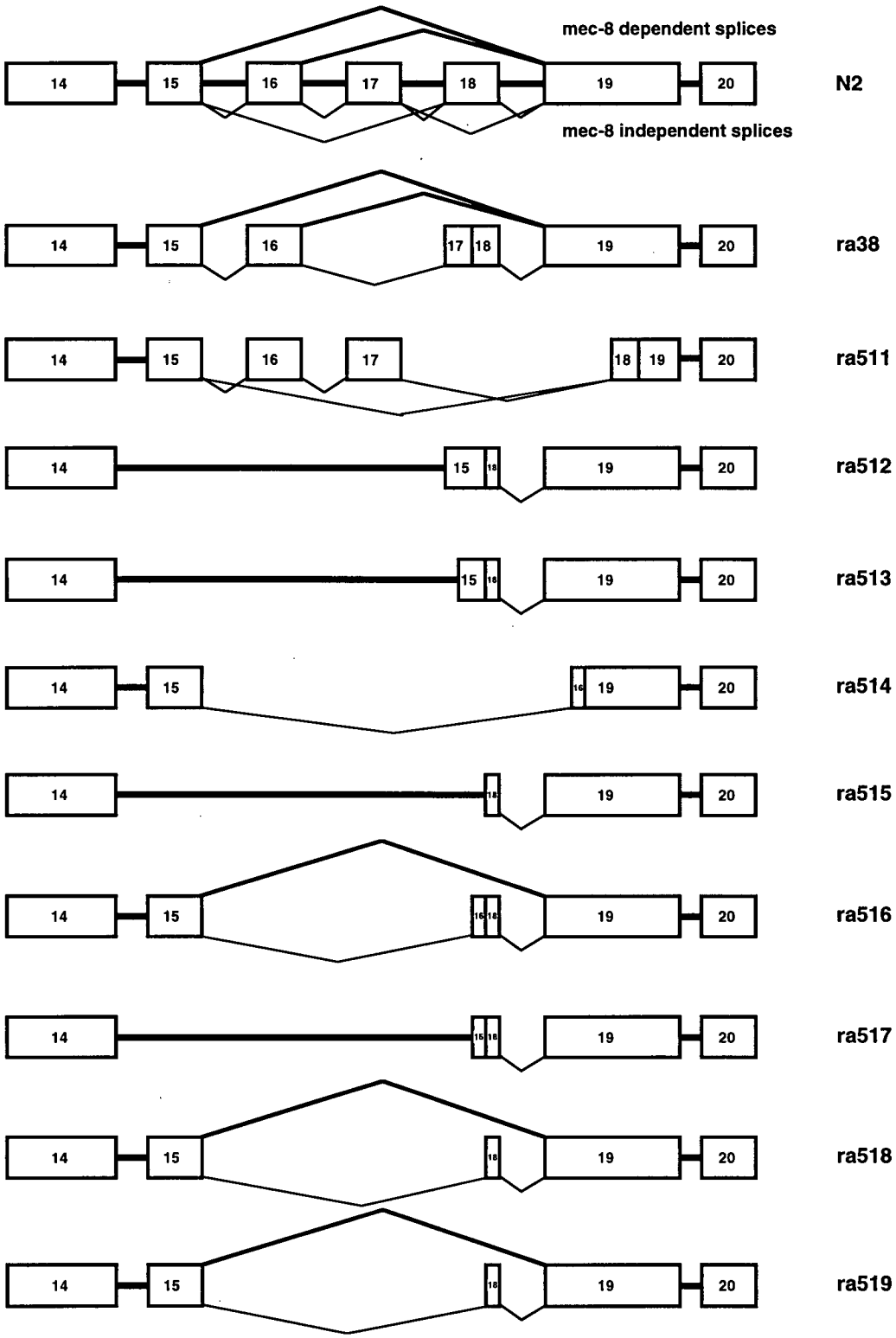
Figure 11.

Potential splice choices occuring in the alternatively-spliced region

The *mec-8*-dependent and independent splices determined through cDNA and RT-PCR data in wild-type worms(Rogalski *et al* 1993, 1995; Lundquist *et al*, 1996). *mec-8*-dependent splicing is eliminated in all revertants except *ra38, ra516, ra518* and *ra519*. The exon 15-19 splice is present in all four of these mutants but exon 16-19 is only present in *ra38*. Surprisingly, *ra516, ra518* and *ra519* are not lethal when MEC-8 is absent unlike *ra38* which is lethal.

Figure 11.

Potential splice choices in *unc-52*



DISCUSSION

The *unc-52* gene encodes a large multi-domain protein found in the BMs underlying the muscle tissues of the nematode *C. elegans* (Rogalski *et al*, 1993) and is required for the proper initiation of myofilament assembly. This is a complex locus that undergoes extensive transcriptional regulation (Rogalski *et al*, 1993, 1995; Lundquist *et al*, 1996). Much of this regulation occurs by alternative-splicing of exons within the domain IV-encoding portion of the gene and this leads to several distinct long isoforms of the protein. Several exons of domain IV encode a single immunoglobulin unit, exceptions are exons 14 and 19 which encode more than one Ig, and exons 24 and 25 which together form a single Ig unit. The combinatorial nature of the Ig units may be dependent on the type of isoform required during development. Expression of *unc-52* varies over time and space (Rogalski *et al*, 1995; G. P. Mullen, unpub. results). Two groups of isoforms are produced based on the presence or absence of domain IV. Short variants lack domain IV while long forms contain this domain. The long isoforms are specifically localized to the BMs of the body-wall muscles during embryogenesis. The primary goal of this thesis has been to determine the importance of domain IV for perlecan function and muscle development. The specific questions that have been addressed include, 1) Is domain IV critical for body wall muscle development? 2) Are certain Ig units more critical than others? and 3) Is overall length of domain IV a crucial factor? To address these questions, several mutations within the *unc-52* locus that possess specific alterations have

been analyzed. A comparison of the location of these lesions with the phenotypes of animals homozygous for these mutations leads us to a fuller appreciation of the complexity of *unc-52* and a greater understanding of the function of this complex gene during muscle development.

I. *Domain IV is essential for UNC-52 function during embryogenesis*

The identification and characterization of mutations within the *unc-52* gene demonstrate that UNC-52 is an important and essential component for myofilament assembly within muscle cells and for muscle growth in *C. elegans* (Brenner, 1974; Mackenzie *et al*, 1978; Waterston *et al*, 1980; Gilchrist and Moerman, 1992; Rogalski *et al*, 1993, 1995; Hresko *et al*, 1994; Williams and Waterston, 1994). The observation that *st549* is a null allele and that it leads to the loss of both short and long isoforms of UNC-52 led to the speculation that lethality results when animals have no UNC-52 isoforms (Rogalski *et al*, 1993). The true cause of embryonic lethality is more complicated and more subtle than this simple explanation. Antibodies to different regions of UNC-52 reveal that long and short isoforms of this protein have different cellular distributions during morphogenesis in the developing embryo. Long isoforms appear to be the predominant product from developing body wall muscle cells while short isoforms are the major product from other muscle tissue during late embryogenesis (G.P. Mullen and D.G. Moerman, unpub. results). Although *st549(null)* animals lack both types of isoforms and do not have any UNC-52 protein (Rogalski *et al*, 1993), other Pat alleles like *st560* and *ra112* lack only the body wall-specific long isoforms. What these three mutations have in common is that they all lead to

disorganization of the myofilament array within body wall muscle (Figure 5, panels B, C, D) and thus paralysis at a critical stage of development when elongation of the embryo should commence (refer to Figure 3).

Complete absence of all *unc-52* gene products is therefore not responsible for the lethality. Rather, it is the lack of a complete domain IV-containing body wall muscle-specific isoform that leads to embryonic lethality. The lethality results in a lack of body wall muscle function which is a direct result of a failure to initiate proper myofilament assembly (Francis and Waterston, 1991; Williams and Waterston, 1994). Our model for *unc-52* function is founded on our characterization of the mutant phenotypes and the earlier observation that UNC-52 concentrates at the base of dense bodies and M-lines (Francis and Waterston, 1991). Dense bodies are analogous to Z-lines in vertebrates but also act as the nematode homologue of a vertebrate focal adhesion plaque. We believe UNC-52 aids in the assembly and anchoring of the dense bodies as they are formed during morphogenesis. Whether UNC-52 plays an instructive/signaling role or simply an attachment role in the assembly of integrin complexes and myofilament anchoring within muscle is still not clear--although it probably functions primarily in attachment at later stages. Without a stable focal attachment structure at the membrane/BM interface to act as an anchor, sarcomere units within muscle cells cannot be properly organized (Burridge *et al*, 1988; Volk *et al*, 1990; reviewed in Moerman and Fire, 1997). Many of the Pat mutants described by William and Waterston (1994) affect some aspect of myofilament assembly or anchoring at this critical stage of muscle development and morphogenesis.

At an earlier stage in this project, our interpretation of these results was that they indicated the essential nature of domain IV of *unc-52*. One model we considered was that perhaps domain IV interacts with the integrin complex. New results based on further sequencing in the *unc-52* region suggests that this is unlikely. We now realize that the *unc-52* gene is even larger than we had previously realized (Rogalski *et al*, 1993; see Introduction). We had described a gene with four domains and 26 exons spread over about 15 kb. More recently the genome sequencing consortium has provided us with further sequence 3' of what we thought was the terminus of the gene. Analysis of this region reveals another 11 exons and extends the gene another 10 kb, approximately (T.M. Rogalski, G.P. Mullen and D.G. Moerman, pers. comm.). This new region is similar to domain V of mammalian perlecan and contains several laminin-like segments and EGF-like repeats (Noonan *et al*, 1991; Kallunki and Tryggvason, 1992; Murdoch *et al*, 1992). The significance of this finding is two-fold. First, it indicates that the importance of domain IV probably lies in its role as a linker between domain V and the rest of the molecule. Since the predicted modular structure of domain IV resembles a rod (Noonan *et al*, 1991), it may be present in a passive sense only to help position domain V. Second, the portion of laminin contained within domain V is thought to be capable of interacting with integrin complexes. This laminin portion of domain V may be the real interactive partner with cell surface components and this may be the important difference between long and short isoforms. We are currently attempting to obtain targeted mutations in this domain to test this hypothesis (T. Rogalski, C. Warren, J. Culotti and D. Moerman, pers. comm.).

II. *Effects of removing alternatively spliced exon of unc-52*

Domain IV of unc-52 encodes 14 Ig repeats. This repetitive modular structure is typical of ECM and large transmembrane proteins with large extracellular components (reviewed in Vaugh and Bjorkman, 1996). It is also found in a few intracellular proteins, most notably large structural muscle components (e.g. the twitchin/titin family; Benian *et al*, 1989; Labeit *et al*, 1992). Long tracts of immunoglobulins within proteins have been proposed to act as rulers to establish a periodicity when building larger molecular aggregates (reviewed in Trinick, 1994). To date, only one *bona fide* "ruler" protein has been identified. This is the lambda phage protein, gpH, (Katsura, 1987) which acts to form a scaffold for the formation of the tail region of this phage. My results showing that up to four Igs can be eliminated without affecting the spacing or length of myofilament components, or affecting any other aspect of muscle development, strongly supports the conclusion that domain IV does not act as a ruler. The results are in agreement with earlier studies that showed that intragenic suppression of viable *unc-52* mutants occurs via exon skipping, or by an in-frame deletion fusing two adjacent exons (Gilchrist and Moerman, 1992; Rogalski *et al*, 1995). These combined observations leave us puzzled about the significance of alternative splicing of domain IV since the largest of the deletion revertants eliminate all of the alternatively spliced Igs and yet we see no deleterious effects on the animal. One conclusion that can be derived from these observations is that none of the alternatively spliced Igs have unique or essential functional roles in muscle development.

Might we have been able to obtain even larger deletion revertants that eliminate even more Ig motifs? In other words, how small could one make domain IV and still have animals survive? This is an interesting question but a difficult one to answer. The Tc1 excision screen we utilized to obtain these rearrangements has several restrictions and possible limitations. Perhaps the most serious limitation was that the screen was biased for those genetic events that produced only better moving animals. This occurred with a high frequency ($\sim 10^{-2}$-10^{-3}) and we were able to analyze over 120 revertants, but only a small proportion of these actually had deletions (Table 2). The majority of revertants probably arose as a result of precise excision events although our data does not specifically address this point. The ratio of rearrangements to precise repair is determined by the ability of the homologues to pair, the specific properties of the repair machinery and in these experiments, our ability to detect the phenotype (see Engels *et al* 1990; Gloor *et al*, 1991 for data pertaining to the repair process). The *ra515* mutation at over 1.5 kb was the largest deletion we isolated that still segregated wild type animals. The 3.2 kb *ra112* deletion mutant demonstrates that it is possible to obtain larger deletions after Tc1 excision, but in this instance we did not demand viability.

The deletion breakpoints in the deletion revertants appear to preferentially extend 5′ rather than 3′ from the transposon insertion site (Figure 7). While our sample size is small, we speculate that this implies a restraint on material that can be removed 3′ of the Tc1 insertion site and still maintain viability. As mentioned earlier UNC-52 may interact with integrins on the transmembrane surface of the muscle cells. Interestingly, 600 bp

3' of the Tc1 insertion site within exon 18 there is encoded within exon 19 an RGD (arg-gln-asp) motif, a demonstrated ligand for integrin receptors (Lawler and Hynes, 1989; reviewed in Ruoslahti, 1996). If one considers that the average deletion size in the revertants is over 1200 bp and presuming an equal distribution around the excision site, then by inspection, the RGD site should be removed. Since it is not removed by any of the viable deletion revertants, we suggest that this motif or some other aspect of the region 3' to exon 18 is critical for domain IV function in UNC-52. We recognize that this is a difficult point to stress because our sample size is small. This possible bias in excision direction and the fact that we did not recover any deletion revertants that extend beyond the alternatively spliced region further suggests that perhaps a smaller domain IV is not compatible with viability.

III. *Effects of mec-8;unc-52(deletion-revertant) mutants on muscle development*

The *mec-8* locus produces a putative RNA-binding protein which has been shown to promote alternative splicing of two *unc-52* pre-mRNA transcripts, exon 15 - 19 and exon 16 -19 (Lundquist, 1994; 1996). Accumulation of other alternatively-spliced *unc-52* transcripts is *mec-8*-independent. The MEC-8-dependent *unc-52* splice products (15 - 19 and 16 - 19) are not absolutely required for embryogenesis in *unc-52(+)* animals since *mec-8* mutants are capable of normal muscle development. However, *mec-8(null)* and *unc-52(paralyzed)* double homozygotes exhibit a synthetic lethal phenotype similar to the Pat alleles of *unc-52* (Lundquist and Herman, 1994). Like true Pat alleles, these double mutants produce

only truncated UNC-52 isoforms (Lundquist *et al*, 1996). This is because the mutated exon that would normally be spliced out during embryogenesis has become the default expression pattern in these embryos. We attempted, through the use of double mutants between *mec-8* and the deletion revertants of *unc-52*, to see if we could eliminate the necessity for MEC-8 processing of *unc-52* transcripts. A secondary reason for these experiments was to see if we could help define the target region within *unc-52* transcripts that MEC-8 acts upon.

Our reasoning in these experiments was that if MEC-8 only interacts with the exon 15 through 19 region of *unc-52*, then by removing this region we should liberate *unc-52* from any requirement for MEC-8. We constructed three *mec-8; unc-52(deletion-revertant)* mutants that are viable and have wild type muscle structure and movement (Table 4). This indicates that indeed the processing of *unc-52* transcripts in these examples is now independent of *mec-8* function (Figure 11). It also further confirms that this region is probably the only portion of *unc-52* that MEC-8 regulates. The phenotype of *mec-8* mutants is complex and includes mechanosensory as well as chemosensory defects and low penetrance cold-sensitive embryonic arrest (Chalfie and Sulston, 1981; Perkins *et al*, 1986; Lundquist and Herman, 1994). This latter phenotype has been correlated with defects in the attachment of body wall muscle to adjacent hypodermis and cuticle (Lundquist and Herman, 1994). The double mutants still exhibit all three Mec phenotypes indicating that MEC-8 must have other target transcripts besides *unc-52*. In particular, since these double mutants are still weakly penetrant for the cold-sensitive lethal phenotype, MEC-8 must interact with other target genes that

affect muscle attachment. Suppressor screens looking for rescue of the Mec-8; Unc-52 synthetic lethal phenotype have found revertants which map to two independent loci (see Smu mutants in Lundquist and Herman, 1994). More recently, a new locus has been defined that when mutated forms a synthetic lethal with *mec-8* but not with *unc-52* paralyzed mutants (Bob Herman, pers. comm.).

Not all double mutants between *mec-8* and *unc-52* revertants are viable; the *mec-8; unc-52(ra38)* double mutant is a synthetic lethal (B. Herman, pers. comm.) as is *mec-8; unc-52(ra507)* (Table 4). While we are not certain why these two constructs form synthetic lethals, we suspect the chimeric (*ra38*) or altered (*ra507*) Ig formed in these revertants may not be compatible with *unc-52* function. Synthetic lethal *mec-8; unc-52(ra38)* embryos accumulate high levels of UNC-52 within the cell, but do not release any detectable protein to the ECM (G. Mullen unpub. results). Each of these altered Igs contains three cysteines (Figure 9) which may be incompatible with proper Ig folding and therefore lead to their accumulation within the intracellular trafficking pathway of the cell. These observations imply that *mec-8* may be a useful tool to determine whether chimeric exons of *unc-52* are functional at the protein level.

IV. *Summary*

A common theme among many extracellular proteins, including perlecan, laminins, and fibronectin is the incorporation of a finite number of modular domains into large multi-domain polypeptides. Over half of the 40 modules which are typical for extracellular proteins have so far been detected in mammalian BMs (Bairoch, 1995). A series of repeating units based on the Ig-

superfamily motif is one such module popular within extracellular molecules. It has been suggested that strings of multiple Igs are important for homophilic interactions (Ranheim *et al*, 1996) and/or heterophilic binding (Brummerdorf and Rathjen, 1996) and/or spacing of globular domains although there does not appear to be any direct evidence in the literature supporting this concept. It has been shown that the alternatively spliced exons of domain IV are not important for any of these functions.

Alternative-splicing is a key mechanism for generating different isoforms of a protein. Most reviews of alternative splicing stress that the purpose for generating this structural diversity is to yield functional diversity. It does seem reasonable that each isoform generated through regulated alternative-splicing should have a distinct function. An excellent example comes from work on the ECM protein, agrin. Agrin is a heparan sulfate proteoglycan that induces aggregation of acetylcholine receptors (AChRs) at the neuromuscular synapse. This aggregating activity is modulated by alternative splicing and includes muscle and neural-specific isoforms (Hamshere *et al*, 1991). Studies have demonstrated that different isoforms have different binding specificities and targets (Stone and Nikolics, 1995). We have shown that the differential distribution of long and short isoforms of UNC-52 does have functional implications, but this does not appear to be true for the variants within domain IV. At this point we are left with no satisfactory explanation for the alternative-splicing within domain IV.

My experiments on *mec-8* and *unc-52* interactions have further elucidated a complex regulatory interaction. MEC-8 regulation of *unc-52* is clearly limited to the exon 15 to 19 interval and one

could interpret my results to argue that the MEC-8 binding region lies between exon 18 and 19. While this is in agreement with earlier speculations (Lundqist *et al*, 1996) other approaches will be necessary to confirm this idea. Mapping the binding site using *unc-52* mini-genes tagged with GFP is in progress (A. Davies and B. Herman, pers. comm.). The new deletion revertant strains could be useful reagents for these experiments.

This study has helped to further resolve the molecular interactions involving the *unc-52* locus. Particularly useful has been the demonstration of the essential importance of domain IV to UNC-52 function. If our model that perlecan/UNC-52 acts as a molecular 'glue' adhering muscle cells to the hypodermis has any validity, then my results point to the regions of the molecule important for this attachment. Using a combined genetic and molecular approach it has been shown that it is possible to derive a simpler functional perlecan variant which itself may be a useful tool for further genetic studies.

REFERENCES

Bairoch, A., 1995 Nomenclature of extracellular domains. *SWISS-PROT Protein Sequences Data Bank*, Release 31.0

Barstead, R.J., L. Kleinman and R.H. Waterston, 1991 Cloning, sequencing, and mapping of an alpha-actinin gene from the nematode *Caenorhabditis elegans*. *Cell Motil. Cytoskeleton* **20**: 69-78.

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**(6245):45-50.

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

Burridge, K., K. Fath, T. Kelly, G. Nuckolls and C. Turner, 1988 Focal adhesions: transmembrane junctions between extracellular matrix and the cytoskeleton. *Annu. Rev. Cell Biol.* **4**: 487-525.

Burridge, K. and M. Chrzanowska-Wodnicka, 1996 Focal adhesions, contractility, and signaling. *Annu. Rev. Cell Biol.* **12**: 463-519.

Calderwood, D.A. Tuckwell, D.S. Eble, J. Kuhn and M.J. Humphries, 1997 The integrin alpha1 A-domain is a ligand binding site for collagens and laminin. *J. Biol. Chemistry*. **272**(19):12311-12317.

Chakravarti, S., T. Horchar, B. Jefferson, G.W. Laurie and J.R. Hassell, 1995 Recombinant domain III of perlecan promotes cell attachment through its RGDS sequence. *J. Biol. Chemistry*. **270**(1):404-409.

Chalfie M. and J. Sulston, 1981 Developmental genetics of the mechanosensory neurons of *Caenorhabditis elegans*. *Developmental Biology* **82**(2):358-370.

Couchman, J.R., R. Kapoor, M. Sthanam and R.R. Wu, 1996 Perlecan and basement membrane-chondroitin sulfate proteoglycan (bamacan) are two basement membrane chondroitin/dermatan sulfate proteoglycans in the Engelbreth-Holm-Swarm tumor matrix. *J. Biol. Chemistry* **271**(16): 9595-9602.

Doege K., M. Sasaki and Y. Yamada, 1990 Rat and human cartilage proteoglycan (aggrecan) gene structure. Biochem. Soc. Trans **18**(2): 200-202.

Engels, W. R., D.M. Johnson-Schlitz, W.B, Eggleston and J. Sved, 1990 High-frequency P element loss in *Drosophila* is homolog dependent. *Cell* **62**: 515-525.

Epstein H.F., D.L. Casey and I. Ortiz, 1993 Myosin and paramyosin of *Caenorhabditis elegans* embryos assemble into nascent structures distinct from thick filaments and multi-filament assemblages. *J Cell. Biol.* **122**(4):845-58.

Francis, G.R., and R.H. Waterston, 1985 Muscle organisation in *C. elegans*: localisation of proteins implicated in thin filament attachment and I-band organisation. *J. Cell Biol.* **101**: 1532-49.

Francis, R., and R.H. Waterston, 1991 Muscle cell attachment in *Caenorhabditis elegans*. *J. Cell Biol.* **114**: 465-479.

Gilchrist, E.J., and D.G. Moerman, 1992 Mutations in the *sup-38* gene of *Caenorhabditis elegans* suppress muscle-attachment defects in *unc-52* mutants. *Genetics* **132**: 431-442.

Gloor, G.B., N.A. Nassif, D.M. Johnson-Schlitz, C.R. Preston and W.R. Engels, 1991 Targeted gene replacement in *Drosophila* via P element-induced gap repair. *Science* **253**(5024):1110-1117.

Hamshere, M., G. Dickson and I. Eperon, 1991 The muscle specific domain of mouse N-CAM: Structure and alternative splicing patterns. *Nucleic Acids Res.* **19**: 4709-4716.

Hayashi, K., J.A. Madri and P.D. Yurchenco, 1992 Endothelial cells interact with the core protein of basement membrane perlecan through beta 1 and beta 3 integrins: an adhesion modulated by glycosaminoglycan. *J. Cell Biol.* **119**(4):945-959.

Heikkila, P. and R. Soininen, 1996 The type IV collagen gene family. *Contributions to Nephrology* **117**: 105-129.

Herman R.K., J.E. Madl and C.K. Kari, 1979 Duplications in *Caenorhabditis elegans*. *Genetics* **92**: 419-435.

Herman R.K. and E.M. Hedgecock, 1990 Limitation of the size of the vulval primordium of *Caenorhabditis elegans* by *lin-15* expression in surrounding hypodermis. *Nature* **348**(6297):169-71.

Hresko, M.C., B.D. Williams and R.H. Waterston, 1994 Assembly of body wall muscle and muscle cell attachment structures in *Caenorhabditis elegans*. *J. Cell Biol.* **124**: 491-506.

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.

Inoue S. 1994 Basic structure of basement membranes is a fine network of "cords," irregular anastomosing strands. *Microscopy Research & Technique* **28**(1):29-47.

Jockusch, B.M., P. Bubeck, K. Giehl, M. Kroemker and J. Moeschner, 1995 The molecular architecture of focal adhesions. *Annu. Rev. Cell Dev. Biol.* **11**: 379-416.

Kallunki, P. and K. Tryggvason, 1992 Human basement membrane heparan sulfate proteoglycan core protein: A 467-kD protein containing multiple domains resembling elements of the low density lipoprotein receptor, laminin, neural cell adhesion molecules and epidermal growth factor. *J. Cell Biol.* **116**: 559-571.

Katsura I. 1987 Determination of bacteriophage lambda tail length by a protein ruler. *Nature* **327**(6117):73-75.

Kiff J.E., D.G. Moerman, L.A. Schriefer and Waterston RH. 1988 Transposon-induced deletions in *unc-22* of *C. elegans* associated with almost normal gene activity. *Nature* **331**(6157):631-633.

Labeit, S., M. Gautel, A. Lakey and J. Trinick, 1992 Towards a molecular understanding of titin. *EMBO J.* **11**(5):1711-1716.

Lawler J. and R.O. Hynes, 1989 An integrin receptor on normal and thrombasthenic platelets that binds thrombospondin. *Blood* **74**(6):2022-2027.

Lundquist, E.A. and R.K. Herman, 1994 The *mec-8* gene of *Caenorhabditis elegans* affects muscle and sensory neuron function and interacts with three other genes: *unc-52, smu-1* and *smu-2*. *Genetics* **138**: 83-101.

Lundquist, E.A., R.K. Herman, T.M. Rogalski, G.P. Mullen, D.G. Moerman, 1996 The *mec-8* gene of *C. elegans* encodes a protein with two RNA recognition motifs and regulates alternative splcing of *unc-52* transcripts. *Development* **122**: 1601-1610.

Mackenzie J.M. Jr., R.L. Garcea, J.M. Zengel and H.F. Epstein, 1978 Muscle development in *Caenorhabditis elegans*: mutants exhibiting retarded sarcomere construction. *Cell* **15**(3):751-762.

Miller, D.M., I. Ortiz, G.C. Berliner and H.F. Epstein. 1983 Differential localization of two myosins within nematode thick filaments. *Cell* **34**: 477-490.

Miller, D.M., F.E. Stockdale and J. Karn, 1986 Immunological identification of the genes encoding the four myosin heavy chain isofoms of *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* **83**: 2305-2309.

Moerman, D.G., J.E. Kiff and R.H. Waterston, 1991 Germline excision of the transposable element Tc1 in *C. elegans*. *Nucleic Acids Res* **19**: 5669-5672.

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*. *Devl. Biol.* **173**: 228-242.

Moerman, D.G., and A.F. Fire, 1997 Muscle: structure, function and development, pp. in Nematodes II, edited by D. L. Riddle. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Murdoch A.D., G.R. Dodge, I. Cohen, R.S. Tuan and R.V. Iozzo, 1992 Primary structure of the human heparan sulfate proteoglycan from basement membrane (HSPG2/perlecan). A chimeric molecule with multiple domains homologous to the low density lipoprotein receptor, laminin, neural cell adhesion molecules, and epidermal growth factor. *J. Biol. Chem.* **267**(12):8544-8557.

Nielson, K., B. Scott, J.C. Bauer and K. Kretz, 1994 *Strategies* **7**(3): 64-65.

Noonan, D.M., A. Fulle, P. Valente, S. Cai and E. Horigan, 1991 The complete sequence of perlecan, a basement membrane heparan sulfate proteoglycan, reveals extensive similarity with laminin A chain, low density lipoprotein-receptor, and the neural cell adhesion molecule. *J. Biol. Chem.* **266**: 22939-22947.

Perkins L.A., E.M. Hedgecock, J.N. Thomson and J.G. Culotti, 1986 Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Devl. Biol.* **117**(2):456-487.

Pfaff, M., W. Gohring, J.C. Brown and R. Timpl, 1994 Binding of purified collagen receptors (a1B1, a2B1) and RGD-dependent integrins to laminins and laminin fragments. *Eur. J. Biochem.* **225**: 975-984.

Riddle, D.L. 1997 *C. elegans* II Cold Spring Harbor Laboratory, New York.

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.

Rogalski, T.M., E.J. Gilchrist, G.M. 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.

Ruoslahti E. 1996 RGD and other recognition sequences for integrins. *Annual Review of Cell & Developmental Biology* **12**:697-715.

Schnabel R., H. Hutter, D.G. Moerman and H. Schnabel, 1997 Assessing normal embryogenesis in *Caenorhabditis elegans* using a 4D microscope: variability of development and regional specification. *Devl. Biol.* **184**(2):234-265.

Sibley, M.H., J.J. Johnson, C.C. Mello and J.M. Kramer, 1993 Genetic identification, sequence, and alternative splicing of the *Caenorhabditis elegans* alpha(IV) collagen gene. *J. Cell Biol.* **123**: 255-264.

Stone, D.M. and K. Nikolics, 1995 Tissue- and age-specific expression patterns of alternatively spliced agrin mRNA transcripts in embryonic rat suggest novel developmental roles. *J. Neuroscience* **15**(10):6767-6778.

Sulston, J.E. and H.R. Horvitz, 1977 Postembyonic cell lneages of the nematode *Caenorhabditis elegans*. *Devl. Biol.* **82**: 41-55.

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

Timpl, R., 1993 Proteoglycans of basement membranes. *Experientia* **49**(5):417-428.

Timpl, R. and J.C. Brown, 1996 Supramolecular assembly of basement membranes. *BioEssays* **18**(2): 123-132.

Trinick J. Titin and nebulin: protein rulers in muscle? 1994 *Trends in Biochemical Sciences* **19**(10):405-409.

van den Brule, F.A., C. Buicu, M. Baldet, M.E. Sobel, D.N. Cooper, P. Marschal and V. Castronovo, 1995 Galectin-1 modulates human melanoma cell adhesion to laminin. *Biochem & Biophysical Res. Comm.* **209**(2):760-767.

Vaughn, D.E. and P.J. Bjorkman, 1996 The (Greek) key to structures of neural cell adhesion molecules. *Neuron* **16**(2): 261-273.

Volk, T., L.I. Fessler and J.H. Fessler, 1990 A role for integrin in the formation of sarcomeric cytoarchitecture. *Cell* **63**: 525-536.

Waterston, R.H., J.N. Thomson and S. Brenner, 1980 Mutants with altered muscle structure in *Caenorhabditis elegans*. *Dev. Biol.* 77: 279-302.

Waterston, R. H., 1988 Muscle, pp. 281-335 in The nematode Caenorhabditis elegans., edited by W. B. Wood. Cold Spring Harbor Press, Cold Spring Harbor.

Wood, W. B., 1988 The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, New York.

Williams, B. and R.H. Waterston, 1994 Genes critical for muscle development and function in *Caenorhabditis elegans* identified through lethal mutations. *J. Cell Biol.* **124**: 475-490.

Yamaguchi, Y. and E. Ruoslahti, 1988 Expression of human proteoglycan in Chinese hamster ovary cells inhibits cell proliferation. *Nature*. **336**(6196): 244-246.

Yanagishita, M., 1993 Function of proteoglycans in the extracellular matrix. *Acta Pathologica Japonica* **43**(6): 283-293.

Yurchenco, P.D. and J.C. Schittny, 1990 Molecular architecture of basement membranes. *FASEB J.* **4**: 1577-1590.

APPENDIX

I attempted to sequence the molecular lesions responsible for the other Pat alleles. The mutants *unc-52(st546)* and *unc-52(st578)* both exhibit the same immunostaining patterns as *unc-52(st560)* but their respective mutations are not in the interval between exons 10 and 16. For completeness sake, the following primers from the Moerman Lab primer stocks (sequence available) should be tried again on amplified DNA from the respective strains:

st546 - p27 and p28 only as all other regions have been saturated,

st578 - exon 14 should be re-sequenced.

If not in these regions, the mutations most likely occur downstream of the alternatively-spliced exons, possibly exon 19 or even more 3′ in domain V.

In addition, the mutation for *unc-52(st572)* was not defined. It exhibits a unique immunostaining pattern compared to the above mentioned mutants. Using GM1, there is only faint staining of the BM over the pharynx suggesting that the mutation may be in exon 10 since this exon is common to all UNC-52 isoforms. No alterations were found in exon 10 or the 3′ end of exon 9. Because this mutant may be a representation of a missense mutation and therefore occurring anywhere in the amino terminal region, further sequencing 5′ of exon 9/10 could isolate it.

Also, the other potential null mutation, *unc-52(ra401)*, was predicted to be in the interval between exons 6-9 (since there is no UNC-52 immunoreactivity in this mutant consistent with st549 staining which is in exon 7). The following nucleotide sequences around exon 6 and 7 do not contain alterations 4570-4692, 4910-5000, 5040-5178, 5570-5725.