Mutations in the unc-52 Gene Responsible for Body Wall Muscle Defects in Adult Caenorhabditis elegans Are Located in Alternatively Spliced Exons

Teresa M. Rogalski, Erin J. Gilchrist, Gregory P. Mullen and Donald G. Moerman

Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

Manuscript received July 25, 1994
Accepted for publication September 24, 1994

ABSTRACT

The unc-52 gene in Caenorhabditis elegans produces several large proteins that function in the basement membrane underlying muscle cells. Mutations in this gene result in defects in myofilament assembly and in the attachment of the myofilament lattice to the muscle cell membrane. The st549 and ut111 alleles of unc-52 produce a lethal (Pat) terminal phenotype whereas the e444, e669, e998, e1012 and e1421 mutations result in viable, paralyzed animals. We have identified the sequence alterations responsible for these mutant phenotypes. The st549 allele has a premature stop codon in exon 7 that should result in the complete elimination of unc-52 gene function, and the ut111 allele has a Tcl transposon inserted into the second exon of the gene. The five remaining mutations are clustered in a small interval containing three adjacent, alternatively spliced exons (16, 17 and 18). These mutations affect some, but not all of the unc-52-encoded proteins. Thirteen intragenic revertants of the e669, e998, e1012 and e1421 alleles have also been sequenced. The majority of these carry the original mutation plus a G to A transition in the conserved splice acceptor site of the affected exon. This result suggests that reversion of the mutant phenotype in these strains may be the result of exon-skipping.

The unc-52 gene in Caenorhabditis elegans produces several large proteins that are homologous to perlecans, the mammalian basement membrane heparan sulfate proteoglycan (Noonan et al. 1991; Kallunki and Truvnassson 1992; Murdoch et al. 1992). Perlecans are synthesized by a wide variety of cell types and has been found in all vertebrate basement membranes examined (Martin and Timpl 1987). The UNC-52 proteins function in the basement membranes of muscle cells (Francis and Waterston 1991; Rogalski et al. 1993). These extracellular proteins play an important role in directing myofilament assembly and in maintaining the ordered structure of the myofilament lattice (MacKenzie et al. 1978; Waterston et al. 1980; Gilchrist and Moerman 1992; Rogalski et al. 1993; Williams and Waterston 1994; Hresko et al. 1994).

In the body-wall muscle of C. elegans the myofilament lattice lies just beneath the cell surface facing the hypodermis and is anchored to the basement membrane through a series of lateral attachments (Waterston et al. 1980; Francis and Waterston 1985, 1991; Waterston 1988). The thick filaments of the lattice are held in register by the proteins of the M-line analogues, whereas the thin filaments are anchored by dense bodies and attachment plaques (Waterston et al. 1980). A transmembrane complex links the dense bodies and M-line analogues to the basement membrane (Francis and Waterston 1985), which is interposed between the muscle cells and the overlying hypodermis. The unc-52 encoded components of the basement membrane are concentrated over the dense bodies and M-lines (Francis and Waterston 1991). These components play a role in muscle attachment late in development (MacKenzie et al. 1978; Waterston et al. 1980) and may also act as extracellular cues, or anchors, for the assembly of nascent sarcomeres during embryogenesis (Rogalski et al. 1993; Hresko et al. 1994; Williams and Waterston, 1994).

The unc-52 gene consists of 26 exons and produces a number of alternatively spliced transcripts (Rogalski et al. 1993). The longest potential open reading frame encodes a 2483 amino acid protein with a signal peptide and four recognizable domains. The first domain is unique to the unc-52 polypeptide whereas the three remaining domains contain sequences found in the LDL-receptor (domain II), laminin (domain III) and N-CAM (domain IV). Two monoclonal antibodies (mAbs), MH2 and MH3 (Francis and Waterston 1991), recognize an epitope in domain IV of the unc-52 protein sequence (Rogalski et al. 1993). The proteins recognized by these mAbs are synthesized in embryogenesis (Rogalski et al. 1993; Hresko et al. 1994) and are present in the basement membranes of all contractile tissues in adult hermaphrodites (Francis and Waterston 1991; G. P. Mullen and D. G. Moerman, unpublished results).
The class 1 mutations are clustered in a small interval responsible for the different unc-52 mutant phenotypes. However, these mutants are not paralyzed. The results of these stages of myofilament lattice assembly (ROGALSKI et al. subset of the unc-52-encoded proteins. The paralyzing phenotype is due to the gradual disruption of the body-wall muscle cells posterior to the pharynx which begins at the third or fourth larval stage (Mackenzie et al. 1978; Waterston et al. 1980). The dense bodies in affected cells are fractured and the myofilaments are no longer anchored to the cell membrane (Waterston et al. 1980). The class 2 mutation, st49, produces a Pat (paralyzed, arrested elongation at twofold) terminal phenotype (Rogalski et al. 1993; Williams and Waterston 1994). Muscle cells in these mutants have no organized A or I bands, suggesting a failure in the earliest stages of myofilament lattice assembly (Rogalski et al. 1993; Hresko et al. 1994; Williams and Waterston 1994). Embryos homozygous for the class 3 allele, ut111, also arrest development at the “twofold” stage; however, these mutants are not paralyzed. The results of complementation tests revealed that the class 3 allele and the class 1 alleles complement (K. Kondo and I. Katsura, personal communication; Gilchrist and Moerman 1992).

In this paper we describe the molecular alterations responsible for the different unc-52 mutant phenotypes. The class 1 mutations are clustered in a small interval containing three adjacent, alternatively spliced exons (16, 17 and 18); the class 2 allele carries a premature translational stop codon in exon 7; and the class 3 allele has a Tct transposon inserted into the second exon of the gene. Using a mAb specific for an exon 19-encoded epitope we show that the class 1 mutations affect only sequence thirteen intragenic revertants of several class 1 alleles (GILCHRIST and MOERMAN 1992). The paralyzed phenotype is due to the gradual disruption of the myofilament lattice assembly in the body-wall muscle cells. The various unc-52 alleles have been divided into three groups based on their mutant phenotypes and complementation patterns (GILCHRIST and MOERMAN 1992). Mutants homozygous for class 1 alleles are viable. They are able to move normally as larvae, but adult hermaphrodites are paralyzed, thin and partially egg-laying defective (Brenner 1974; Gilchrist and Moerman 1992). Muscle cells in these mutants have no organized A or I bands, suggesting a failure in the earliest stages of myofilament lattice assembly (Rogalski et al. 1993; Hresko et al. 1994; Williams and Waterston 1994). Embryos homozygous for the class 3 allele, ut111, also arrest development at the “twofold” stage; however, these mutants are not paralyzed. The results of complementation tests revealed that the class 3 allele and the class 1 alleles complement (K. Kondo and I. Katsura, personal communication; Gilchrist and Moerman 1992). MATERIALS AND METHODS

Nematode strains and culture conditions: Nematode strains were grown and maintained on NG agar plates as described in Brenner (1974). Strains used in this work include the wild-type strain N2 (Brenner 1974); CB444, unc-52(e444); CB669, unc-52(e669); CB998, unc-52(e998); CB1012, unc-52(el012); CB1421, unc-52(e1421); HE250, unc-52(e669ut250); DM1173, unc-52(e998ut1); DM1181, unc-52(e669ut12); DM1182, unc-52(e1012ut1); DM1183, unc-52(e1012ut1); DM1183, unc-52(e1012ut1); DM1184, unc-52(e1012ut1); DM1302, unc-52(e111)/unc-52(e998) and RW6010, unc-52(st549)/unc-52(st549)/unc-52(st549) (Williams and Waterston 1994). The mdp34 duplication used to balance the st549 mutation carries a wild-type copy of the unc-52 gene (Herman et al. 1979).

PCR amplification of genomic DNA: The PCR reactions were performed as described by Barstead et al. (1991). Several (2–10) adult hermaphrodites or mutant embryos were placed in 3 μl lysis buffer (50 mM KCl, 10 mM Tris (pH 8.0), 2.5 mM MgCl2, 0.45% Tween 20, 0.45% NP-40, 60 μg/ml Proteinase K) in the lid of a 0.5 μl microcentrifuge tube. A drop of mineral oil was added to the tube and the contents were briefly spun in a microcentrifuge. The tubes were incubated at 60° for 45 min to digest the worms and then at 95° for 15 min to inactivate the Proteinase K. To amplify the genomic DNA, 25 pmol of forward primer, 25 pmol of reverse primer, 2.5 μl 10X PCR buffer (Promega; 15 mM MgCl2), 2.0 μl dNTP mix (Perkin-Elmer, 10 mm solution), 2.5 units of Taq polymerase (Promega), and dH2O were added to each tube for a final volume of 25 μl. The mixture was amplified using a Perkin-Elmer Cetus DNA thermocycler. Amplification conditions were 30 sec at 95°, 60 sec at 55° and 90 sec at 72° for 35 cycles. This was followed by 5 min at 72°. Eight DNA fragments were amplified from st549/st549 mutant embryos using the following primer sets: (1) CAAATGGCTCGTCTGAGTTT (5’ upstream) and GGAGTCCGGACAGACTCTCT (exon 3); (2) CGCGTGGCTACATGTTAG (exon 2) and CGTGAACATGCCCTGCA (exon 4); (3) AACGACAGTGACAGCAAG (exon 4) and CAAAGACTCTGCGTGGCT (exon 5); (4) TGCTGCTGCTTGGGCT (5’ upsteam) and TTCTAACAGGGAGCTGTCACACG (intron 4) and TTCTAACAGGGAGCTGTCACACG (exon 6); (5) GACACACAAATGGCCTAG (intron 4) and AAAGTGTGACGCTCCCTCCTT (exon 6); (5) TGAGTACCGCATCGG (exon 5) and CGATCCCGAGAGCTATAATC (exon 7); (6) CAACTAAACCAAGGATG (exon 6) and CCAACATCCGTGGCTTGCA (exon 9); (7) AGAACCCGCCAGGCGTCTTCA (exon 9) and CCAACGACACTCTCAGTGCT (exon 11). Six DNA fragments were amplified from st549/st549 mutant embryos using the following primer sets: (1) CAAATGGCTCGCTGAGTTT (5’ upstream) and TTCTGCACACTGGAGGGA- ACTCT (exon 2); (2) same upstream primer as fragment 1 and GGAGTCCGGACAGACTCTCT (exon 3); (3) CGCGTGGCTACATGTTAG (exon 1) and same downstream primer as fragment 2; (4) same upstream primer as fragment 3 and TGCTGAAGCTGGCTCTCTCC (exon 4); (5) GAACACACTGCTGAACTTGCC (Tcl 1) and same downstream primer as fragment 2; (6) same downstream primer as fragment 5 and same downstream primer as fragment 4. Four DNA fragments were amplified from the class 1 mutants and revertants using the following primer sets: (1) TTATCGTCAAGTCGCGTGAG (exon 13) and TGGGCTGCA (exon 15) and TGGGTCCAGCATGTTCTCCGGA (exon 18); (2) GACATCCAGATGTCCATTAG (exon 16) and same downstream primer as fragment 1; (3) same upstream primer as fragment 2 and AACTGATGTCGCCCGTCAAC (exon 19); (4) GTGATT- CAGTGATACATGA (exon 17) and CCGTGGCAGAAGTCC- GAAAT (exon 19).

Reverse transcription: RNA was isolated from mixed populations of worms using the technique developed by Zarkower and Hodgkin (1992). Worms were washed off of two or three 60-mm plates with M9 buffer into 1.5-ml microfuge tubes and allowed to settle on ice. The M9 buffer was removed and 50 μl of Proteinase K buffer (50 mM NaCl, 50 mM Tris pH 7.5, 5 mM EDTA, 0.5% SDS and 200 μg/ml Proteinase K) was added. The tubes were incubated at 60° for 30–60 min to allow the worms to digest and then 150 μl of solution A (10 mM EDTA pH 8.0; 0.5% SDS) and 200 μl of solution B (10 mM EDTA pH 8.0; 160 mM NaOAc) were added. Nucleic acid was then extracted (3×) with phenol/chloroform. The aqueous layer was transferred to a 1.5-ml tube containing 44 μl ice cold 1 mM Tris, pH 7.5, and 18 μl 5 M NaCl, and the...
RNA was precipitated with two volumes of ethanol. The RNA was resuspended in 20–30 μl H2O and stored at −20°. Reverse transcription was performed with 5 μl of the resuspended RNA as previously described (ROGALSKI et al. 1993). The procedure for amplifying the cDNA products is also described in ROGALSKI et al. (1993). The downstream primers used in the reverse transcription step were CCCGG (exon 15) and GACATCCAAGTG'TCAGC (exon 16).

Sequence analysis: The PCR-amplified genomic and cDNA fragments were not gel-purified. They were sequenced directly using the BRL dsDNA Cycle Sequencing System and unc-52 primers. A sequencing reaction mix consisted of 1–4 μl of the PCR reaction, 4.5 μl 10× Taq sequencing buffer (BRL), 0.5 μl Taq polymerase (BRL or Promega), 5 μl 10× end-labeled primer (1 pmol) and H2O to a total volume of 36 μl. Four 0.5-μl eppendorf tubes were labeled C, T, A or G, and 2.0 μl of the appropriate termination mix (BRL) and 8 μl of the sequencing reaction mix were added to each tube along with one drop of mineral oil. The sequencing reactions were performed in a Perkin-Elmer Cetus DNA thermocycler. The conditions used were 30 sec at 95°, 30 sec at 55° and 60 sec at 70° for 30 cycles and then 30 sec at 95° and 60 sec at 70° for 10 cycles. At the end of the 30 cycles 5 μl of stop buffer (BRL) was added to each tube. The reactions were heated to 95° for 5 min before loading onto a standard 6% acrylamide sequencing gel. The 1.2-kb genomic containing exons 16, 17 and 18 was completely sequenced for all populations of embryos from the N2, CB444, CB669, CB998, and CB1012 and DM102 strains. Embryos were fixed and stained using methods described previously (GOH and BOGAERT 1991; ROGALSKI et al. 1993). Briefly, embryos were released from adult hermaphrodites by treatment with a mixture of hypochlorite and NaOH, fixed in 3% formaldehyde followed by 100% methanol and stained with the MI2 mAb. The secondary antibody was Texas red–conjugated goat anti-mouse IgG.

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 200-nm intervals and combined using the PROJECT function. For reproduction, image files were transferred to a Macintosh computer, converted to TIFF file format using NIH Image 1.54 and arranged and annotated using Adobe Photoshop 2.5. The computer images were printed on a Linotronic 3300 printer.

For polarized light microscopy, live worms were viewed following the procedure described by WATERSTON et al. (1980). Photomicrographs were prepared using Kodak TMAX400 film on a Zeiss Axiophot Photomicroscope (Carl Zeiss D-7082 OberKochen).

RESULTS

The class 2 allele represents the null state of unc-52: The class 2 mutation, st549, was identified by directly sequencing PCR-amplified DNA from homozygous mutant embryos. We began sequencing at the 5′ end of the gene ~100 bp upstream of the putative initiator methionine and continued downstream until the first sequence alteration, a G to A transition, was identified (Table 1). This mutation introduces a premature termination codon into exon 7 that encodes part of the second laminin-like repeat of domain III (Figure 1). Since exon 7 is present in all unc-52 transcripts detected (ROGALSKI et al. 1993), the st549 allele should result in the complete elimination of unc-52 gene function.

The class 3 allele has a Tc1 insertion: The lethal phenotype produced by the class 3 allele, utl11, is due to the insertion of a Tc1 transposon into exon 2, which encodes the first immunoglobulin repeat (IgR1) of the UNC-52 proteins (Table 1; Figure 1). The location of the Tc1 element was determined by amplifying DNA from animals of genotype unc-52(e998)/unc-52(utl11) using primers that generated PCR fragments covering the entire gene. One set of primers produced a second, larger fragment in addition to the wild-type–sized fragment that was expected (data not shown). Further analysis of this region using an upstream primer that recognized the Tc1 element and a downstream primer from unc-52, demonstrated that the utl11 mutation was caused by the insertion of a Tc1 element into exon 2 (Figure 1). DNA from e998/e998 or wild-type animals did not produce a PCR product with these primers, whereas DNA from utl11 homozygous mutants consistently produced a 220-bp fragment (data not shown). The insertion site of the transposon was identified by sequencing this PCR product (Table 1).

The class 1 mutations are clustered in alternatively spliced exons: We have previously shown that the class
TABLE 1
Sequence alterations in unc-52 mutants

<table>
<thead>
<tr>
<th>unc-52 allele</th>
<th>Phenotypic class</th>
<th>Position*</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ut111</td>
<td>3</td>
<td>Exon 2</td>
<td>Tcl insertion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2228)</td>
<td>CATACA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exon 7</td>
<td>Termination codon</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5730)</td>
<td>TGG−TGA</td>
</tr>
<tr>
<td>st549</td>
<td>2</td>
<td>Exon 16</td>
<td>Splice donor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(19647)</td>
<td>/gtagc−/gtaa</td>
</tr>
<tr>
<td>e1421</td>
<td>1</td>
<td>Exon 17</td>
<td>Termination codon</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(10769)</td>
<td>CAG−TAG</td>
</tr>
<tr>
<td>e669</td>
<td>1</td>
<td>Exon 17</td>
<td>Termination codon</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(10829)</td>
<td>CAG−TAG</td>
</tr>
<tr>
<td>e1012</td>
<td>1</td>
<td>Exon 18</td>
<td>Termination codon</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(11254)</td>
<td>CGA−TGA</td>
</tr>
<tr>
<td>e444</td>
<td>1</td>
<td>Exon 18</td>
<td>Termination codon</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(11301)</td>
<td>TGG−TGA</td>
</tr>
<tr>
<td>e998</td>
<td>1</td>
<td>Exon 18</td>
<td>Tcl insertion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(11396)</td>
<td>AATACG</td>
</tr>
</tbody>
</table>

*Nucleotide numbering (in parentheses) is from the unc-52 genomic sequence, Genbank accession number L13458. The / indicates an intron/exon boundary; upper case letters represent exon sequences and lower case letters represent intron sequences. The nucleotides altered by the mutations are in bold and are underlined.

*This second alteration changing a threonine codon (AGG) to a lysine codon (AAG) was detected in this strain at position 19,935 in exon 17. As this alteration is also present in all of the e998 revertants including the m38 deletion, it cannot be responsible for the class 1 mutant phenotype.

*This mutation was identified by Rogalski et al. (1993).

1 mutant phenotype produced by the st196 allele is the result of a Tcl-transposon insertion into the exon encoding IgR10 of domain IV (Rogalski et al. 1993; Figure 1). We have identified the sequence alterations in five additional class 1 alleles by directly sequencing PCR-amplified DNA from homozygous mutant hermaphrodites. Since the st196 mutation was found in exon 18, we began our analysis of the other class 1 alleles by sequencing this exon, as well as the two adjacent, upstream exons. Five mutations were identified in this 1.2-kb interval. Four of these are point mutations that introduce a stop codon into either exon 17 (e669 and e1012) or exon 18 (e444 and e998), whereas the fifth, e1421, is a point mutation that alters the splice donor site of exon 16 (Table 1; Figures 1 and 2).

All six of the class 1 mutations that we have identified are located in alternatively spliced exons. Each of these alterations will affect some, but not all, of the UNC-52 proteins. Our earlier analysis of unc-52 cDNA clones (Rogalski et al. 1993) revealed that exons 16 and 17 are removed from one of the mRNA transcripts (Figure 2). We have examined splicing in this region of the unc-52 gene in more detail by generating additional cDNAs using reverse transcription and PCR. The cDNA fragments in Figure 2 were obtained from wild-type RNA using various primers from exons 15 through 19, and the alternative splicing patterns shown were confirmed by sequencing each fragment. This analysis reveals that the three exons affected by the class 1 mutations are all alternatively spliced and that a minimum of six different unc-52 mRNAs are apparently generated by these splicing events (Figure 2). These three exons each encode one complete immunoglobulin repeat in domain IV (IgR8, IgR9 and IgR10 in Figure 1). The relative abundance of the different mRNAs has not be determined.

An exon 19-encoded epitope recognized by the MH2 mAb is present in class 1 and class 3 unc-52 mutant embryos: The MH2 mAb described by Francis and Waterston (1991) recognizes an epitope in domain IV of the unc-52-encoded protein sequence (Rogalski et al. 1993). On Western blots, MH2 specifically recognizes the unc-52-encoded portion of the fusion product synthesized by the pGEX clone DM#178 (Rogalski et al. 1993; Figure 3). We have further refined the position of the MH2 epitope by expressing smaller regions of the unc-52-encoded protein sequence (ROGAISKI et al. 1993). An incomplete immuno globulin repeat in domain IV (IgR8, IgR9 and IgR10 in Figure 1). The relative abundance of the different mRNAs has not be determined.

Previous studies demonstrated that MH2 stains the dorsal and ventral muscle quadrants in wild-type embryos, but fails to stain embryos homozygous for the st196 allele of unc-52 (Rogalski et al. 1993; Hresko et al. 1994). We have stained class 1 and class 3 embryos with this mAb and the results obtained are presented in Figure 4. The class 1 mutants (4b and 4c) exhibit the same staining intensity and pattern as the wild-type embryo (4a). The embryo in Figure 4b is homozygous for the e1012 mutation in exon 17 and the embryo in Figure 4c is homozygous for the e444 mutation in exon 18. The e669 and e998 mutants also exhibit the wild-type MH2 staining pattern at this early developmental stage (data not shown). The fact that class 1 mutant
Mutations in *C. elegans* unc-52

Figure 1.—Localization of mutations within the *unc-52* gene. The physical map of the *unc-52* gene represents 16.5 kb of sequenced genomic DNA and shows the positions of 26 exons (■). The exons affected by the *ut111, st549, e1012, e444, e998* and *st196* mutations are indicated by asterisks. Also presented is a diagram of the largest predicted UNC-52 protein showing the organization of the different domains. Domain I is a short, acidic region located at the amino terminus (solid area). Domain II contains three copies of a cysteine-rich sequence that occurs in the amino-terminal ligand-binding domain of the LDL-receptor protein (crosshatched areas). Domain III contains sequences similar to the amino-terminal region of laminin. This domain contains two copies of a larger cysteine-free globular sequence (●) and seven copies of a cysteine-rich element (vertical stripes). Domain IV consists of 14 immunoglobulin C2-like repeats (shown hatched and numbered 3–16). There are also two of these repeats flanking domain II (numbered 1 and 2). For a more detailed description of the protein domains see Rogalski et al. 1993.

Embryos stain with MH2, a mAb specific for an epitope in exon 19, confirms our earlier conclusion that mutations in exons 17 and 18 do not disrupt all of the domain IV containing *unc-52* proteins. The presence of MH2 reactive material in class 1 mutants indicates that at least some of the alternatively spliced mRNAs we have detected by RT-PCR are translated.

The class 3 (*ut111::Tc1*) embryo is shown in Figure 4d. Antibody staining is detected in this mutant; however, it is greatly reduced relative to wild type and appears disorganized and patchy. The identification of UNC-52 protein in *ut111* embryos is consistent with the observation that *ut111* complements all class 1 mutations.

Intragenic revertants of class 1 mutations alter intron/exon boundaries: Intragenic revertants of four different class 1 *Unc-52* mutants have been isolated as part of a genetic characterization of the *unc-52* locus (Gilchrist and Moerman 1992). These rare individuals were obtained from the progeny of mutagenized class 1 hermaphrodites, and the mutations responsible for their revertant phenotypes were shown to be closely linked to the *unc-52* locus. We have determined that 12 of these revertants carry a second mutation within the *unc-52* gene that results in improved body-wall muscle organization.

Sequencing PCR-amplified DNA from revertant strains revealed two types of reversion events (Table 2). One, *ra38*, is a small in frame deletion that removes the site of the *e998* mutation (Figure 5). The endpoints of this deficiency are located in exons 17 and 18, and both breaks occur within the same 5-bp sequence in each exon (see Table 2). This finding suggests that *ra38* may have resulted from an unequal crossover event. The 11 other strains examined still carry the original class 1 mutation (either *e669, e998, e1012* or *e1421*) plus an additional alteration. In nine of these we detected a G to A transition at the conserved splice acceptor site of the affected exon (Table 2; Figure 5). The two exceptions were the *e1012* revertants. Although *ra15* and *ra16* alter a splice acceptor site, it is not the splice acceptor of the exon containing the *e1012* mutation, but instead the splice acceptor for the upstream exon (Table 2; Figure 5). Preliminary results from RT-PCR experiments suggest that transcripts carrying class 1 mutations are never (or only rarely) made in these revertant strains (data not shown). This result implies that mutations in the splice acceptor sites of exons 16, 17 and 18
DISCUSSION

The *unc-52* gene in *C. elegans* produces a number of different polypeptides as a result of the alternative splicing of at least 6 of its 26 exons. We have shown that transcripts encoding proteins with three different carboxy-terminal amino acid sequences can be generated by the utilization of two different poly(A) addition sites and by the differential splicing of exons 21 and 22 (ROGALSKI et al. 1993). All of the exons that encode domain IV of the protein sequence are located between the two poly(A) addition sites (see Figure 1). Thus, at least one *unc-52* transcript encodes a shortened polypeptide completely lacking domain IV (ROGALSKI et al. 1993). In addition, exon 6, which encodes the first laminin-like subdomain of domain III, is not present in all *unc-52* mRNAs (ROGALSKI et al. 1993). The analysis of the differential splicing of exons 16, 17 and 18 reported here has identified at least six transcripts that encode proteins containing some portion of domain IV. The size and complexity of the *unc-52* mRNAs make it difficult to determine the number of polypeptides produced by this gene. To address this question, we are preparing a number of exon specific antisera that should permit us to identify all of the *unc-52* gene products (G. P. MULLEN and D. G. MOERMAN, unpublished results). From our studies thus far we can conclude that alternative splicing leads to proteins with variability in domains III and IV and to proteins with different carboxy-terminal amino acids.

A total of eight *unc-52* mutations have been identified. Two of these result from Tc1 insertions, and five introduce premature translational stop codons into the...

**Figure 2.**—Alternative splicing of exons 16, 17 and 18. The top line represents genomic DNA and shows the location of the six class 1 *unc-52* mutations. Partial cDNAs representing alternatively spliced transcripts are shown below. All of the cDNAs shown here were generated by reverse transcription and PCR and have been sequenced. The cDNA indicated by an asterisk had been identified previously (ROGALSKI et al. 1993). Exons 16, 17 and 18 each encode a single immunoglobulin repeat (see Figure 1), and any or all can be removed from a transcript without altering the reading frame of the message.

may prevent these exons from being included in any *unc-52* transcript. We have also sequenced PCR-amplified DNA from animals carrying *su250*, a spontaneous, intragenic revertant of *e669* (MACKENZIE et al. 1978). The only alteration in the sequenced region, aside from the *e669* point mutation in exon 17, is a single base change in the adjacent, upstream intron (Table 2; Figure 5). The localization of the intragenic reversion events to the same exons as the class 1 mutations provides conclusive evidence that we have identified the sequence alterations responsible for the class 1 mutant phenotypes.

All of the reversion events described above were selected because they result in animals that no longer exhibit the paralysis typical of adult class 1 animals. The *e1421ra37, e669ra12* and *ra38* revertants move as well as wild-type animals, and their body wall muscle appears essentially normal when viewed with polarizing optics. In contrast, *e1012ra16, e669su250* and *e998ra1* animals do not exhibit wild-type movement, and their muscle cells contain fewer A-bands than do muscle cells in wild-type animals of a comparable age. In wild-type hermaphrodites each cell normally contains 7–10 sarcomeres (WATERSTON 1988; Figure 6a), but in these three *unc-52* intragenic revertants only 4–6 sarcomeres could be detected in most muscle cells (Figure 6b). In addition, some myofilament disorganization was observed (Figure 6c) and the muscle was easily disrupted when the animals were manipulated for microscopic observation. Adult *e1012ra16* hermaphrodites grown at 15° move very well, whereas the *e669su250* and *e998ra1* animals are much slower. In fact, most of the older *e998ra1* adults eventually become paralyzed at this temperature. When grown at 25°, all three strains exhibit a paralyzed phenotype similar to that observed in class 1 mutants. In contrast, the *e1421ra37, e669ra12* and *ra38* revertant animals move well at either temperature.
Mutations in C. elegans unc-52

Figure 3.—MH2 recognizes an antigen encoded by exon 19 of the unc-52 gene. (a) Diagrammatical representations of the partial polypeptides expressed as GST-fusions by the pGEX clones DM#178 (ROGALSKI et al. 1993), DM#181, DM#182, DM#183 and DM#190. The DM#178 clone contains a 1.47-kb alternatively spliced cDNA fragment that encodes IgR6, IgR7, IgR10, IgR11, IgR12 and IgR13 in domain IV of the unc-52 protein sequence (■). The other four clones encode portions of this region (indicated by horizontal lines). (b) Affinity purified GST-Unc-52 fusion proteins from DM#178 (lane 1), DM#185 (lane 2), DM#182 (lane 3), DM#181 (lane 4) and DM#190 (lane 5) were separated on an 8% acrylamide gel, transferred to Hybond-ECL nitrocellulose and immunoreacted with an anti-GST sera or the MH2 mAb. MH2 specifically recognizes the fusion products synthesized by DM#178, DM#182 and DM#190. The vertical lines in (a) define the boundaries of the immunoreactive region.

protein sequence. The remaining mutation, el421, alters a conserved base in the splice donor site of an intron. This same G to A transition in an intron of the Adenovirus E1A gene leads to the complete abolition of splicing at this site (SOLNICK 1981). This occurs because the U1 snRNA can no longer pair efficiently with the mutated splice site (ZHANG and WEINER 1986). C. elegans U1 snRNA is similar in structure to that of other eukaryotes (THOMAS et al. 1990) and includes a 5' end that is complementary to the 5' splice site consensus sequence (FIELDS 1990). Therefore, we expect that the el421 alteration reduces or abolishes splicing to the 3' end of exon 16. The e669 amber mutation in exon 17 is suppressible by the tRNA<sup>sup</sup> amber suppressors, sup-5 and sup-7 (WATERSTON 1981; WILLS et al. 1983). However, the e1012 amber mutation that is also in exon 17 is not. The relative positions of these two residues in the UNC-52 protein may explain this difference in behavior. The e669 mutation is within the linker region between two immunoglobulin-like repeats whereas the

Figure 4.—Immunofluorescence antibody staining of wild-type and unc-52 mutant embryos visualized by confocal microscopy. (a) Wild-type, (b) unc-52(e1012), (c) unc-52(e444) and (d) unc-52(ut111) mutant embryos stained with the MH2 mAb. Each image is a projection of a series of optical sections, allowing all four muscle quadrants to be viewed at once. MH2 stains the dorsal and ventral muscle quadrants in the wild-type and class 1 mutant embryos. a, b, c, d: dorsal left, ventral left, dorsal right and ventral right, respectively. MH2 staining in the class 3 (ut111) mutant is disorganized, and the intensity is reduced. The intensity of the images was adjusted for reproduction and should not be regarded as quantitative. Bar, 10 μm.
e1012 mutation is adjacent to the first conserved cysteine in an immunoglobulin-like repeat (Rogalski et al. 1993). The amino acid substituted by these tRNA<sup>am</sup> amber suppressors may be incompatible with this latter site.

The class 2 unc-52 mutation, st549, introduces a termination codon into exon 7. Embryos homozygous for this mutation fail to stain with the MH2 mAb (Rogalski et al. 1993) and therefore, do not produce any of the larger proteins. These mutants also fail to stain with a polyclonal antibody that recognizes sequences that should be present in all of the UNC-52 proteins (G. P. Mullen and D. G. Moerman, unpublished results). Thus, st549 is a null mutation, and the Pat terminal phenotype of this allele results from the complete absence of UNC-52 protein. Hresko et al. (1994) have recently shown that st549 mutant embryos exhibit defects in the localization of β-integrin and vinculin, components of the dense bodies and M-lines in developing body wall muscle cells. This finding reveals that some UNC-52 isoforms are required for the initiation of myofilament assembly.

The class 3 allele, ut111, produces a lethal terminal phenotype that is similar to the null phenotype of unc-52. However, in contrast to the st549 mutants, ut111 embryos exhibit staining with the MH2 antibody. Thus, ut111::Tc1 is not a null mutation, but rather a severe hypomorph. The residual protein activity present in this mutant may explain the ability of ut111 to complement class 1 alleles. It may also explain why the ut111 terminal phenotype is more similar to a mild Pat phenotype than to the severe Pat phenotype of the st549 mutant (Williams and Waterston 1994). There are at least three possible explanations for the leaky nature of ut111 Tc1 mutants. First, full-length proteins could be synthesized by the class 3 mutant if the exon carrying the Tc1 element is alternatively spliced. Exon 2, which encodes the first immunoglobulin repeat (IgR1), could be spliced out of a message without altering the reading frame. However, as we have been unable to detect any alternative splicing in this region of the unc-52 gene in wild-type hermaphrodites this seems unlikely (E. J. Gilchrist, T. M. Rogalski and D. G. Moerman, unpublished results). Somatic excision of the Tc1 element, either at the DNA or RNA level could also explain the presence of full-length UNC-52 protein in this mutant. Somatic excision of Tc1 from the genome occurs at a very low frequency (Emmons et al. 1983), which means that only a few body wall muscle cells should have detectable protein (see Eide and Anderson 1985; Moerman et al. 1988). Our observations are not consistent with this explanation since we see UNC-52 protein and its export to adjacent to many muscle cells in all four muscle

### TABLE 2

<table>
<thead>
<tr>
<th>unc-52 allele</th>
<th>Intragenic revertant</th>
<th>Site&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>e998&lt;sup&gt;a&lt;/sup&gt;</td>
<td>ra38</td>
<td>Exon 17 (11050)</td>
<td>311-bp deletion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGAAC&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exon 18 (11341)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ra1, ra3, ra4, ra6, ra7</td>
<td>Splice acceptor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ttcag/-ttcag/</td>
<td></td>
</tr>
<tr>
<td>e669</td>
<td>ra12</td>
<td>Exon 17 (10760)</td>
<td>Splice acceptor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ttcag/-ttcag/</td>
<td></td>
</tr>
<tr>
<td>su250</td>
<td></td>
<td>Intron 16 (10692)</td>
<td>?</td>
</tr>
<tr>
<td>e1012</td>
<td>ra15, ra16</td>
<td>Exon 16 (10345)</td>
<td>Splice acceptor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gcaagg/cgag</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ttcag/-ttcag/</td>
<td></td>
</tr>
<tr>
<td>e1421</td>
<td>ra19, ra22, ra37</td>
<td>Exon 16 (10345)</td>
<td>Splice acceptor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ttcag/-ttcag/</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Nucleotide numbering (in parentheses) is from the genomic sequence of unc-52, Genbank accession number L13458. The / indicates an intron/exon boundary; upper case letters represent exon sequences and lower case letters represent intron sequences. The nucleotides altered by the mutations are in bold and are underlined.

<sup>1</sup>The e998 revertants retain the C to A base change in exon 17 that is present in the original e998 mutant (see Table 1). The ra38 deletion revertant also carries a G to A transition at position 11,356 in exon 18. This would change an Asp codon (GAT) to an Asn codon (AAT).

<sup>2</sup>The sequence of the 5-bp repeat in exons 17 and 18 where the ra38 deletion endpoints occur.

![FIGURE 5.—Intragenic revertants of class 1 unc-52 mutants](image-url)
The muscle cell in Fig. 6(a) is intact, but has fewer mutant hermaphrodites. Polarized light micrographs of wild-type muscle cell in Fig. 6(b) and (c) body-wall muscle. The arrows point to dense hemicords and the arrowheads point to A bands. Bar, 5 μm.

![Figure 6](image-url)

Mutations in the unc-52 gene produce a very mild phenotype compared with that produced by the null allele (Gilchrist and Moerman 1992). Embryogenesis in these mutants appears to be normal, and the mutant embryos show the wild-type pattern of staining with the MH2 antibody. The class 1 mutations are clustered in three adjacent exons that are alternatively spliced in wild-type hermaphrodites. Each of these exons encodes a complete immunoglobulin repeat (IgR8, IgR9 and IgR10), and any or all can be removed from a transcript without altering the reading frame of the message. All six class 1 alleles are completely recessive to wild type, indicating that they do not produce truncated proteins that interfere with wild-type gene function. Our analysis suggests that these mutations eliminate some, but not all, of the domain IV containing proteins, and this is confirmed by our finding that class 1 mutant embryos express the MH2 epitope. We suspect that mutations in other constitutive regions of this domain would be lethal. The Unc phenotype produced by the class 1 mutations could be due to the absence of specific proteins, or simply to a reduction in the total amount of UNC-52 protein. Most class 1 alleles manifest their phenotype during larval development when there is an increase in muscle cell mass. The surface-to-volume ratio in muscle cells is altered dramatically and this leads to increased tension on the adhesion complexes of muscle and on muscle/hypodermal attachment points. We believe that the overall level of domain IV containing proteins is critical for the assembly of new sarcomeres and for maintaining muscle attachment during this period of postembryonic growth.

We previously demonstrated that the class 1 unc-52 mutant phenotype can be suppressed by mutations in an extragenic suppressor locus, sup-38, or by intragenic reversion events (Gilchrist and Moerman 1992). We have analyzed 13 intragenic events and have found that 11 facilitate suppression by altering the invariant AG at the splice acceptor site of either the exon harboring the initial mutation or, in the case of the e1102 revertants, the adjacent, upstream exon. The dinucleotide AG is an invariant feature of eukaryotic splice acceptor sites (Mount 1982) and has been found in all C. elegans introns analyzed to date (Fields 1990). Altering the splice acceptor site can lead to one of several consequences at the transcript level. The affected intron may remain unspliced (Cladaras et al. 1987), the splicing machinery may utilize a cryptic AG in the vicinity of the mutated AG (Aebi et al. 1986; Kupper et al. 1988; Su and Lin 1990), or the exon with the altered site may be skipped (Tromp and Prokor 1988). Another alternative, the utilization of the mutated AA 3′ splice acceptor site, has recently been discovered in C. elegans (Aroian et al. 1993). Of these four possible consequences, only exon skipping can account for the suppression of the original mutations in the unc-52 revertants (see Figure 5). The mutation responsible for suppression of the e669 class 1 phenotype in e669su250 animals appears to be a base change in the adjacent, upstream intron. Although the su250 alteration does not affect a known functional region within this intron, it seems likely that this revertant also affects splicing of the mutated exon. Exon skipping as a form of intragenic suppression should be relatively rare as it requires redundancy of functional and structural information in contiguous in-frame exons. This particular region of the unc-52 gene clearly meets these requirements.
The \textit{e1421ra37}, \textit{ra38} and \textit{e669ra12} revertant animals have virtually wild-type muscle structure and move well suggesting that enough UNC-52 protein is provided to maintain muscle attachment during postembryonic growth. In contrast, the \textit{e1012ra16}, \textit{e669ra250} and \textit{e998ra1} revertants are only partially suppressed. This weak suppression could be due to reduced levels of UNC-52 protein in these animals, or to the absence of a specific IgR necessary for muscle attachment. The fact that the \textit{ra38} deletion revertant is wild type leads us to suspect that the former explanation is perhaps more plausible. Also, reversion of the mutant phenotype in these partial revertants is temperature dependent, occurring only at low temperature (15–20\degreeCelsius). It may be that the splicing events occurring in these animals are less efficient at the higher temperature or, perhaps, it is the process of myofilament assembly and attachment that is somehow being affected.

The \textit{unc-52}-encoded proteins are homologues of perlecain, the mammalian basement membrane heparan sulfate proteoglycan (Noonan et al. 1991; Kallunki and Tryggvason 1992; Murdoch et al. 1992). The genomic sequence of the gene encoding the human perlecain core protein has recently been completed (Cohen et al. 1993), and similar to \textit{unc-52}, the immunoglobulin-like repeats that comprise domain IV of the human protein are dispersed across many exons. While the precise pattern of exon/intron junctions is not conserved between the human and nematode genes, the theme of multiple combinatorial immunoglobulin possibilities is conserved. This similarity in organization between these two phylogenetically disparate organisms suggests there are functional constraints on domain IV organization. Our study of \textit{unc-52} in \textit{C. elegans} illustrates what two of these functional constraints may be: initiation of myofilament assembly and maintenance of myofilament organization in muscle cells. How the regulation of a particular immunoglobulin repeat is related to these functions is a problem that remains to be resolved.

We thank Linda Matsuuchi, Mary Gilbert and Ken Norman for comments on the manuscript; Jason Bush for excellent technical assistance; Michael Weiss for his guidance with confocal microscopy and image processing; Ross Francis for the 8H2 mAb and Alice Rushforth and Philip Anderson for permission to cite unpublished results. We also thank Kazunori Kondo and I. Katsura for providing the \textit{utll} mutation and Ben Williams for providing the \textit{st549} mutation. Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center, which is funded by the National Institutes of Health National Center for Research Resources. This work was supported by grants from the Medical Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada to D.G.M.

\textbf{LITERATURE CITED}


Rogalski, T. M., B. D. Williams, G. P. Mullen and D. G. Moerman, 1993 The products of the \textit{unc-52} gene in \textit{Caenorhabditis elegans}
are homologous to the core protein of the mammalian basement membrane heparan sulfate proteoglycan. Genes Dev. 7: 1471–1484.


Communicating editor: R. K. Herman