Molecular and Genetic Analyses of the *Caenorhabditis elegans* *dpy-2* and *dpy-10* Collagen Genes: A Variety of Molecular Alterations Affect Organismal Morphology

Adam D. Levy, Jie Yang, and James M. Kramer

Department of Cell, Molecular, and Structural Biology, Northwestern University Medical School, Chicago, Illinois 60611

Submitted April 12, 1993; Accepted June 9, 1993

We have identified and cloned the *Caenorhabditis elegans* *dpy-2* and *dpy-10* genes and determined that they encode collagens. Genetic data suggested that these genes are important in morphogenesis and possibly other developmental events. These data include the morphologic phenotypes exhibited by mutants, unusual genetic interactions with the *sqt-1* collagen gene, and suppression of mutations in the *glp-1* and *mup-1* genes. The proximity of the *dpy-2* and *dpy-10* genes (3.5 kilobase) and the structural similarity of their encoded proteins (41% amino acid identity) indicate that *dpy-2* and *dpy-10* are the result of a gene duplication event. The genes do not, however, appear to be functionally redundant, because a *dpy-10* null mutant is not rescued by the *dpy-2* gene. In addition, full complementation between *dpy-2* and *dpy-10* can be demonstrated with all recessive alleles tested in trans. Sequence analysis of several mutant alleles of each gene was performed to determine the nature of the molecular defects that can cause the morphologic phenotypes. Glycine substitutions within the Gly-X-Y portion of the collagens can result in dumpy (Dpy), dumpy, left roller (DLRol), or temperature-sensitive DLRol phenotypes. *dpy-10(cn64)*, a dominant temperature-sensitive DLRol allele, creates an Arg-to-Cys substitution in the amino non-Gly-X-Y portion of the protein. Three *dpy-10* alleles contain Tc1 insertions in the coding region of the gene. *dpy-10(cg36)* (DLRol) creates a nonsense codon near the end of the Gly-X-Y region. The nature of this mutation, combined with genetic data, indicates that DLRol is the null phenotype of *dpy-10*. The Dpy phenotype results from reduced function of the *dpy-10* collagen gene. Our results indicate that a variety of molecular defects in these collagens can result in severe morphologic changes in *C. elegans*.

INTRODUCTION

Structural elements that can help determine organismal morphology may be found in the extracellular matrix (reviewed in Birk *et al.*, 1991; Hay, 1991; Toole, 1991). Components of extracellular structures interact with other extracellular elements and transmembrane factors in the determination of morphology. The proper assembly of extracellular structures is therefore crucial to the development of normal morphology. The analysis of the composition and assembly of the extracellular cuticle in the nematode *Caenorhabditis elegans* can be carried out at the genetic and molecular levels and provides the opportunity to assess the structure and func-

tion of individual components of an extracellular structure in vivo.

The cuticle is the collagenous exoskeleton of *C. elegans* that is shed and resynthesized at each of the four larval molts (Singh and Sulston, 1977; Cox *et al.*, 1981). The four larval and the adult cuticles can be shown to be distinct on the basis of structural, biochemical, and/or genetic criteria. Because the cuticle serves as the animal's exoskeleton, alterations in the cuticle might be expected to have drastic effects on the overall morphology of the animal. Mutations in over 40 genes have been described that cause gross alterations in the morphology of the animal (Brenner, 1974; Higgins and Hirsh, 1977; Cox *et al.*, 1980; Kusch and Edgar, 1986). These abnormal

morphologies fall into five classes: dumpy (short and fat), roller (helically twisted), long, squat (dominant roller alleles and recessive dumpy alleles), and blister (blisters on cuticle). Four such genes that affect organismal morphology have been isolated and shown to encode collagens: *sqt-1* (Kramer *et al.*, 1988), *dpy-13* (von Mende *et al.*, 1988), *rol-6* (Kramer *et al.*, 1990), and *dpy-7* (Johnstone *et al.*, 1993). Defective collagens are therefore able to cause these morphologic defects.

C. elegans contains between 50 and 150 collagen genes (Kramer *et al.*, 1982; Cox *et al.*, 1984, 1985). Over 30 of these genes have been isolated and, except for the two type IV collagen genes (Guo and Kramer, 1989), each has been shown to encode an ~30-kDa product that contains several short interruptions of the Gly-X-Y repeat domain (Kramer *et al.*, 1982, 1988, 1990; von Mende *et al.*, 1988; Cox *et al.*, 1989; Bird, 1992; Johnstone *et al.*, 1993). It is likely that most, or all, of these small collagens are localized to the cuticle. An antibody directed against the carboxyl-terminal domain of a *Haemonchus contortus* collagen that is homologous to *C. elegans* collagens *col-1* and *col-2* reacts with cuticles from *H. contortus* and *C. elegans* (Cox *et al.*, 1990). Also, antibodies specific for *sqt-1* and *rol-6* react with Western blots of *C. elegans* cuticle extracts (Yang and Kramer, unpublished data). Because of the large number of collagen genes in the *C. elegans* genome, an exhaustive study of all *C. elegans* collagens is not feasible. Instead, we are approaching the analysis of cuticle collagen assembly by studying genes that affect organismal morphology and also display unusual genetic properties or interactions. These genetic interactions may reflect physical interactions of the gene products and may therefore be useful in understanding the components and mechanisms that underlie cuticle assembly and the critical factors that determine morphology.

dpy-2 and *dpy-10* are two genes that can alter morphology. Mutations in these genes can cause a variety of phenotypes: dumpy, nonroller (Dpy); nondumpy, left roller (Lrol); and dumpy, left roller (DLrol) (Brenner, 1974; Hosono, 1980; Cox *et al.*, 1985; Kusch and Edgar, 1986). The phenotypes seen in these mutants are similar to the phenotypes exhibited in alleles of *sqt-1*, *dpy-13*, and *rol-6*. Interactions between *dpy-2* and *dpy-10* and other genes important in development have been observed. *sc13* is a recessive left roller allele of the *sqt-1* collagen gene. Both *dpy-10(e128)* and *dpy-2(e489)*, recessive Dpy alleles, provide genetic backgrounds in which *sc13/+* animals are DLrol (Kusch and Edgar, 1986). This effect was termed cryptic dominance because *sc13* appears to act dominantly in the *dpy* background.

Two alleles of each gene, *dpy-2(e8)* and *q292* and *dpy-10(e128)* and *q291*, have been shown to suppress a temperature-sensitive allele of *glp-1* (Maine and Kimble, 1989), a gene that encodes a putative transmembrane receptor in the gonad (Yochem and Greenwald, 1989). The phenotypes displayed in *glp-1* mutants are: defec-

tive embryonic induction of pharyngeal cells, defective embryonic hypodermal development, and defective induction of germline proliferation. *glp-1*; *dpy-2* and *glp-1*; *dpy-10* double mutants are partially rescued for all of the *glp-1* defects at the restrictive temperature. *dpy-10(e128)* and *dpy-2(e8)* also suppress three alleles of *mup-1*, a gene that has a role in muscle development and that may encode an extracellular protein (Bogaert and Goh, 1991).

Analysis of *dpy-2* and *dpy-10* may help us understand the basis for these interactions as well as add to our knowledge of the composition and assembly of the cuticle. In the present investigation we report the molecular identification and characterization of *dpy-2* and *dpy-10*. We present a detailed analysis of the molecular defects that cause the aberrant morphologies seen in mutants of these two genes. From this mutational analysis, it is clear that the activity of *dpy-10* collagen within the cuticle plays a major role in the determination of organismal morphology. In addition, we provide strong evidence that the null phenotype of *dpy-10* is DLrol, with Dpy representing the partial loss-of-function phenotype.

MATERIALS AND METHODS

C. elegans Strains and Genetics

General *C. elegans* maintenance and handling was performed as described (Brenner, 1974). Unless otherwise indicated, all experiments were performed at 20°C. Strains of worms containing the following alleles were used in these studies: *dpy-2(e8)*, *e489*, *sc38*, and *q292*II; *dpy-10(e128)*, *cn64*, *sc48*, *cg36*, *cg37*, *q291*, *q323*, *m457*, and *m481m482*II; *unc-4(e120)*II; *sqt-1(sc13)*, *sc1*, *sc103*, *sc101*, and *sc99*II; and *mnDf39*II, a deficiency maintained as a balanced strain across from the *mnC1* [*dpy-10(e128)* *unc-52(e444)*] chromosome (Sigurdson *et al.*, 1984). Strains containing Tc1 insertions in *dpy-10* (DR1028, DR1048, and JK1012) and the Tc1 revertant (DR1049) were isolated from mutator backgrounds (Riddle, personal communication; Schedl, personal communication). Some strains were obtained from the *Caenorhabditis* Genetics Center, which is supported by the National Institutes of Health Division of Research Resources.

Two alleles of *dpy-10*, *cg36* and *cg37*, were isolated in this study through the following screen. Wild-type N2 males were treated with 50 mM ethylmethanesulfonate (EMS) for 4 h and then mated to *e128e120* hermaphrodites. F1 progeny were screened for Dpy non-Unc animals. Two such animals were isolated during the screening of ~5000 F1 cross progeny. F2 hermaphrodite progeny of these animals were picked to individual plates to isolate alleles *cg36* and *cg37*. These strains were backcrossed to N2 three times before further analysis.

The following alleles of *dpy-2* and *dpy-10* were tested as trans heterozygotes to determine complementation: *q291/q292*, *q291/sc38*, *cn64/sc38*, *cg37/e489*, *cg37/sc38*, *e128/e489*, *e128/sc38*.

Identification, Cloning, and Sequencing of *dpy-2* and *dpy-10*

Molecular techniques, including DNA and RNA analyses, were performed as described (Sambrook *et al.*, 1989). DNAs from cosmids were digested with *Eco*RI, blotted onto Zeta Probe nylon membranes, and hybridized according to manufacturer's instructions (Bio-Rad Laboratories, Richmond, CA). Probes were generated by nick-translating plasmid clones containing the *sqt-1* collagen gene (Kramer *et*

al., 1988) using alpha-³²P dATP. Subclones from cosmid ZK857 were constructed through a "shotgun" approach by digesting cosmid DNA with *Hind*III, ligating into *Hind*III-digested pBS plasmid (Stratagene, LaJolla, CA), and transforming into JM109 *Escherichia coli* cells. Colonies were probed with *sqt-1*-containing clones (Kramer *et al.*, 1988) under low stringency, yielding plasmids pAL1 and pAL5. These plasmid subclones were nick-translated and probed to blots containing *Eco*RI digested genomic DNAs from strains JK1012, DR1048, and DR1049. The modified T7 DNA polymerases Sequenase v1.0 and v2.0 were used for all DNA sequencing with gene-specific oligonucleotide primers according to manufacturer's instructions (United States Biochemical, Cleveland, OH).

cDNA clones were isolated from a lambda phage cDNA library (Barstead and Waterston, 1989) using gene-specific plasmid probes. DNAs from these clones were subcloned into pBS and sequenced. cDNAs covering the entire coding region were isolated. For *dpy-2*, the cDNA clone sequenced covered nucleotides from position 60 to 1750 (polyadenylation site) on Figure 4. For *dpy-10*, the cDNA clone sequenced covered nucleotides from position 110 to 1895 (polyadenylation site) on Figure 5.

Analysis of Mutant Alleles

Polymerase chain reaction (PCR) was performed as described (Innis and Gelfand, 1990). Oligonucleotide primers were constructed that flank the *dpy-2* and *dpy-10* genes and were used for amplification of genomic DNAs isolated from mutant animals. These primers contain built-in restriction sites to facilitate cloning. In some instances, PCR was performed directly on adult mutant worms (Barstead and Waterston, personal communication). PCR fragments were then cloned into pBS and sequenced. To ensure that mutant sequences obtained were not a result of misincorporation by *Taq* DNA polymerase, DNAs from five independently isolated subclones of each mutant were sequenced simultaneously.

Primers used for the amplification of mutant DNAs were constructed as follows. Numbers in parentheses correspond to numbering of nucleotides in Figures 4 and 5. Only primer D10B contained an internal restriction site (*Eco*RI); other primers contained restriction sites that were added to the 5' end to facilitate cloning.

dpy-2: Primer D2A: 5'*Kpn*I-(114)GGATC
GGATACATTTGACAG(-95)3'

Primer D2B: 5' *Pst*I-(1838)TACAGGTCATCTC(1826)3'

dpy-10: Primer D10A: 5'*Bam*HI-(467)CCTA
CCGTAGTACTTCTCTTC(-447)3'

Primer D10B: 5'(1920)GAGAATTTT
GAGAATTCAAA(1901)3'

For the sequence analysis of Tc1 containing *dpy-10* strains (DR1028, DR1048, and JK1012) and the Tc1 revertant strain DR1049, PCR was performed using the following primers (numbering as in Figure 5):

Primer D10C: 5'(580)ATCTACCGGTGTGTAC(596)3'

Primer D10D: 5'(949)TGGTCCTGGAATGCAACATCC(929)3'

PCR products were then sequenced directly using the end-labeled primer:

Primer D10E: 5'(778)GTCGTTTGATCTCTGAA(762)3'

For analysis of relative amounts of Tc1 excision, primers D10C and D10E were used. These primers generate a 198-base pair (bp) fragment when no element is present and a 1910-bp fragment when a Tc1 is present. DNAs prepared from mixed populations of N2 worms were used as PCR templates. The difference in relative amounts of excision was also seen when PCR was performed on individual adult N2 worms.

RESULTS

Isolation and Identification of the *dpy-2* and *dpy-10* Genes

dpy-2 and *dpy-10* have been genetically mapped to the same point in the cluster of LG II, 0.1–0.2 map units to the left of *tra-2* (Hodgkin *et al.*, 1988). Examples of the phenotypes caused by mutations in these genes are shown in Figure 1. Although recombinants between *dpy-2* and *dpy-10* have not been isolated, the genes have been assigned to independent loci based on complementation (Brenner, 1974). The metric in this region of the genome (~1000 kilobase [kb] of DNA per map unit) allowed us to estimate the physical distance between *tra-2*, which has been genetically and physically mapped (Hodgkin and Brenner, 1977; Okkema and Kimble, 1991), and *dpy-2/dpy-10* as ~100–200 kb. We selected a series of overlapping cosmid clones (Coulson *et al.*, 1986, 1988) covering this region of the genome for analysis. Because genetic data suggested *dpy-2* and *dpy-10* might encode collagens, we probed DNAs from these cosmids with a clone containing the *sqt-1* collagen gene (Kramer *et al.*, 1988). Cross-hybridization between most collagen genes occurs under low stringency hybridization conditions because of the shared Gly-X-Y encoding sequences common to all collagens. One cosmid, ZK857, was found to contain two collagen genes by this method. Figure 2 depicts the region of ZK857 that contains the two collagens and some of the clones used for their analysis.

To determine if one of these collagen genes was *dpy-10*, we utilized two *dpy-10* mutant strains isolated from a mutator background, DR1048 (recessive Dpy) and JK1012 (recessive DLRol), and a phenotypic revertant of DR1048, DR1049. Mutator strains cause high frequency germline transposition of *C. elegans* transposable elements (Eide and Anderson, 1985). Southern blots of DNAs from these strains were probed with plasmid clones pAL1 and pAL5 (Figure 3). pAL1 and pAL5 each contain one of the collagen genes from cosmid ZK857. The pAL5 probe revealed a 1.6-kb band shift in strains DR1048 and JK1012, the size expected for insertion of the *C. elegans* transposon Tc1 (Emmons *et al.*, 1983). DNA from revertant strain DR1049 showed the wild-type band size. No alteration of the wild-type band pattern was seen when the probe containing the other collagen gene, pAL1, was hybridized to the same blot. These results suggest that the pAL5 clone contains the *dpy-10* gene and that the mutator-derived *dpy-10* alleles resulted from 1.6-kb insertions into or near the *dpy-10* gene. DNA from DR1049 shows the wild-type hybridization pattern, indicating that it was generated by the excision of the 1.6-kb insertion in strain DR1048. Sequencing of restriction fragments confirmed that the 1.6-kb insertions present in strains DR1048 and JK1012 are Tc1 elements in the coding region of *dpy-10* (see below). Sequence analysis of the other collagen gene from *dpy-*

Figure 1. Nomarski micrographs of wild-type and mutant adult animals. (A) N2, wild-type; (B) *dpy-10(e128)*, Dpy; (C) *dpy-10(cn64)*, DLRol; (D) *dpy-10(e128) sqt-1(sc101)*, pinhead Dpy. Notice the narrowing of the head in the pinhead Dpy relative to the Dpy and DLRol animals.

2 mutants demonstrated that this neighboring collagen gene is *dpy-2* (see below).

Structure of the *dpy-2* and *dpy-10* Genes and Their Products

The wild-type sequences of *dpy-2* and *dpy-10* are presented in Figures 4 and 5. The structures of the genes are compared schematically in Figure 6. cDNA clones of each gene were isolated from a lambda phage library (Barstead and Waterston, 1989) to confirm exon sequences and to determine intron-exon structures and polyadenylation sites. The regions covered by the cDNAs are indicated in the legends of Figures 4 and 5. Primer extension using end-labeled probes was performed to determine transcription start sites, indicated as position +1. Direct RNA sequencing revealed that *dpy-2* transcripts are trans-spliced with the SL1 splice leader (Krause and Hirsh, 1987); *dpy-10* is not trans-spliced. Seventy-five percent of *C. elegans* introns are between 45 and 59 bases long (Emmons, 1988). The

first intron of *dpy-2* (180 nt) and the first intron of *dpy-10* (342 nt) represent relatively long *C. elegans* introns. The position of intron 2 is conserved between *dpy-2* and *dpy-10*.

The number and spacing of cysteine residues is of interest because disulfide bonding is important in cuticle

Figure 2. Restriction map of a portion of cosmid ZK857 containing two collagen genes. Boxed above the restriction map are the collagen genes, determined to be *dpy-10* and *dpy-2* (see Figure 3 and text). Below the restriction map are some of the subclones used in this study. H, HindIII; V, EcoRV; S, Sal I; E, EcoRI; B, BamHI; X, Xho I.

Figure 3. Analysis of genomic DNAs from strains containing mutator-induced *dpy-10* alleles. DNAs were digested with *Eco*RI, electrophoresed, and blotted. Lanes 1 and 5, N2 (wild-type); lanes 2 and 6, JK1012; lanes 3 and 7, DR1048; lanes 4 and 8, DR1049. Lanes 1–4 were probed with pAL5; lanes 5–8 were probed with pAL1.

assembly and integrity (Cox *et al.*, 1981). There are three conserved sets of cysteines in cuticle collagens. On the amino-terminal side of the Gly-X-Y repeats there are three cysteines, within the first Gly-X-Y interruption there are either two or three cysteines, and on the carboxyl-terminal side of the Gly-X-Y repeats there are two cysteines. The positions of these cysteines in *dpy-2* and *dpy-10* are identical but different from all other sequenced collagens. The carboxyl non-Gly-X-Y portions of *dpy-2* and *dpy-10* are considerably longer, 60 and 51 amino acids, respectively, than any of the other 15 sequenced cuticle collagens, which have carboxyl non-Gly-X-Y tails of <25 amino acids. These two collagens have 40.8% overall identity at the amino acid sequence level (Figure 7). Amino acid conservation is higher around the three sets of conserved cysteines. Eight out of nine amino acids are identical in the regions surrounding the first two sets of conserved cysteines, and seven out of nine amino acids are identical surrounding the last set of conserved cysteines. The first Gly-X-Y interruption of *dpy-2* occurs on the carboxyl side of the fourth Gly-X-Y triplet and consists of a Ala-Leu-Pro sequence. This interruption is unusual because all other sequenced cuticle collagens have 10 or 11 Gly-X-Y repeats before the first interruption.

Genetics of *dpy-2* and *dpy-10*

To determine if there are interactions between *dpy-2* and *dpy-10*, we tested trans heterozygotes [*dpy-2*(*x*) + / + *dpy-10*(*y*)] for complementation. With all alleles tested (see MATERIALS AND METHODS), animals were

wild-type, indicating full complementation between mutant alleles of these two genes.

We then tested heteroallelic combinations of *dpy-10* alleles to determine possible relationships between phenotype and level of wild-type *dpy-10* function (Table 1). Several generalizations can be drawn from this analysis. *Dpy* alleles in trans to other *Dpy* alleles always yield *Dpy* animals. *Dpy* alleles in trans to *DLRol* alleles can yield both *Dpys* and *DLRols*. Often, an intermediate phenotype is observed in which animals lift their heads in an upward twisting motion, stop, and then back up. We have termed this phenotype “lifter” (Lif). Lif animals sometimes show a related behavior that consists of twisting the head downward (the beginning of a normal roll), causing the head to be bent back. Usually the animals will then stop, back up, and continue forward normally. We call this behavior “head catching” (Hdc). Animals displaying the Lif and Hdc traits often appear to be on the verge of rolling but rarely do. *DLRol* alleles in trans to other *DLRol* alleles invariably give *DLRols*. *cn64* (Hosono, 1980) is a dominant left roller allele that displays the left roller phenotype only when in trans to the wild-type allele or the *m481m482* allele. *m481m482*, the allele contained in revertant strain DR1049, fails to fully complement other *dpy-10* mutant alleles despite its wild-type phenotype. Phenotypes are also given for each allele in trans to a deficiency of the entire *dpy-2/dpy-10* region. *Dpy* alleles tend to acquire more roller (Lif or Hdc) character in trans to a deficiency; *DLRol* alleles remain unchanged. The *cg36* allele behaves like a deficiency in trans to all other alleles tested, indicating that it may be a null allele.

The observation that *dpy-10*(*e128*) *sqt-1*(*sc13*)/*dpy-10*(*e128*)⁺ animals are *DLRol* (Kusch and Edgar, 1986) prompted us to ask how the *e128* background affected other *sqt-1* alleles (Table 2). In the *e128* background, all non-null *sqt-1* heterozygous animals display the *DLRol* phenotype similar to the more severe *dpy-10* alleles. *sqt-1* alleles tested include *sc13*, a recessive left roller; *sc101*, a recessive long; and *sc99*, an allele with no phenotype that fails to complement other *sqt-1* alleles. Surprisingly, even the dominant right roller allele *sc1* displayed the *DLRol* phenotype as a heterozygote in the *e128* background. Thus, the *dpy-10* mutant phenotype is displayed preferentially over that of the *sqt-1* mutant. A null allele, *sqt-1*(*sc103*), however, displayed the *Dpy* phenotype as a heterozygote in the *e128* background, indicating that it is the presence of the mutant *sqt-1* collagen, not merely the absence of the normal *sqt-1* collagen, that is involved in this interaction. Homozygous double mutant animals displayed an unusual phenotype we term “pinhead.” The majority of the body of these animals is *Dpy*; however, the tip of the head through the region of the pharynx is much thinner, approximating wild-type width (see Figure 1D). The pinhead body shape is easily distinguished from that of a normal *Dpy* animal.

```

-56   aaa cag aat ctt ctt cca taa ctt ctt att ttc ctt tct ttt ctg ttt aat ttc act+1 ctt ttt tta taa gct ttc att tta gca gta aca ttt gaa
41   aat taa acc tca aat atg aaa ctt tca gcg ATC AAA TCG CAA ACG AGT GGG TAC GTG GAC CCA CAA ACA GAG GAA TAT GAT GAG gtg agc atg att
1   Met Lys Ser Gln Thr Ser Gly Tyr Val Asp Pro Gln Thr Glu Glu Tyr Asp Glu
137  aat tat taa aat ttt cag att ata act gtg tag aac aat ctt ggt gaa ttt tgt agc tct tcg caa tgt aga att ttt aat ttc tga aaa tat cat
233  ctt aca cat ttt cca aaa agc att tga aga ttt cat att tta tac ttc aaa aat aca ata ttt tat ttt cag TCT CGT GAA AAG AGT TAT CAA GTT
19   Ser Arg Glu Lys Ser Tyr Gln Val
329  CTG ATT GCC GTT TCC GGA GTT CTT TCG ATC ACA TCT CTT CTT ATT CTT GTT TTA TTC GTC CCT TCA ATG TAC AAT TAT GTC GAT AAC ATT TCA CGG
27   Leu Ile Ala Val Ser Gly Val Leu Ser Ile Thr Ser Leu Leu Ile Leu Val Leu Phe Val Pro Ser Met Tyr Asn Tyr Val Asp Asn Ile Ser Arg
425  TTC TCA AGG CGA GAT TTC GAA TTT TGT CAG gta aat ttt ctt ata cat ata agt tta att tcc gaa ctt tct gat ttt ata g GCA TCC GCA AAT
59   Phe Ser Arg Arg Asp Phe Glu Phe Cys Gln Ala Ser Ala Asn
519  GAT TTA GAA ACT GAA ATG ATG TCT GTT AGA GAA GGG CTT TTG AGA GGA AGA AAT GTA ACG AAA AGG GCT GCT GGA TAT GGA CAT TAT AAT CCA TCA
73   Asp Leu Glu Thr Glu Met Met Ser Val Arg Glu Gly Leu Leu Arg Gly Arg Asn Val Thr Lys Arg Ala Ala Gly Tyr Gly His Tyr Asn Pro Ser
615  ATG CTG GCT GCC GAT TCA CCA CAA TTT CAA GAG TGT CCA Ggt aag att tat aaa caa ttt tga aaa tct tag atg tta aat aat tta ata tttacag
105  Met Leu Ala Ala Asp Ser Pro Gln Phe Gln Glu Cys Pro
712  CC TCG TGT ATT CCG GGA GAA CGT GGA CCA TCA GGA GAC TCT GGA CTT CCA GCT CTT CCA GGA GCA CCT GGA CCC GAT GGA GCA CCA GGA CGA CCA
118  Ala Cys Cys Ile Pro Gly Glu Arg Gly Pro Ser Gly Asp Ser Gly Leu Pro Ala Leu Pro Gly Ala Pro Gly Pro Asp Gly Ala Pro Gly Arg Pro
807  GGA ACC ACT CCA AAT GCC TCA TGT ATT CCA GAA AGA GTT TTC GAG CCA CCA CCA TGT CTT CCA TGT CCA CAA GGA CCA GCT GGA GTT CCA GGA CAT
150  Gly Thr Thr Pro Asn Ala Ser Cys Ile Pro Glu Arg Val Phe Glu Pro Pro Cys Leu Pro Cys Pro Gln Gly Val Val Pro Gly His
903  CCA GGA TTC CCA GGT GAT CCA GGA GAG TAT GGA ATT GGA GGA AGA CCA GGA TCC GAT GGA ATG CCA GGG AAA CCA GGA GAT CCA GGA CTC GCA GGA
182  Pro Gly Phe Pro Gly Asp Pro Gly Glu Tyr Gly Ile Gly Gly Arg Pro Gly Ser Asp Gly Net Pro Gly Lys Pro Gly Asp Pro Gly lew Ala Gly
999  CCA ATT GGA CCA CCA GGA GAA TCT GGA CCA ATT GGA GAT AAG GGA AGA ACA CCA GAA GCA CAT GTT ATT CCA GGA CCA CCT GGA GAT TCT GGA TTA
214  Pro Ile Gly Pro Pro Gly Glu Ser Gly Pro Ile Gly Asp Lys Gly Arg Thr Pro Glu Ala His Val Ile Pro Gly Pro Pro Gly Asp Ser Gly Leu
1095 CCA GGA CCA TGG GGA CCA CCA GGA TCA GCT GGT ATG CCT GGA GAA GAT GGA TAC GCA GGA ACA CCT GGA GAG AAG GGA TGG CCA GGA CCG CCA GGA
246  Pro Gly Pro Trp Gly Pro Gly Ser Ala Gly Met Pro Gly Glu Asp Gly Tyr Ala Gly Thr Pro Gly Glu Lys Gly Trp Pro Gly Pro Pro Gly
1191 GCA CCA GGA CCA GGT GGA ATG CCT GGA CCA AAT GGT CCA ACT GGA GAA CAA GGA CCT TCT GGA ACT CCT GGT ACT TGT GTT TGC CAA GAT ACT GAA
278  Ala Pro Gly Pro Gly Gly Met Pro Gly Pro Asn Gly Pro Thr Gly Glu Gln Gly Pro Ser Gly Thr Pro Gly Thr Cys Val Cys Gln Asp Thr Glu
1287 GTA GTT ATG AAT GAT GAA AAG GGA CGT GTT CCA GCT CCA CGT GAT AAT gta agt taa tgt ttg aac tgt att ttt aaa ctt aat cag aaa aac caa
310  Val Val Met Asn Asp Glu Lys Gly Arg Val Pro Ala Pro Arg Asp Asn
1383 act ata cta aaa att tca g GTT GCT CCG GGA GCC ACG GGT GGA TCA TAT GAA CCA CAA GGA GGA AAT GAT GCA GCT GCT CCA TCA GCA ATT TCT
326  Val Ala Pro Gly Ala Thr Gly Gly Ser Tyr Glu Pro Gln Gly Gly Asn Asp Ala Ala Ala Pro Ser Ala Ile Ser
1477 CAG AGT CAA AAC CGA CGG ATA CGG AAA TGG TGA atc agt cag taa tca taa gaa taa tgg agg ata tta tca ctt gcg taa att cac aca ata gtt
351  Gln Ser Gln Asn Arg Arg Ile Arg Lys Trp *
1573 gat tct ctc tgc atc ctt ttt tga cgg tac ctc att atg atc tag ccc gat ttc ggc agt tgc taa act aag ggt ttt ctc ttt aat ttt ctt cgt
1669 tgt ttc ttt cca aaa att caa aca ttc cac agt tgt tta tga cca tct aat aag tgt aat tat taa taa atc tta cat gat aaa agt gca tac aaa
1765 tta aat aaa

```

Figure 4. Nucleotide and deduced amino acid sequences of *dpy-2*. Numbering of nucleotides uses start of transcription (underlined) as +1. Gly-X-Y repeats are underlined; 5' and 3' noncoding regions and introns are in lowercase letters. Intron/exon structure was determined by sequencing a cDNA clone that covered nucleotides 60–1750 (polyadenylation site). Poly A⁺ addition signal and site are underlined. SL1 is trans-spliced to the *dpy-2* transcript at nucleotide position +69.

Analysis of Point Mutations

To determine the molecular nature of the *dpy-2* and *dpy-10* mutant alleles, PCR generated fragments derived from mutant DNAs and containing either the entire *dpy-2* or entire *dpy-10* gene were sequenced (Table 3). All of the nucleotide substitutions are GC-to-AT transitions; the most common mutation induced by the chemical mutagen EMS. Figure 8 summarizes the location of each point mutation sequenced. Substitutions for glycines within the Gly-X-Y regions were found in 7 out of 10 point mutations. Three of these seven are temperature-sensitive, displaying the Dpy phenotype at 15°C and the DLrol phenotype at 25°C. All three temperature-sensitive alleles are *dpy-2* mutations. The phenotypes generated by the glycine substitution mutations are position dependent. For example, a Gly-to-Arg substitution at amino acid position 247 in *dpy-2* confers the tem-

perature-sensitive DLrol phenotype whereas the equivalent substitution six amino acids away at position 253 confers a strict Dpy phenotype.

One nonsense mutation was identified, *dpy-10(cg36)*, that creates an ochre codon in the 48th (out of 50) Gly-X-Y triplet. We believe this allele to be null for several reasons. The mutant collagen would lack the conserved carboxyl-terminal domain that has been shown to be required for assembly of vertebrate fibrillar collagens (Rosenbloom *et al.*, 1976; Gelman *et al.*, 1979). Also, *cg36* behaves like a deficiency in trans to other *dpy-10* alleles (see Table 1). The phenotype of *cg36* animals is DLrol. Interestingly, the severity of the roller and dumpy phenotypes in this mutant are not as severe as that seen in some other mutants. For example, the mis-sense mutant *cn64* is a stronger LRol and a more severe Dpy than *cg36*.

```

-100  gga gta aaa cta aaa agg acg aaa acc aaa agt aag ttg cag ttt taa att gtt ttt ttc acc ata tca aat tat aat att cca gcg aca tac ctg
      +1
-4    aca gtc gac caa agt tct tgt gcg ttc ctt tgg gcc ggc caa gct tat tga gaa aag gcc gaa agg ctt ctt aca cat ttt caa gtg aag act tga
93    gaa aag aaa ctc aag aaa ttg aag tta aga aga cca ata tct tta cgt ttt tct tcg ata caa ata caa tca gca aga ATG AAA AAC AAC GCC AAA
1     Met Lys Asn Asn Ala Lys
189   GAA GAG tga gtt gaa ggg gct ttt aaa gaa caa cta att tta ata tta ttc cca gta tgt aga cag att taa atg ttc aaa cca gtt tta ctc taa
7     Glu Asp
286   aaa ttt gtt taa aaa aac gga agc aga caa gca gac aat cag aca acc aat ttt ttt ggt cta aaa att ttt gtc tca aat tat tgt tta ata tta
382   ccc tat ttt ctc tat gta tgg cat tca agt ata ttt tcc gat caa tca att tga gtg caa ctc caa tag ggg cgc aat act aat aga gaa act atg
478   ttt act aat tca gag tca tca tct cgc cgc att cat tga ata aat taa tta att tt   tc agC TAT CGA ACG TTC TCG CTG ACA ACG AAC TAT TCG
9     Tyr Arg Thr Phe Ser Leu Thr Thr Asn Tyr Ser
572   CGT CAG ATG ATC TAC CGG TGT GTC ACC GGT CTT CAA ATC GGA TTT AGC TTA TTT TCA TTC ATC ATT GTC TGT GTT GCT CTC CCA ATT ATG TAC AAT
20    Arg Gln Met Ile Tyr Arg Cys Val Thr Gly Leu Gln Ile Gly Phe Ser Leu Phe Ser Phe Ile Val Cys Val Ala Leu Pro Ile Met Tyr Asn
668   CAT GTC CAG AAT ACT ATT ACC TAT GTC GAC AGA GAA ATG GCT TAT TGC GAA gta aga ttt tta tat ttc tat tgt ata aaa ttt ttt tca act ctt
52    His Val Gln Asn Thr Ile Thr Tyr Val Asp Arg Glu Met Ala Tyr Cys Glu
764   cag AGA TCA AAC GAC GAG GCT GCA CTT GAA CTT CAA TAC GGC AAG ATG AGA ATG ACT GGA AAC CGT ACC GCT CGT GGT GCC TAT GGT AGC GGA GCT
69    Arg Ser Asn Asp Glu Ala Ala Leu Glu Leu Gln Tyr Gly Lus Met Arg Met Thr Gly Asn Arg Thr Ala Arg Gly Ala Tyr Gly Ser Gly Ala
860   TCA CAT GGC TTC AGA CCA ACA GCC TAT GGA GAT GAA ATC ACT GGA GCT CCA CTT GAA ACT GAA TGT CCA GGA TGT TGC ATT CCA GGA CCA CCA GGT
100   Ser His Gly Phe Arg Pro Thr Ala Tyr Gly Asp Glu Ile Thr Gly Ala Pro Leu Glu Thr Glu Cys Pro Gly Cys Cys Ile Pro Gly Pro Pro Gly
956   CCA CGT GGA TCA TCT GGA ACA CCA GGA AAA CCT GGA CTT CCA GGA AAC GCT GGA AAG CCA GGA ATG CCA GGA ACT ACT CCA AAT CAA ACA TGC CCA
132   Pro Arg Gly Ser Ser Gly Thr Pro Gly Lys Pro Gly Leu Pro Gly Asn Ala Gly Lys Pro Gly Met Pro Gly Thr Thr Pro Asn Gln Thr Cys Pro
1052  CTC AAC CAA GTT CGT GAG CCA CCA CCG TGC CGT CCA TGT CCA AAG GGA CCA CCA GGA ATC AAG GGA TGG CCA GGA TTC CCA GGA GAT GTT GGG CCA
164   Leu Asn Gln Val Arg Glu Pro Pro Pro Cys Arg Pro Cys Pro Lys Gly Pro Pro Gly Ile Lys Gly Trp Pro Gly Phe Pro Gly Asp Val Gly Pro
1148  CCA GGA CCA CCA GGA CTT AAG GGA ATT GAT GGT GAA GAC GGA GCA CCA GGA GAG ACT GGA CCA ACT GGA CCA CCA GGA TAC AGA GGA GGA CCA GGA
196   Pro Gly Pro Pro Gly Leu Lys Gly Ile Asp Gly Glu Asp Gly Ala Pro Gly Gly Thr Gly Pro Thr Gly Pro Pro Gly Tyr Arg Gly Gly Pro Gly
1244  GCT CCT GGA GAT AAG GGA CCA ACT CCA GAA GGT GAC TTG AAG GAG GGA CCA CCA GGA GAC GAA GGA CCA CCA GGA CCA ATT GGA GCT CCA GGA ATG
228   Ala Pro Gly Asp Lys Gly Pro Thr Pro Glu Gly Asp Leu Lys Glu Gly Pro Pro Gly Asp Glu Gly Pro Pro Gly Pro Ile Gly Ala Pro Gly Met
1340  CCT GGA TTA CCA GGA AGA AAC GGA TTG ACC GGA GGA CAA GGA GAA CGA GGA TGG CCG GGA GTT TCT GGA GAA TCT GGT GAG CCA GGA TAT CCA GGA
260   Pro Gly Leu Pro Gly Arg Asn Gly Leu Thr Gly Gly Gln Gly Glu Arg Gly Trp Pro Gly Val Ser Gly Gly Ser Gly Glu Pro Gly Tyr Pro Gly
1436  CCA GAA GGA CCA ATG GGA GGA CAA GGA CCA CCA Ggt gag tct tga aca act att ttt att cca aaa ctt att gaa att tcc ag   GA GAG CCA GGA
292   Pro Glu Gly Pro Met Gly Gly Gln Gly Pro Pro Gly Glu Pro Gly
1530  GTC TGT GTC TGT CAA AAC GTT GAT AGT ATT CTT TTG ATC AAT CCA GGA CCA CAG CCA CGC ATT CGT GCC GAT GAC TAC AAC AGT GAT GAC GGA TAT
307   Val Cys Val Cys Gln Asn Val Asp Ser Ile Leu Leu Ile Asn Pro Gly Pro Gln Pro Arg Ile Arg Ala Asp Asp Tyr Asn Ser Asp Asp Gly Tyr
1626  GGT GGA AGT CGT GGA GGA GGA GAT CGT GCC GGA TAC CAA GGT TAC GGT AGA AAG TAG tca ttc gac ttt cag ata ttc ttc tat caa tta aca tta
339   Gly Gly Ser Arg Gly Gly Gly Asp Arg Ala Gly Tyr Gln Gly Tyr Gly Arg Lys *
1722  aaa aga cct cga tga ctt ctc att caa aat gcg tgt aca ttg ttc att tgt ttg ttt ccc aaa tac att ccc act tgg ctt tta gta atg ttt ttg
1818  acc tgt ttg att ctc gta gtt att ctc atg atg aca tgg ttc ttt aat aat aat aat aat aat aaa gat aca tta ata att ttt gaa ttc tca aaa
1914  att ctc aaa g

```

Figure 5. Nucleotide and deduced amino acid sequences of *dpy-10*. Numbering of nucleotides uses the start of transcription (underlined) as +1. Gly-X-Y repeats are underlined; 5' and 3' noncoding regions and introns are in lowercase letters. Intron/exon structure was determined by sequencing a cDNA clone that covered nucleotides 110–1895 (polyadenylation site). Poly A+ addition signal and site are underlined. The *dpy-10* transcript is not trans-spliced.

The *dpy-10(e128)* allele creates an alteration in the splice acceptor of intron 2. The consensus splice acceptor dinucleotide AG is changed to AA. If unspliced, the 16 amino acids encoded by the intron would presumably be present in the mature *dpy-10* collagen in *e128* animals because the intron contains no stop codons and retains the correct reading frame. Because the phenotype of this mutant is non-null (it is a strict Dpy), some level of normal *dpy-10* function must be present in these animals. We do not know if this is due to the presence of the insertion-containing collagen, the presence of low levels of normally spliced product, or the use of a cryptic splice site that allows the production of a partially functional collagen. Recent observations (Aroian *et al.*, 1993) indicate that several different *dpy-10* transcripts are

present in *e128* animals. Some transcripts are normally spliced despite the mutation at the splice acceptor, others are spliced to acceptors near the correct site, and some transcripts are unspliced for intron 2. The level of *dpy-10* transcripts that are spliced to the correct site is ~10% that of wild-type.

The *cn64* allele results in an arginine to cysteine substitution in the amino half of the *dpy-10* collagen. This mutant displays the left roller phenotype in the heterozygous state. As a homozygote, it is temperature-sensitive, displaying the Dpy phenotype at 15°C and the DLRol phenotype at 25°C. The arginine altered in this mutant is the second of two highly conserved arginines found in this region of most cuticle collagens. The consensus sequence in this region is

Figure 6. Structural comparison of the *dpy-2* and *dpy-10* collagen genes. Thin lines indicate noncoding regions; thick lines indicate coding regions. ▨ represent Gly-X-Y encoding regions, ■ represent interruptions of Gly-X-Y repeats. Introns are indicated by ▽ with size given in nucleotides (nt). Numbers given in parentheses represent nucleotide position with numbering as in Figures 4 and 5.

Arg-X-X-Arg-Gln-Ala-Tyr, where the second X is often an arginine as well.

Mutator Strain-induced Alleles of *dpy-10*

Analysis of mutator strain-induced alleles of *dpy-10* indicated the presence of single Tc1 elements in strains JK1012, DR1028, and DR1048. To determine the location of Tc1 insertions in these strains, the elements were mapped relative to restriction sites on genomic DNA blots. All three elements appeared to be inserted at or very near the same site. Oligonucleotide primers that flank this region were used to amplify genomic DNA from each of these three strains as well as the DR1049 revertant strain. The nucleotide positions of the Tc1 insertions were then determined by direct sequencing of the PCR-amplified fragments (Table 4). The Tc1 elements in JK1012 and DR1028 are inserted between the same two nucleotides and are in the same orientation (forward) with respect to the Tc1 open reading frames (Rosenzweig *et al.*, 1983) (Figure 9). The Tc1 element in strain DR1048 is 33 nucleotides upstream of the ones in JK1012 and DR1028 and is in the opposite orientation. The excision event that generated strain DR1049 was imprecise; it removed three nucleotides on the 5' side of the insertion, deleting an isoleucine codon. Therefore, allele *m481m482* represents a deletion mutant despite its wild-type phenotype. This may explain why it does not fully complement other *dpy-10* alleles.

Strains DR1028 and JK1012 have Tc1 elements inserted into *dpy-10* in the same place and in the same orientation yet they have different phenotypes: DR1028 is Dpy, whereas JK1012 is DLRol. One possible explanation for the different phenotypes of these strains is different amounts of somatic excision. Somatic excision of Tc1 is known to occur at a frequency of 1–5% (Eide and Anderson, 1985). Animals that contain more cells in which the Tc1 element has excised may produce more

functional *dpy-10* collagen. If this were the case, we would expect the strain with the nonnull phenotype (Dpy, strain DR1028) to undergo more excision. To test this hypothesis, PCR was performed on DNAs obtained from adult DR1028, DR1048, JK1012, and DR1049 animals using primers that flank the insertion site. The relative amounts of excision that occurred in each strain were then compared (Figure 10). There was significantly less excision in strain JK1012 relative to the other two Tc1 alleles. We estimate the difference to be between 5- and 20-fold. Strains DR1028 and DR1048, both Dpys, appear to have approximately equal amounts of Tc1 excision. These observations are consistent with the notion that the different phenotypes exhibited by these strains are caused by different amounts of excision.

DISCUSSION

To understand the assembly of the collagens that comprise the *C. elegans* cuticle, we have undertaken studies of loci that affect the morphology of this nematode. We have shown here that the *dpy-2* and *dpy-10* genes encode collagens that have major influence on the shape of the organism. The overall structure of the genes is similar to the other previously sequenced presumptive cuticle collagens (Kramer *et al.*, 1982, 1988, 1990; von Mende *et al.*, 1988; Cox *et al.*, 1989; Johnstone *et al.*, 1993). Each cuticle collagen gene encodes a 30–35 kDa protein that contains an amino non-Gly-X-Y region, a Gly-X-Y repeat region that has several short interruptions, and a short carboxyl non-Gly-X-Y region. The domain organization of these collagens is similar to the vertebrate types IX and XII collagens (reviewed in Burgeson, 1988).

The *dpy-2* and *dpy-10* collagens share 40.8% amino acid identity and have identical patterns of conserved

<i>dpy-10</i>	- MK--NNAKDYRTFWLTNYSROMIYRCVLTGIGQIFSLFSFIIVCVLPI -48
<i>dpy-2</i>	- MKSQTSGYVDPQT--EEYDESREKSYQVLIAVSGVLSITSLILVLFVPS -48
<i>dpy-10</i>	- MYNHVQNTITYVDREMAICERSNDEAALELQYGGKQMTGNRTARGAYGSG -98
<i>dpy-2</i>	- MYNYVDNISRFSSRDFFPCQASANDLETSMVSVREGLLRGNVT-KRAAG -97
<i>dpy-10</i>	- ASHGFRPTAYGDEITGAPLETECPGCCIPGPPGPRGSGTPGKPLPGNA -148
<i>dpy-2</i>	- YGH-YNPSMLAAD---SPQFQECFACCTPGERGSGDSGLPALPGARPD -143
<i>dpy-10</i>	- GKPGMPGTTTNPQTCPLNQVREPPPCRCPCPKGPGIKGMPGPGDVGPPGP -198
<i>dpy-2</i>	- GAPGRPGTTPNASCIPERVFEPPCLPCQPCPGPRGVPGHPGPGDPEGYI -193
<i>dpy-10</i>	- PGLKIGDGEDGAPGETGPTGPPGYRGPGAPGDKGPTPEGDLKEGPPGDE -248
<i>dpy-2</i>	- GGRPGSDGMPKPGDPLAGPIGPPGESGPIGDKGRTPEAHVIPGPPGDS -243
<i>dpy-10</i>	- GPPGPIGAPGMPGLPGRNLTGGQGERGWPGVSGESGEPGYPGPGPMGG -298
<i>dpy-2</i>	- GLPGFMGPPGSGAGMGEDGYAGTPGEKQWPPGAPGPGMGPPNGPTGE -293
<i>dpy-10</i>	- QGPPGEPGVEVCQNVDSILLINPG--PPRIRADDYNSDDGY---GGSRG -343
<i>dpy-2</i>	- QGPGSTPGTGVQCQDTEVMNDEKGRVAPRPNVAPGATGSGSYEPGGGND -343
<i>dpy-10</i>	- GGDRAQYQGYGRK -356
<i>dpy-2</i>	- AAPSAISQSNRRIRKW -360

Figure 7. Amino acid sequence alignment of *dpy-2* and *dpy-10*. Colons indicate identical amino acids. Conserved cysteines are indicated in bold and Gly-X-Y repeats are underlined.

Table 1. *dpy-10* heteroallelic phenotypes^a

	<i>e128</i>	<i>q291</i>	<i>m457</i>	<i>m481</i>	<i>cn64</i>	<i>m481</i> <i>m482</i>	<i>cg37</i>	<i>sc48</i>	<i>q323</i>	<i>cg36</i>	<i>Df</i> ^b
+	WT	WT	WT	WT	LRol	WT	WT	WT	WT	WT	WT
<i>e128</i>	Dpy	Dpy	Dpy	Dpy	Dpy	wDpy Lif	Dpy	Dpy LRol	Dpy Lif	Dpy Lif	Dpy Lif
<i>q291</i>		Dpy	—	Dpy	Dpy Lif	wDpy LRol	Dpy	Dpy wLRol	Dpy Lif	Dpy wLRol	Dpy wLRol
<i>m457</i>			Dpy	Dpy	Dpy	wDpy	Dpy Lif	—	Dpy Lif	Dpy Lif	Dpy Lif
<i>m481</i>				Dpy	Dpy	—	Dpy Lif	Dpy LRol	—	Dpy Lif	Dpy Lif
<i>cn64</i>					Dpy LRol	LRol	Dpy	—	Dpy LRol	Dpy Lif	—
<i>m481m482</i>						WT	wDpy wLRol	Lif	—	Lif	—
<i>cg37</i>							Dpy wLRol	Dpy wLRol	Dpy wLRol	Dpy wLRol	Dpy wLRol
<i>sc48</i>							Dpy LRol	Dpy LRol	Dpy LRol	Dpy LRol	Dpy LRol
<i>q323</i>									Dpy LRol	—	Dpy LRol
<i>cg36</i>										Dpy wLRol	Dpy wLRol

^a Trans-heterozygous animals were examined as adults at 20°C. Dpy, dumpy; LRol, left roller; Lif, lifter and/or head catcher (see text); w, weak; WT, wild-type; —, not determined

^b *mnDf39* was used to construct *dpy-10/Df* animals.

cysteines. No other sequenced cuticle collagen shares the *dpy-2/dpy-10* cysteine pattern, indicating that they represent a novel collagen subfamily. The proximity of the genes (3.5 kb) combined with their structural similarities provides evidence that they are likely the result of a gene duplication event.

Given this, one might ask if the functions of these collagens overlap in any way. If this were the case, one prediction that can be tested genetically is that double heterozygotes of *dpy-2* and *dpy-10* would display some

Table 2. *dpy-10(e128)* interactions with *sqt-1*^a

<i>sqt-1</i>	<i>dpy-10</i> genotype		
	+/+	<i>e128</i> /+	<i>e128/e128</i>
+/+	WT	WT	Dpy
<i>sc1</i> /+	RRol	RRol	DLRol
<i>sc1/sc1</i>	Dpy	Dpy	phDpy ^b
<i>sc99</i> /+	WT	WT	DLRol
<i>sc99/sc99</i>	WT	WT	phDpy
<i>sc101</i> /+	WT	WT	DLRol
<i>sc101/sc101</i>	Lon	Lon	phDpy
<i>sc13</i> /+	WT	WT	DLRol
<i>sc13/sc13</i>	LRol	LRol	phDLRol
<i>sc103</i> /+	WT	WT	Dpy
<i>sc103/sc103</i>	WT	WT	Dpy

^a WT, wild-type; Dpy, dumpy; RRol, right roller; LRol, left roller; Lon, long; DLRol, dumpy, left roller; ph, pin-head.

^b *e128 sqt-1* double mutant animals display a distinct Dpy phenotype in which the head is noticeably thinner than the rest of the animal. These phDpys are easily distinguishable from the *e128* Dpy and from the *sc1* Dpy (see Figure 1D).

Table 3. *dpy-2* and *dpy-10* point mutations

Allele	Nucleotide ^a	Amino acid ^b	Phenotype ^c	
			15°C	25°C
<i>dpy-2(sc38)</i>	G ₇₄₅ → A	Gly ₁₂₉ → Glu	Dpy	DLRol
<i>dpy-2(e8)</i>	G ₉₀₆ → A	Gly ₁₈₃ → Arg	Dpy	DLRol
<i>dpy-2(e489)</i>	G ₁₀₉₈ → A	Gly ₂₄₇ → Arg	Dpy	DLRol
<i>dpy-2(q292)</i>	G ₁₁₁₆ → A	Gly ₂₅₃ → Arg	Dpy	Dpy
<i>dpy-10(e128)</i>	G ₇₆₆ → A	Splice ^d	Dpy	Dpy
<i>dpy-10(cn64)</i>	C ₈₃₆ → T	Arg ₉₂ → Cys	Dpy	DLRol
<i>dpy-10(cn64)/+</i>			LRol	LRol
<i>dpy-10(sc48)</i>	G ₉₇₂ → A	Gly ₁₃₇ → Glu	DLRol	DLRol
<i>dpy-10(cg37)</i>	G ₁₀₁₇ → A	Gly ₁₅₂ → Glu	DLRol	DLRol
<i>dpy-10(q291)</i>	C ₁₂₂₃ → A	Gly ₂₂₁ → Arg	Dpy	Dpy
<i>dpy-10(cg36)</i>	C ₁₄₅₇ → T	Gln ₂₉₉ → ochre	DLRol	DLRol

^a Nucleotide position of mutation. Numbers correspond to the numbering in Figures 4 and 5.

^b Amino acid substitution caused by mutation. Numbers correspond to numbering in Figures 4 and 5.

^c Dpy, dumpy; DLRol, dumpy, left roller; LRol, left roller.

^d The mutation in *e128* is in the splice acceptor dinucleotide AG of intron 2.

Figure 8. Schematic representation of *dpy-2* and *dpy-10* point mutations. A composite *dpy-2/dpy-10* collagen gene is depicted. Thick lines represents coding regions, $\square$, Gly-X-Y repeats, $\blacksquare$, Gly-X-Y interruptions, and vertical lines represent cysteine residues. *dpy-2* mutations are depicted above the collagen drawing; *dpy-10* mutations are below. The allele name is given in the upper box, the amino acid substitution is in the middle box, and the phenotype is indicated in the lower box. *e128* is a g-to-a nucleotide substitution in intron 2 and *cg36* creates an ochre codon. Dpy, dumpy; DLRol, dumpy, left roller; LRol, left roller; ts, temperature-sensitive; dom, dominant.

of the same morphologic defects seen in animals containing two mutant copies of either *dpy-2* or *dpy-10*. In other words, if the two collagens functioned together structurally, disruptions in any two of the four alleles present might result in altered organismal morphology. However, double heterozygotes for all recessive alleles tested did not display any morphologic alterations. This result suggests that despite their structural similarity, these two collagen genes and/or their encoded proteins have functionally diverged to a significant degree.

The phenotypes generated by various heteroallelic combinations of *dpy-10* revealed several things. First, we observed no instances of intragenic complementation; animals containing defects at two different locations in *dpy-10* still displayed abnormal morphologies. These results indicate that the alleles we have in hand do not represent defects in functionally independent

Table 4. Mutator-induced alleles of *dpy-10*

Strain	Allele	Insertion site ^a	Orientation ^b	Phenotype ^c
JK1012	<i>q323</i>	T ₇₈₈ -A ₇₈₉	Forward	DLRol
DR1028	<i>m457</i>	T ₇₈₈ -A ₇₈₉	Forward	Dpy
DR1048	<i>m481</i>	T ₇₅₅ -A ₇₅₆	Reverse	Dpy
DR1049	<i>m481m482</i>	Excision ^d	—	WT

^a Nucleotide position of the Tc1 insertions, numbering as in Figure 5.

^b Orientation of the Tc1 open reading frame with respect to the reading frame of *dpy-10*.

^c DLRol, dumpy, left roller; Dpy, dumpy; WT, wild-type.

^d The Tc1 element in DR1048 excised yielding strain DR1049.

Figure 9. Sites of Tc1 insertions in mutator-derived alleles of *dpy-10*. The nucleotide and amino acid sequence of *dpy-10* in the region that contains the insertions is shown. Numbering corresponds to that used in Figure 5. Sites of Tc1 insertions in strains DR1028, DR1048, and JK1012 are indicated with arrows. Large arrows indicate the direction of the Tc1 open reading frame. The phenotypes of these strains are shown in the lower portion of each box. The boxed isoleucine codon was removed in the excision event that generated strain DR1049 from DR1048. Dpy, dumpy; DLRol, dumpy, left roller; WT, wild-type.

domains of the *dpy-10* collagen. Second, a phenotype of intermediate severity, in which animals begin but fail to complete a roll, is often observed in animals carrying one Dpy and one DLRol allele. This phenotype has two manifestations: Lif, in which the roll begins to twist the head upward off the surface of the agar plate, and Hdc, in which the roll begins downward into the plate. These animals very rarely complete a roll. These observations indicate that there may be a continuum of phenotypes

Figure 10. Analysis of excision in *dpy-10* Tc1 mutant strains. Shown is an ethidium bromide stained gel of PCR products generated using primers that flank the region of the Tc1 insertions (see MATERIALS AND METHODS). Lane 1, size standards; lane 2, N2 (wild-type); lane 3, JK1012; lane 4, DR1028; lane 5, DR1048; lane 6, DR1049. The 1810-bp product results from amplification of a Tc1-containing template, whereas the 198-bp product is from amplification of a Tc1-excised template. The lower band in lane 6 (strain DR1049) is actually 195 bp (see text and Figure 9).

with respect to rolling. In addition, observations of heteroallelic phenotypes are consistent with the null phenotype of *dpy-10* being DLRol (see below).

The genetic and molecular analyses of *dpy-2* and *dpy-10* mutant alleles provide a further basis for understanding the functions of the encoded collagens.

Null Phenotype

Several lines of evidence presented here support the idea that DLRol is the null phenotype of *dpy-10*. The DLRol allele *cg36* is a nonsense mutation and behaves like a deficiency across from the other *dpy-10* alleles. The protein encoded by this allele would lack the complete carboxyl-terminal domain with its conserved cysteines. Biochemical studies of vertebrate collagen assembly has shown that the carboxyl-terminal extensions must be present in order for assembly to occur (Rosenbloom *et al.*, 1976; Gelman *et al.*, 1979). Therefore, it is unlikely that the *cg36*-encoded collagen would be able to assemble. Interestingly, this allele does not cause the strongest roller nor the most severe dumpy phenotype. The alleles that produce more severe phenotypes may be producing gene products that get into the cuticle and act as a structural poison.

Other observations presented here are consistent with the idea that DLRol is the null phenotype. *Dpy* alleles across from a deficiency often display an intermediate *Lif* or *Hdc* phenotype. This clearly means that *Dpy* is not null and that the reduction in amount of gene product is likely responsible for the increased roller character. Further, temperature-sensitive alleles always display the *Dpy* phenotype at 15°C and the DLRol phenotype at 25°C. We would expect that the mutant collagens are less stable at the higher temperature because of the inherent temperature dependence of triple helix formation (Prockop *et al.*, 1989). Finally, we find less somatic Tc1 excision in the DLRol strain JK1012 than in the *Dpy* strain DR1028. Our interpretation is that the higher amount of excision in DR1028 results in more functional copies of the *dpy-10* gene in these animals and hence a higher level of *dpy-10* function. We have not, however, addressed the degree to which Tc1 elements are excised accurately in either strain.

dpy-10 differs from the *sqt-1* and *rol-6* collagens whose null phenotypes are both approximately wild-type (Kusch and Edgar, 1986; Kramer, unpublished data). This contrast may reflect functional differences between these collagens during cuticle assembly. The presence of *dpy-10* collagen is required for the attainment of proper body shape. The functions of the *sqt-1* and *rol-6* collagens may be more subtle; in their absence, no fundamental change is observed, at least at the level of light microscopy. It is possible that *dpy-2* and *dpy-10* act as major building blocks in the cuticle structure. Other collagens, such as *sqt-1* and *rol-6* may provide minor structural functions that when absent do not

cause gross morphologic changes. Another reason for this difference could be that the *dpy-2/dpy-10* collagens function in different layers of the cuticle or at different times during development than the *sqt-1/rol-6* collagens. Also, the functions of the *sqt-1/rol-6* collagens may be redundant, so that their absence does not cause gross structural changes.

Glycine Substitutions

Mutations in both *dpy-2* and *dpy-10* result in the substitution of glycine with other amino acids within the triple helical portion of the protein. Glycine is required at every third position in the triple helical domain of each collagen chain because its small side group is the only one that can be accommodated in the center of the triple helix. Bulkier side groups interfere with triple helix assembly and stability. The molecular defect seen in many cases of the human disease osteogenesis imperfecta is a glycine substitution in the alpha 1 chain of type I collagen (reviewed in Prockop *et al.*, 1989). Biochemical studies of these glycine-substituted collagens has revealed that the melting temperature for these collagens can be significantly lower than that of normal type I collagen. Substitution of arginine for glycine within the human alpha 1 (I) triple helix, the same substitution found here in alleles *e8*, *e489*, *q292*, and *q291*, can result in lethal osteogenesis imperfecta (Lamande *et al.*, 1990; Bateman *et al.*, 1987, 1988; Wallis *et al.*, 1990). Glycine substitutions within the triple helical portions of other human collagens are also associated with clinical disorders such as dwarfism (type II) (Vissing *et al.*, 1989) and Ehlers-Danlos syndrome (type III) (Tromp *et al.*, 1989). The clinical manifestations of these disorders may reflect the tissue distributions of the particular collagens involved. That the same substitutions in *C. elegans* collagens result in abnormal morphologies indicates that the functioning of collagens in vertebrates and invertebrates may share at least some features.

It is interesting that certain glycine substitutions result in *Dpy* phenotypes, whereas others produce DLRol animals. This situation may reflect differences in the amount of *dpy-2* or *dpy-10* collagen that can assemble into the cuticle. Under this model, animals displaying the null (DLRol) phenotype contain less *dpy-2* or *dpy-10* collagen in their cuticles than *Dpy* animals. We do not know the rules that govern which positions, when substituted, yield *Dpy* and which yield DLRol. It is clear that identical amino acid substitutions at sites close together can result in different phenotypes. For example, *e489*, an allele that creates a Gly-to-Arg substitution at position 247 in *dpy-2*, displays a temperature-sensitive DLRol phenotype whereas the same substitution at position 253, *q292*, results in a strict *Dpy* phenotype. The underlying molecular property that governs the amount of functional collagen in the cuticle may be the degree of thermal instability introduced by different substitutions.

Studies of mutations in human type I collagen indicate that positional effects play a role in determining the degree of thermal instability produced in glycine substitutions. For example, a glycine to serine substitution in $\alpha 1(I)$ at position 598 drastically decreases thermal stability (Westerhausen *et al.*, 1990), whereas a glycine to serine substitution at position 844 of the same molecule displays normal thermal stability (Pack *et al.*, 1989). In our mutants, the collagens produced from the DLrol alleles may be less thermally stable than the collagens from the Dpy alleles. The phenotypes of temperature-sensitive alleles are consistent with the argument that less functional collagen causes the DLrol phenotype. Temperature-sensitive alleles always display the Dpy phenotype at 15°C and the DLrol phenotype at 25°C. This shift in temperature is apparently enough to destabilize the partially functional collagen (Dpy phenotype) to the point where little or no function is present (null = DLrol phenotype).

Certain mutants display a more severe roller phenotype than that displayed in the null mutant. This observation argues against a simple "amount dictates phenotype" model. Because the phenotype of these mutants (e.g., *sc48*) is more severe than null, the collagens produced in these animals must assemble into the cuticle and disrupt its structure, acting as structural poisons. Thus, we have provided evidence for two possible mechanisms that explain Dpy and DLrol phenotypes in these mutants. These models are not mutually exclusive; it may be that the phenotypes generated in some animals are the result of a decreased level of normally functioning collagen in combination with the presence of a structurally poisonous product in the cuticle.

One interesting "glycine substitution" that is not a mutation occurs in the wild-type *dpy-2* collagen in what would have been the fifth Gly-X-Y triplet. The amino acid at this position is alanine, whose side chain is next smallest after glycine. None of the other sequenced cuticle collagens have an interruption within the first ten Gly-X-Y repeats. Apparently, the function of *dpy-2* is normal despite this "glycine substitution." We do not know the configuration the triple helix adopts at this site. It may be that the helix can still form in this unique case. Alternatively, this may represent a true interruption that attains a nontriple helical configuration. If so, *dpy-2* is the only known cuticle collagen to contain an interruption within the first block of 10 Gly-X-Y repeats.

Splice Acceptor Mutant

dpy-10(e128) has an alteration in intron 2 that changes the highly conserved AG splice acceptor dinucleotide (Mount, 1982; Shapiro and Senapathy, 1987) to AA. The effects on splicing of mutations in the AG dinucleotide have been examined in a variety of systems (Aebi *et al.*, 1986; Mitchell *et al.*, 1986; Cladaras *et al.*, 1987;

Tromp and Prockop, 1988; Su and Lin, 1990; Carstens *et al.*, 1991). In these studies, one or more of three effects is seen. 1) The entire next exon is skipped and the 3' splice site of the next intron is used. 2) The utilization of one or more cryptic 3' splice sites that have an AG is seen. 3) The intron remains unspliced. The effect on 3' splice site utilization in this mutant has been investigated (Aroian *et al.*, 1993). Surprisingly, a variety of 3' splice sites are used in the 50% of transcripts that undergo intron 2 splicing, including the mutant AA site. Because *e128* does not display the null phenotype (it is Dpy), at least some normal *dpy-10* function is provided in this mutant. This phenotype is consistent with the finding that some normal splicing is occurring. Because *dpy-10(+)/Df* animals are wild-type, the amount of *dpy-10* collagen present in *e128* animals must be <50% that of normal animals. *e128* animals were found to have ~10% of the normal level of properly spliced *dpy-10* transcripts (Aroian *et al.*, 1993), a low enough level to possibly produce the Dpy phenotype.

Dominant Mutant

The *dpy-10(cn64)* mutation results in an arginine to cysteine substitution preceding the Gly-X-Y repeats, and it causes a dominant left roller phenotype. This arginine is the fourth residue in a conserved set of amino acids found in this region of all cuticle collagens. The consensus for this conserved sequence is Arg-X-X-Arg-Gln-Ala-Tyr. When this same arginine is changed to a cysteine in the *sqt-1* and *rol-6* collagens, it also causes a dominant phenotype (Kramer, unpublished data). However, the *sqt-1* and *rol-6* mutants display a right roller phenotype. Thus, the equivalent mutation in different collagens can cause the animal to roll in opposite directions. Each of the five dominant mutants thus far sequenced in *sqt-1*, *rol-6*, and *dpy-10* results from the replacement of arginine with cysteine in this conserved region. Extensive disulfide bonding occurs between collagens in the cuticle (Cox *et al.*, 1981). The dominant phenotype exhibited by these mutants may be a result of the participation of the new cysteines in ectopic disulfide bonds. Alternatively, the arginine to cysteine replacement may destroy an important endoprotease cleavage site necessary for proper processing of the collagen molecule. Arg-X-X-Arg is recognized as a cleavage site by members of the subtilisin-like protease family (Barr, 1991). It is also possible that the dominant effects of these substitutions results from a combination of the loss of cleavage and presence of an abnormal cysteine.

Tc1 Mutants

The sequence analysis of *dpy-10* mutants created by insertions of Tc1 elements revealed an unusual situation: the insertions in the Dpy strain DR1028 and the DLrol strain JK1012 are at the same site and in the same orientation despite the different phenotypes exhibited.

There are several potential explanations for this phenomenon. A second closely linked mutation that interacts with *dpy-10* may be present in one of the strains. This second-site mutation would have to be tightly linked to *dpy-10* because both strains have been backcrossed to wild-type at least four times. An alternative explanation is a difference in amount of somatic excision between the Tc1 elements in these two strains. A Tc1 that excises at a higher frequency would generate somatic mosaic animals that could potentially produce more functional copies of *dpy-10*. We found evidence that this is in fact the case. DR1028 appears to have ~10 times more somatic excision than JK1012 as analyzed by PCR. The DLRol (null) phenotype exhibited by strain JK1012 is consistent with the finding that there is much less excision in this strain. A third Tc1 mutant, strain DR1048, contains a Tc1 insertion 33 nucleotides upstream of the ones in DR1028 and JK1012. The phenotype of this strain is Dpy, and the amount of excision is similar to that in DR1028. We do not know the reason for the difference in amount of somatic excision, but there may be sequence differences in the Tc1 elements in these strains.

glp-1 and *mup-1* Suppression

Mutations in several morphologic genes, including alleles of the collagen encoding genes *sqt-1*, *dpy-2*, and *dpy-10*, have been shown to suppress mutations in *glp-1* (Maine and Kimble, 1989) and *mup-1* (Bogaert and Goh, 1991). The *glp-1* gene product is a transmembrane receptor that functions in inductive interactions in the germline. Two somatic cells within the gonad, the distal tip cells, stimulate germ cells to proliferate by a mechanism that requires *glp-1* gene function (Austin and Kimble, 1987; Priess *et al.*, 1987). The *mup-1* gene is required for the proper positioning of body wall muscles during development (Bogaert and Goh, 1991). The alleles of *glp-1* and *mup-1* that are suppressible by mutations in the collagen genes are temperature-sensitive, indicating that they are partial loss-of-function mutations. The collagen gene mutations that suppress the *glp-1* and *mup-1* mutants fall into three molecular classes. The first class consists of glycine substitutions, e.g., *q291* and *q292*, which suppress *glp-1*. The second is the splicing mutant *e128*, which suppresses both *glp-1* and *mup-1*. Third, Maine and Kimble (1989) found that *sqt-1(e1350)*, an arginine-to-cysteine substitution (Kramer, unpublished data), can suppress *glp-1* mutants. In addition, the *sqt-1(e1350)* mutation has been found to partially suppress temperature-sensitive mutations in the dauer constitutive gene *daf-2*. The wide variety of molecular defects that can suppress, combined with the multiple genes that can act as suppressors, as well as be suppressed, makes a model of suppression via direct interaction of the gene products unlikely. Instead, suppression may be occurring through a more general

mechanism. One possibility is that the mutant collagens could accumulate in an abnormal unfolded state and induce a stress response similar to heat shock. The chaperone functions that are induced by such stress might then enable a greater amount of the mutant *glp-1* or *mup-1* proteins to fold, such that partial suppression occurs. Consistent with this hypothesis is the fact that only temperature-sensitive *glp-1* and *mup-1* alleles are suppressed, null alleles are not. Our model requires that some *glp-1* and *mup-1* protein must be produced for suppression to occur. The temperature sensitivity of the suppressible alleles indicates that partially functional gene products are being made.

This model is consistent with observations that mutant actin in *Drosophila* can induce the heat shock response (Parker-Thornberg and Bonner, 1987) and that in *E. coli* the GroEL heat shock protein can suppress a number of temperature-sensitive mutations in unrelated genes (Van Dyk *et al.*, 1989). If this model is correct, suppression may not be occurring through direct physical interactions but rather may represent a general physiological response to the mutant proteins.

ACKNOWLEDGMENTS

We thank Alan Coulson and John Sulston for generously providing cosmid clones and contig map data. We are grateful to Tim Schedl, Don Riddle, and Patrice Alberts for providing worm strains derived from mutator backgrounds. We also thank Robert Bartstead and Robert Waterston for providing the library used to isolate cDNA clones of *dpy-2* and *dpy-10*. The nucleotide sequences reported in this work have been submitted to the GenBank database, accession numbers L12706 (*dpy-2*) and L12692 (*dpy-10*). This work was supported by Public Health Service grant HD-22028 to J.M.K. from the National Institutes of Health.

REFERENCES

- Aebi, M., Hornig, H., Padgett, R.A., Reiser, J., and Weissman, C. (1986). Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* 47, 555-565.
- Aroian, R.V., Levy, A.D., Koga, M., Ohshima, Y., Kramer, J.M., and Sternberg, P.W. (1993). Splicing in *Caenorhabditis elegans* does not require an AG at the 3' splice acceptor site. *Mol. Cell. Biol.* 13, 626-637.
- Austin, J., and Kimble, J. (1987). *glp-1* is required in the germ line for regulation of the decision between mitosis and meiosis in *C. elegans*. *Cell* 51, 589-599.
- Barr, P.J. (1991). Mammalian subtilisins: the long-sought dibasic processing endoproteases. *Cell* 66, 1-3.
- Barstead, R.J., and Waterston, R.H. (1989). The basal component of the nematode dense-body is vinculin. *J. Biol. Chem.* 264, 10177-10185.
- Bateman, J.F., Chan, D., Walker, I.D., Rogers, J.G., and Cole, W.G. (1987). Lethal perinatal osteogenesis imperfecta due to the substitution of arginine for glycine at residue 391 of the alpha1(I) chain of type I collagen. *J. Biol. Chem.* 262, 7021-7027.
- Bateman, J.F., Lamande, S.F., Dahl, H.M., Chan, D., and Cole, W.G. (1988). Substitution of arginine for glycine in the collagen alpha1(I) chain in lethal osteogenesis imperfecta. *J. Biol. Chem.* 263, 11627-11630.

- Bird, D.M. (1992). Sequence comparison of the *Caenorhabditis elegans* *dpy-13* and *col-34* genes, and their deduced collagen products. *Gene* 120, 261–266.
- Birk, D.E., Silver, F.H., and Trelstad, R.L. (1991). Matrix assembly. In: *Cell Biology of Extracellular Matrix*, ed. E.D. Hay, New York: Plenum Press, 221–254.
- Bogaert, T., and Goh, P. (1991). Positioning and maintenance of embryonic body wall muscle attachments in *C. elegans* requires the *mup-1* gene. *Development* 111, 667–681.
- Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* 77, 71–94.
- Carstens, R.P., Fenton, W.A., and Rosenberg, L.R. (1991). Identification of RNA splicing errors resulting in human ornithine transcarbamylase deficiency. *Am. J. Hum. Genet.* 48, 1105–1114.
- Cladaras, C., Hadzopoulou-Cladaras, M., Felber, B.K., Pavlakis, G., and Zannis, V.I. (1987). The molecular basis of a familial apoE deficiency. *J. Biol. Chem.* 262, 2310–2315.
- Cox, G.N., Carr, S., Kramer, J.M., and Hirsh, D. (1985). Genetic mapping of *Caenorhabditis elegans* collagens genes using DNA polymorphisms as phenotypic markers. *Genetics* 109, 513–528.
- Cox, G.N., Kramer, J.M., and Hirsh, D. (1984). Number and organization of collagen genes in *Caenorhabditis elegans*. *Mol. Cell. Biol.* 4, 2389–2395.
- Cox, G.N., Kramer, J.M., Hirsh, D.I., Fields, C.A., and Rosenzweig, B. (1989). Sequence comparison of developmentally regulated collagen genes of *Caenorhabditis elegans*. *Gene* 76, 331–344.
- Cox, G.N., Kusch, M., and Edgar, R.S. (1981). Cuticle of *Caenorhabditis elegans*: its isolation and partial characterization. *J. Cell Biol.* 90, 7–17.
- Cox, G.N., Laufer, J.S., Kusch, M., and Edgar, R.S. (1980). Genetic and phenotypic characterization of roller mutants of *Caenorhabditis elegans*. *Genetics* 95, 317–339.
- Cox, G.N., Shamansky, L.M., and Bolsvenue, R.J. (1990). Haemonchus contortus: evidence that the 3A3 collagen gene is a member of an evolutionarily conserved family of nematode cuticle collagens. *Exp. Parasitol.* 70, 175–185.
- Coulson, A., Sulston, J.E., Brenner, S., and Karn, J. (1986). Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 83, 7821–7825.
- Coulson, A., Waterston, R., Kiff, J., Sulston, J.E., and Kohara, Y. (1988). Genome linking with artificial chromosomes. *Nature* 335, 184–186.
- Eide, D., and Anderson, P. (1985). Transposition of Tc1 in the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 82, 1756–1760.
- Emmons, S.W. (1988). The genome. In: *The Nematode Caenorhabditis elegans*, ed. W.B. Wood, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 47–80.
- Emmons, S.W., Yesner, L., Ruan, K., and Katzenberg, D. (1983). Evidence for a transposon in *Caenorhabditis elegans*. *Cell* 32, 55–65.
- Gelman, R.A., Poppke, D.C., and Piez, K.A. (1979). Collagen fibril formation in vitro. *J. Biol. Chem.* 251, 11741–11745.
- Guo, X., and Kramer, J.M. (1989). The two *Caenorhabditis elegans* basement membrane (type IV) collagen genes are located on separate chromosomes. *J. Biol. Chem.* 264, 17574–17582.
- Hay, E.D. (1991). Collagens and other matrix glycoproteins in embryogenesis. In: *Cell Biology of Extracellular Matrix*, ed. E.D. Hay, New York: Plenum Press, 419–462.
- Higgins, B.J., and Hirsh, D. (1977). Roller mutants of the nematode *Caenorhabditis elegans*. *Mol. Gen. Genet.* 150, 63–72.
- Hodgkin, J.A., and Brenner, S. (1977). Mutations causing transformation of sexual phenotype in the nematode *Caenorhabditis elegans*. *Genetics* 86, 275–287.
- Hodgkin, J., Edgley, M., Riddle, D.L., and Albertson, D.G. (1988). Genetic map. In: *The Nematode Caenorhabditis elegans*, ed. W.B. Wood, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 560–580.
- Hosono, R. (1980). A study of morphology of *Caenorhabditis elegans*: a mutant of *Caenorhabditis elegans* with dumpy and temperature-sensitive roller phenotype. *J. Exp. Zool.* 213, 61–67.
- Innis, M.A., and Gelfand, D.H. (1990). Optimization of PCRs. In: *PCR Protocols: A Guide to Methods and Applications*, ed. M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White, San Diego, CA: Academic Press, 3–12.
- Johnstone, I.L., Shafi, Y., and Barry, J.D. (1993). Molecular analysis of mutations in the *Caenorhabditis elegans* collagen gene *dpy-7*. *EMBO J.* 11, 3857–3863.
- Kramer, J.M., Cox, G.N., and Hirsh, D. (1982). Comparisons of the complete sequences of two collagen genes from *Caenorhabditis elegans*. *Cell* 30, 599–606.
- Kramer, J.M., French, R.P., Park, E., and Johnson, J.J. (1990). The *Caenorhabditis elegans* *rol-6* gene, which interacts with the *sqt-1* collagen gene to determine organismal morphology, encodes a collagen. *Mol. Cell. Biol.* 10, 2081–2089.
- Kramer, J.M., Johnson, J.J., Edgar, R.S., Basch, C., and Roberts, S. (1988). The *sqt-1* gene of *C. elegans* encodes a collagen critical for organismal morphogenesis. *Cell* 55, 555–565.
- Krause, M., and Hirsh, D. (1987). A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 49, 753–761.
- Kusch, M., and Edgar, R.S. (1986). Genetic studies of unusual loci that affect body shape of the nematode *Caenorhabditis elegans* and may code for cuticle structural proteins. *Genetics* 113, 621–639.
- Lamande, S.R., Dahl, H.M., Cole, W.G., and Bateman, J.F. (1989). Characterization of point mutations in the collagen COL1A1 and COL1A2 genes causing lethal perinatal osteogenesis imperfecta. *J. Biol. Chem.* 264, 15809–15812.
- Maine, E.M., and Kimble, J. (1989). Identification of genes that interact with *glp-1*, a gene required for inductive cell interactions in *Caenorhabditis elegans*. *Development* 105, 133–143.
- Mitchell, P.J., Urlaub, G., and Chasin, L. (1986). Spontaneous splicing mutations at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol. Cell. Biol.* 11, 1926–1935.
- Mount, S.M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.* 10, 459–472.
- Okkema, P.G., and Kimble, J. (1991). Molecular analysis of *tra-2*, a sex determination gene in *C. elegans*. *EMBO J.* 10, 171–176.
- Pack, M., Constantinou, C.D., Kalin, K., Nielsen, K.B., and Prockop, D.J. (1989). Substitution of serine for alpha1(I)-glycine 844 in a severe variant of osteogenesis imperfecta minimally destabilizes the triple helix of type I procollagen. *J. Biol. Chem.* 264, 19694–19699.
- Parker-Thornberg, J., and Bonner, J.J. (1987). Mutations that induce the heat shock response of *Drosophila*. *Cell* 51, 763–772.
- Priess, J.R., Schnabel, H., and Schnabel, R. (1987). The *glp-1* locus and cellular interactions in early *C. elegans* embryos. *Cell* 51, 601–611.
- Prockop, D.J., Constantinou, C.D., Dombrowski, K.E., Hojima, Y., Kadler, K.E., Kuivaniemi, H., Tromp, G., and Vogel, B.E. (1989). Type I procollagen: the gene-protein system that harbors most of the mutations causing osteogenesis imperfecta and probably more common

heritable disorders of connective tissue. *Am. J. Med. Genet.* 34, 60–67.

Rosenbloom, J., Endo, R., and Harsch, M. (1976). Termination of procollagen synthesis by puromycin. Evidence that assembly and secretion require a COOH-terminal extension. *J. Biol. Chem.* 251, 2070–2074.

Rosenzweig, B., Liao, W., and Hirsh, D. (1983). Sequence of the *C. elegans* transposable element Tc1. *Nucleic Acids Res.* 11, 4201–4209.

Sambrook, E.F., Fritsch, J., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Shapiro, M.B., and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15, 7155–7174.

Sigurdson, D.C., Spanier, G.J., and Herman, R.K. (1984). *Caenorhabditis elegans* deficiency mapping. *Genetics* 108, 331–345.

Singh, R.N., and Sulston, J.E. (1977). Some observations on moulting in *Caenorhabditis elegans*. *Nematologica* 24, 63–71.

Su, T., and Lin, L. (1990). Analysis of a splice acceptor site mutation which produces multiple splicing abnormalities in the human argininosuccinate synthetase locus. *J. Biol. Chem.* 265, 19716–19720.

Toole, B.P. (1991). Proteoglycans and hyaluronan in morphogenesis and differentiation. In: *Cell Biology of the Extracellular Matrix*, ed. E.D. Hay, New York: Plenum Press, 305–342.

Tromp, G., Kuivaniemi, H., Stolle, C., Pope, F.M., and Prockop, D.J. (1989). Single base mutation in the type III procollagen gene that

converts the codon for glycine 883 to aspartate in a mild variant of Ehlers-Danlos syndrome IV. *J. Biol. Chem.* 264, 19313–19317.

Tromp, G., and Prockop, D.J. (1988). Single base mutation in the pro-alpha2(I) collagen gene that causes efficient splicing of RNA from exon 27 to exon 29 and synthesis of a shortened but in-frame pro-alpha2(I) chain. *Proc. Natl. Acad. Sci. USA* 85, 5254–5258.

Van Dyk, T.K., Gatenby, A.A., and LaRossa, R.A. (1989). Demonstration by genetic suppression of interaction of GroE products with many proteins. *Nature* 342, 451–453.

Vissing, H., D'Alessio, M., Lee, B., Ramirez, F., Godfrey, M., and Hollister, D.W. (1989). Glycine to serine substitution in the triple helical domain of pro-alpha1(II) collagen results in a lethal perinatal form of short-limbed dwarfism. *J. Biol. Chem.* 264, 18265–18267.

von Mende, N., Bird, D., Albert, P.S., and Riddle, D.L. (1988). *dpy-13*: a nematode collagen gene that affects body shape. *Cell* 55, 567–576.

Wallis, G.A., Starman, B.J., Schwartz, M.F., and Byers, P.H. (1990). Substitution of arginine for glycine at position 847 in the triple-helical domain of the alpha1(I) chain of type I collagen produces lethal osteogenesis imperfecta. *J. Biol. Chem.* 265, 18628–18633.

Westerhausen, A., Kishi, J., and Prockop, D.J. (1990). Mutations that substitute serine for glycine alpha1-598 and glycine alpha1-631 in type I procollagen. *J. Biol. Chem.* 265, 13995–14000.

Yochem, J., and Greenwald, I.S. (1989). glp-1 and lin-12, genes implicated in distinct cell-cell interactions in *C. elegans*, encode similar transmembrane proteins. *Cell* 58, 553–563.