# Gene Interactions in *Caenorhabditis elegans* Define DPY-31 as a Candidate Procollagen C-Proteinase and SQT-3/ROL-4 as Its Predicted Major Target

**Jacopo Novelli,\* Shawn Ahmed† and Jonathan Hodgkin\*,[1]**

*\*Genetics Unit, Department of Biochemistry, Oxford OX1 3QU, United Kingdom and
†Department of Biology, Chapel Hill, North Carolina 27599-3280*

## ABSTRACT

Zinc metalloproteases of the BMP-1/TOLLOID family (also known as astacins) are extracellular enzymes involved in important developmental processes in metazoans. We report the characterization of the *Caenorhabditis elegans* gene *dpy-31*, which encodes the first essential astacin metalloprotease identified in this organism. Loss-of-function mutations in *dpy-31* result in cuticle defects, abnormal morphology, and embryonic lethality, indicating that *dpy-31* is required for formation of the collagenous exoskeleton. DPY-31 is widely expressed in the hypodermal cells, which are responsible for cuticle secretion. We have investigated the *dpy-31* function through reversion analysis. While complete reversion can be obtained only by intragenic suppressors, reversion of the Dpy-31 lethal phenotype also can be caused by dominant extragenic suppressors. Nine extragenic suppressors carry mutations in the uniquely essential collagen gene *sqt-3*, which we show is the same gene as *rol-4*. Most mutations exhibit the unusual property of exclusively dominant suppression and all affect the sequence of the SQT-3 collagen C terminus. This suggests that DPY-31 is responsible for C-terminal proteolytic processing of collagen trimers and is therefore a structural and functional homolog of vertebrate BMP-1. The results also demonstrate the critical importance of the collagen C-terminal sequence, which is highly conserved among all 49 members of the SQT-3 subfamily.

COLLAGENS constitute a major class of extracellular structural proteins in metazoans. Their importance is nowhere greater than in nematodes, in which the cuticle is largely constructed from an elaborate matrix of collagen molecules. In *Caenorhabditis elegans*, collagens account for ∼80% of the cuticle mass (Cox *et al.* 1981). The *C. elegans* cuticular collagen gene family includes >170 members, which are 1–2 kb long and encode proteins of ∼300 amino acids (aa). *C. elegans* collagen are structurally similar to vertebrate nonfibrillar FACIT collagens (SHAW and OLSEN 1991), although considerably smaller in size. Like their vertebrate counterparts, nematode collagens possess Gly-X-Y repeats, which are required for formation of triple helical trimers. After assembly in the endoplasmic reticulum (ER), collagen trimers are secreted into the extracellular environment where they are highly crosslinked, thus forming a polymeric matrix (for a review on *C. elegans* cuticle collagens, see JOHNSTONE 2000).

Genetic analysis has provided extensive insights into the process of cuticle formation in *C. elegans*. Many morphological mutants affecting the cuticle have been isolated in random and directed mutagenesis screens. These mutants are divided into distinct gene classes on the basis of visible phenotypes. Six types of viable morphological mutants have been described: *dpy* (*dumpy*), which is shorter than wild type (WT); *bli*, which has a *bli*stered cuticle; *rol* (*rol*ler), which has a helically twisted body; *sqt* (*squat*), which has both a dominant Rol and a recessive Dpy phenotype; and *lon*, which is *lon*ger than WT (BRENNER 1974). The molecular identity of most defined morphogenetic loci has now been determined; on one hand, 19 genes in these classes encode cuticle collagens, indicating that, despite the high level of redundancy, many collagens play distinct roles in cuticle structure. On the other hand, four morphogenetic loci, all of which were isolated in the course of the first *C. elegans* mutagenesis screens (BRENNER 1974), encode core enzymatic components of the collagen assembly pathway, which is largely conserved between nematodes and vertebrates (for a recent review, see MYLLYHARJU and KIVIRIKKO 2004). The essential gene *bli-4* encodes a furin proprotein convertase, which is thought to be involved in N-terminal proteolytic cleavage of collagen precursors (THACKER *et al.* 1995). Furin proteases have been implicated in collagen processing in mammals (IMAMURA *et al.* 1998). *dpy-18* codes for one component of the prolyl 4-hydroxylase complex, which is required for collagen chain assembly in vertebrates (WINTER and PAGE 2000). The *dpy-11* gene encodes an ER-bound thioredoxin-like protein, which may also play a role in

[1]*Corresponding author:* Genetics Unit, Department of Biochemistry, South Parks Rd., Oxford OX1 3QU, United Kingdom. E-mail: jah@bioch.ox.ac.uk

collagen assembly, presumably during trimer formation (Ko and Chow 2002). Finally, *bli-3* codes for a dual oxidase, which is the putative catalyst of collagen tyrosine-derived crosslinking (Edens *et al.* 2001; Simmer *et al.* 2003). Herein, we examine a newly defined morphogenetic gene, which was named *dpy-31* after its Dpy mutant phenotype. *dpy-31* mutants also display a lethal phenotype, indicating that this gene is required for an essential aspect of cuticle formation. Upon cloning, *dpy-31* was found to encode a zinc-metalloprotease of the BMP-1 (*bone morphogenetic protein-1*)/TOLLOID family. Proteases of this family are involved in important developmental processes, such as pattern formation, extracellular matrix (ECM) maturation, remodeling, and degradation (Bond and Beynon 1995). In particular, BMP-1 is responsible for proteolytic maturation of the C-propeptides of various collagen precursors. Reversion analysis of *dpy-31* revealed unusual genetic interactions, which implicate the C-terminal region of a uniquely essential collagen, SQT-3, as a major target for processing by DPY-31. Involvement of zinc-metalloproteases in collagen maturation therefore appears to be crucial for collagen deposition and a conserved feature of the metazoan lineage.

## MATERIALS AND METHODS

**Strains:** Maintenance and handling of *C. elegans* were performed as described by Sulston and Hodgkin (1987). Worms were cultured at 20°, unless otherwise stated. Bristol N2 was used as WT.

The following mutant strains were used:

LG II: *rrf-3(pk1426)*;
LG III: *dpy-31(e2770, e2919, e2920, ju345); unc-32(e189); sma-3 (e491); mab-5(e1239); unc-36(e251, e548); lon-1(e185); unc-119 (e2498); nDf16*;
LG IV: *fem-1(hc17)*;
LG V: *sqt-3(e24, e2117, sc63, sc8, sc42, e2639) sqt-3(e2809, e2888, e2889, e2890, e2896, e2901, e2906, e2911); ctDf1; arDf1*;
LG X: *xol-1(y9)*;
Arrays: *kaIs(col-19::GFP)*.

**Mapping and cloning of *dpy-31*:** After isolation in the *mrt-2* mutant background, *dpy-31(e2770)* was mapped to the center of LG III and therefore tightly linked to the *mrt-2* mutation. It was separated from *mrt-2* by constructing a *dpy-31 unc-32* double-mutant chromosome and then by removing the *unc-32* marker by recombination against WT. Three-factor mapping was performed by using the triple-marked strain *sma-3 mab-5 unc-36*, whereby *dpy-31* was placed between the two outside markers at position $-0.80$ (39 recombinants were examined). Snip-SNP mapping was performed as described by Wicks *et al.* (2001). Four double mutants containing *dpy-31* and a flanking marker were constructed. These were the following: *lon-1 dpy-31; sma-3 dpy-31; dpy-31 unc-36*; and *dpy-31 unc-32*. We analyzed, respectively, 13 Lon-1, 1 Sma-3, 11 Unc-36, and 30 Unc-32 recombinants. Snip-SNP mapping defined a genomic region between cosmid clones F31E3 and T20B12. Clones from this interval were injected into the gonad of *e2770* worms as described by Mello and Fire (1995) at an average concentration of 10 µg/ml with 20 µg/ml of pTG96 (a green fluorescent protein (GFP) reporter plasmid; Yochem *et al.* 1998). Following identification of R151 as a rescuing clone, the pD31FL

construct including the genomic sequence of gene R151.5 was generated as below. PCR was performed on clone R151 with the following primers (bases added at the 5′ end are in lowercase; *Pst*I sites are underlined): PstIPrF sense: 5′-aaactg cagTCCAACACCAGACTTTTCTCC-3′ and PstItoh2R antisense: 5′-tatttgacgTCCATCAGATCACCCAGC-3′. The 8.9-kb product was digested with *Pst*I and ligated into vector pGEM-5Zf (Promega, Madison, WI) to generate pD31FL. PD31FL was injected into *e2770* mutants at a concentration of 10 µg/ml with 20 µg/ ml of pTG96. Gene R151.5 was amplified by PCR from *e2770* genomic DNA and sequenced with gene-specific primers.

***dpy-31* reporter gene:** pD31P3.4G was generated as described below. A 3.4-kb PCR product was amplified from cosmid R151 using the following primers: PstIPrF sense (primer used for construct pD31FL) and BaPrR antisense (*Bam*HI site is underlined): 5′-cgggatcccgGCTGAAATTAAAGTTTAAAG-3′. The resulting PCR product was digested and cloned directionally into pPD96.04 (A. Fire, personal communication). pD31P3.4G was injected into *unc-119* worms at a concentration of 20 µg/ml with 20 µg/ml of pDP#MM016b (an *unc-119* rescuing construct; Maduro and Pilgrim 1995). To ascertain consistency in the expression pattern, we generated multiple transgenic *unc-119* lines carrying pDP#MM016b and pD31P3.4G on an extrachromosomal array. Every line was found to display the same temporal and spatial pattern of expression.

pD31P3.4G-N, a version of pD31P3.4G devoid of nuclear localization signal (NLS), was generated by digestion of pD31P3.4G with *Kpn*I followed by religation. PD31P3.4-N was directly injected into N2 worms.

**RNAi of *dpy-31*:** For post-transcriptional silencing of the *dpy-31* gene, RNA interference (RNAi) was performed by double-stranded RNA injection of N2 hermaphrodites (Fire *et al.* 1998). The double-stranded RNA injected corresponded to exons 2 and 3 of *dpy-31* and the intervening sequence between the two exons. The following primers were used to amplify the fragment from N2 genomic DNA (T3 and T7 polymerase recognition sequences are in lowercase): T3toh23F sense 5′-attaaccct cactaaagGTTGGTAGCATGGGATCG-3′ and T7toh23R antisense 5′-aatacgactcactatagTGCGTTTAGACTTAATCC-3′.

**Screens for suppressors of *dpy-31*:** Nine suppressor screens were performed, three using ethyl methane sulfonate (EMS) and six using *N*-ethyl *N*-nitrosourea (ENU). In all screens, the *e2770* allele of *dpy-31* was used. The high level of lethality and infertility associated with *e2770* precludes accurate measurement of the number of genomes tested in each suppressor screen. We estimate that at least $10^6$ (EMS) and at least $2 \times 10^6$ genomes (ENU), in total, were tested.

*EMS mutagenesis:* Mutagenesis of *e2770* was performed as described by Brenner (1974). Following treatment with EMS, worms were plated and grown for 6–10 days before being shifted to restrictive temperature. Plates were then inspected for viable revertants.

*ENU mutagenesis:* We used the protocol devised by De Stasio and Dorman (2001). Procedure was otherwise the same as the one followed in EMS screens.

**Mutation analysis of WT revertants of *e2770*:** Mutations in WT revertants of *e2770* were analyzed either by sequencing or by amplification refractory mutation system-polymerase chain reaction (ARMS-PCR; Newton *et al.* 1989). ARMS-PCR was performed by using a gene-specific reverse primer (LastR: antisense 5′-AATACCACCGCTGTCTGTCC-3′) paired in separate PCRs with three allele-specific forward primers. These were the following (the variable sequence is underlined): LeuF (sense 5′-GCAACTACTGATATGGTTGTACT-3′), which is specific for the Leu codon found in WT; SerF (sense 5′-GCAACTAC TGATATGGTTGTATC-3′), which is specific for the Ser codon found in "Ser" revertants of *e2770*; and ProF (sense 5′-GCAAC TACTGATATGGTTGTACC-3′), which is specific for the Pro codon found in *e2770* mutants.

**Mapping and cloning of *rol-4*:** Five cosmid clones contained within the interval covered by the deficiency *ctDf1* were injected into *rol-4(sc8)* worms at a concentration of 10 μg/ml with 20 μg/ml of pTG96. Clone F23H12 rescued the left-handed Roller (Lrol) phenotype in transgenic *sc8* worms. *sqt-3* was amplified by single-worm PCR (WICKS *et al.* 2001) and sequenced with gene-specific primers. The procedure was similar for all the *sqt-3* suppressors associated with a LRol phenotype, *i.e.*, *sc8*, *e2809*, *e2888*, *e2889*, *e2890*, *e2896*, and *e2901*. The non-Rol *sqt-3(sup)* alleles *e2906* and *e2911* were identified by their failure to complement *sqt-3* alleles. Other alleles assigned to *sqt-3/rol-4* for which the sequence alteration had not been identified were *b238* and *sc42*. No *sqt-3* mutation was found in *b238*; this allele complements *sqt-3/rol-4* alleles and probably defines a distinct locus. *sc42* was sequenced with gene-specific primers and the sequence alteration was determined. Finally, the *e2639* deletion allele of *sqt-3*, which was recovered in a *mut-7* mutator background (M. SKIPPER and J. HODGKIN, unpublished results), was assigned to *sqt-3* on the basis of the results of complementation tests. The *e2639* allele of *sqt-3* was then sequenced with gene-specific primers.

**Isolation of new *dpy-31* alleles:** Additional alleles of *dpy-31* were obtained by crossing EMS-mutagenized N2 males with females of genotype *dpy-31(e2770) unc-36(e548) III; fem-1(hc17) IV; sqt-3(sc8)/+ V; xol-1(y9) X* at 22.5°. The *unc-36* mutation is tightly linked to *dpy-31* and thus acts as marker for the *dpy-31* tester allele. The temperature-sensitive (*ts*) mutation *fem-1(hc17)* results in feminization of hermaphrodite worms at restrictive temperature. The *sc8* dominant suppressor of the *dpy-31* lethality allows viability of homozygotes for *dpy-31*. Finally, the *xol-1* mutation (XO lethal) leads to the death of the male cross-progeny.

Approximately 5600 $F_1$ progeny, of which half (2800) would have carried the *sc8* suppressor, were scored for Dpy non-Unc animals. Candidates were cloned. ARMS-PCR performed using the primers described above was used to identify new alleles of *dpy-31*. Two Dpy lines, which appeared WT at the nucleotide position changed in *e2770* mutants, were analyzed further. Sequencing of *dpy-31* in these strains identified novel mutations in the coding region. The lines were retained and the new alleles of *dpy-31* were named, respectively, *e2919* and *e2920*. *ju345* was assigned to *dpy-31* on the basis of its failure to complement *e2770*.

**Test for the effect of the *sqt-3(sup)* dosage on the viability of *dpy-31* worms:** Worms of genotype *dpy-31(e2770)/unc-32 (189); sqt-3(sup)/+* were identified by inspection of the $F_1$ progeny for the presence of viable Dpy and Unc segregants. Animals of correct genotype were grown at 22.5° and moved daily to a fresh plate. The $F_1$ progeny were scored. We tested the *sqt-3(sup)* alleles *sc8*, *e2889*, *e2901*, *e2906*, and *e2911*. As *unc-32* is tightly linked to *dpy-31*, recombination occurring between the two markers could be neglected. As a control, $F_1$ progeny of *dpy-31/unc-32* double heterozygotes were scored.

## RESULTS

***dpy-31*: phenotypic characterization:** The *dpy-31* locus of *C. elegans* was originally defined by a single mutation, *e2770*, which was recovered as a spontaneous event in a *mrt-2(e2663)* mutator background (AHMED and HODGKIN 2000). In subsequent analysis, two further alleles of the gene, *e2919* and *e2920*, were isolated in a directed EMS screen. A fourth allele of *dpy-31*, *ju345*, was a gift from Andrew Chisholm, who recovered it independently in an EMS screen.

*dpy-31(e2770, e2919, e2920)* worms exhibit a strong recessive Dpy phenotype (Figure 1B) and are *ts* lethal. These *dpy-31* mutants are barely viable at permissive temperature (15°), but totally inviable at restrictive temperature (20°–25°). *ju345* animals display a milder Dpy phenotype and are ~80% viable at 25°, indicating that *ju345* is a hypomorphic allele of *dpy-31*. Embryos and first larval stages (L1) of *dpy-31* are particularly sensitive to temperature upshifts. At 25°, *dpy-31* embryos develop normally until the threefold stage; however, after completing elongation, mutant embryos retract in length and fail to hatch (Figure 1C). Rare hatchlings are arrested as L1 with a distinctive "retracted lumpy" appearance (Figure 1D). This phenotype has been observed in other mutants, such as *sqt-3(e24, e2117)* (PRIESS and HIRSH 1986; VAN DER KEYL *et al.* 1994), *bli-4(s90)* (THACKER *et al.* 1995), and *pdi-2(RNAi)* (WINTER and PAGE 2000). In all these cases, the lethal Dpy phenotype was found to be associated with defects in cuticle synthesis. Inspection of the mutant phenotype thus suggests that DPY-31 plays an essential role in formation of the collagenous exoskeleton.

To further characterize the defect in cuticle structure of *dpy-31* mutants we made use of a strain carrying the transgenic array *kaIs(col-19::GFP)*, which contains a fusion of the collagen gene *col-19* and a GFP reporter (THEIN *et al.* 2003). In adult WT worms, the COL-19::GFP protein localizes to the cortex of annuli, which are circumferential ridges covering the body of the worms, and to the alae, which are tripartite structures running perpendicular to the annuli (Figure 1E).

Construction of a line carrying the *e2770* mutation together with the *kaIs(col-19::GFP)* array revealed that COL-19::GFP assembly is aberrant in *dpy-31* mutants grown at permissive temperature (Figure 1F). In *e2770* worms, COL-19::GFP fails to localize to the alae; instead, in this region an irregular mesh of fibers, extending into the annuli, can be seen. Furthermore, intensely bright patches of COL-19::GFP were observed in the animals' head region. These patches probably represent unsecreted COL-19::GFP accumulating intracellularly in the seam cells (A. PAGE, personal communication). The disruption of COL-19::GFP assembly in *e2770* worms is similar to that observed by THEIN *et al.* (2003) for *dpy-5*, which encodes a cuticle collagen. These observations support a role for *dpy-31* in cuticle formation.

***dpy-31*: mapping and cloning:** Visible marker mapping placed *dpy-31* 0.1 map unit to the right of *sma-3* (−0.90) on LG III (Figure 2). We also found that *dpy-31* is uncovered by the deficiency *nDf16*, extending between positions −1.05 and −0.23. Worms of genotype *e2770/nDf16* were indistinguishable from *e2770* homozygotes, showing that *e2770* behaves as a classic loss of function and indicating that this allele represents the null state of *dpy-31* at 25°.

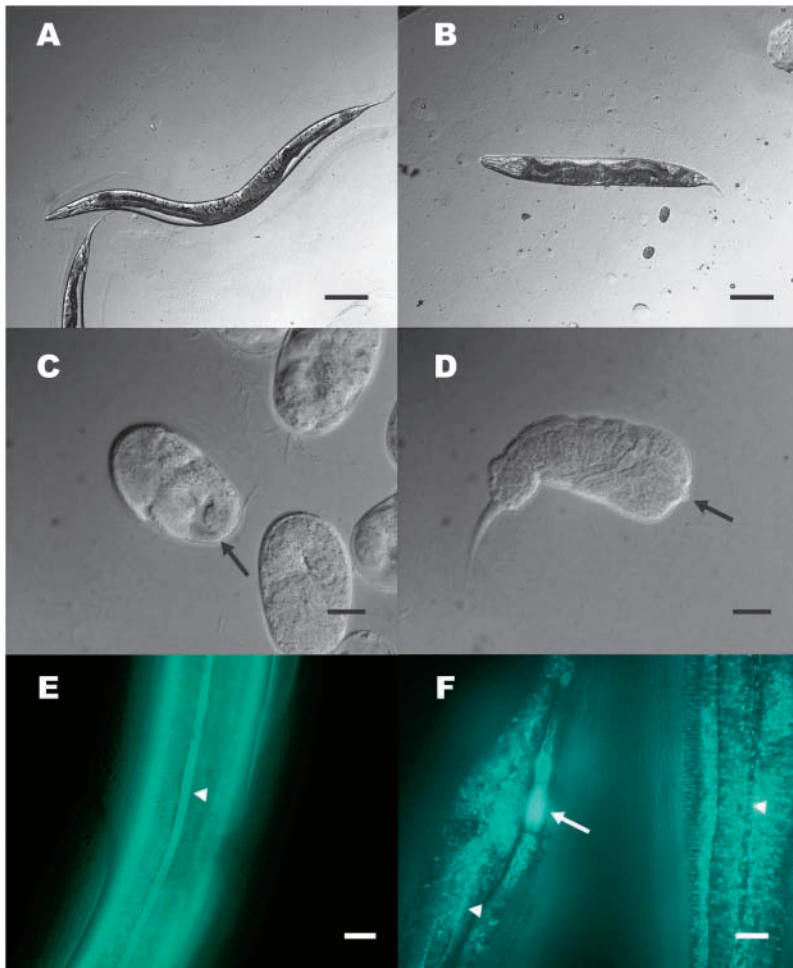To refine the map position, we performed snip-SNP mapping, whereby we reduced the interval to 15 geno-

FIGURE 1.—Phenotype of *dpy-31* mutants. (A) WT adult worm. (B) *dpy-31(e2770)* adult raised at 15°. (C) *e2770* embryos raised at 25°: the cuticle surface in tail region of a dead embryo (solid arrow) appears abnormal. (D) *e2770* arrested L1 larva. Note the severe internal disorganization resulting from hypercontraction. The solid arrow indicates the pharynx. (E) WT worm carrying the *col-19::GFP* reporter gene. The alae (open arrowhead) are visible. The circumferential rings perpendicular to the alae are the annuli. (F) *col-19::GFP* expression in *dpy-31* mutants. Two worms are depicted. An area of concentrated fluorescence (open arrow) can be seen in the head region of a worm. The arrowheads mark the position of the alae. Bars: A and B, 0.1 mm; C–F, 10 μm.

mic clones. Within these we identified a single positive rescuing clone (R151; Figure 2). In further tests, we determined that a WT copy of gene R151.5 is sufficient to rescue the *dpy-31* function. Gene R151.5 had already been named *toh-2* (*tollish-2*) on the basis of the sequence similarity to *Drosophila melanogaster tolloid* (http://www.wormbase.org). In light of current data, the name *toh-2* is misleading; we therefore renamed gene R151.5 after its mutant phenotype, *i.e.*, *dpy-31*.

**Structure of DPY-31 and identification of the DNA lesions in *dpy-31* mutants:** *dpy-31* encodes a predicted zinc-metalloprotease of the BMP-1/TOLLOID family. Peptidases of this family are also known as astacins and are characterized by a zinc-binding catalytic domain, which is highly conserved across metazoans. In addition to the catalytic domain, astacin metalloproteases usually also contain a variable number of C-terminal extensions, which are thought to be required for regulation or modulation of activity. All astacin proteases identified so far are secreted or plasma membrane bound (BOND and BEYNON 1995). TOLLOID is an essential component of pattern formation in Drosophila (MARQUÉS *et al.* 1997). Mammalian BMP-1 is a major ECM regulator, which, in addition to its key role in the BMP-signaling pathway,

is responsible for processing of ECM structural components (PAPPANO *et al.* 2003 and references cited therein). The deduced DPY-31 protein contains a signal peptide (SP: aa 1–16; Figure 3A), a prodomain (PD: aa 16–133), an astacin catalytic domain (CAT: aa 134–327), an epidermal growth factor motif (EGF: aa 344–370), a CUB (*C*1r/*C*1s, embryonic sea urchin protein *U*egf, *B*MP-1) domain (CUB: aa 371–484), and a *t*hrombo*sp*ondin type I repeat (TSP-1: aa 494–539). DPY-31 shows 29% overall identity and 43% similarity to the shorter splicing variant of human BMP-1.

Sequencing of the *e2920, e2919, ju345,* and *e2770* alleles of *dpy-31* identified mutations, respectively, in exons 4, 5, 5, and 6 of the gene. *e2770*, which arose spontaneously in the *mrt-2* mutant background, is a T-to-C transition resulting in a Leu-to-Pro substitution at residue 469 of the encoded protein (L469P; Figure 3, A and B). This type of mutation would not be commonly induced by the standard chemical mutagen EMS, which causes chiefly GC-to-AT transitions (SULSTON and HODGKIN 1987). At the protein level, the *e2770* mutation maps to the CUB domain, which is a protein-protein interaction module composed of 10 β-strands arranged in a jellyroll-type topology (ROMÃO *et al.* 1997). The Pro substitution
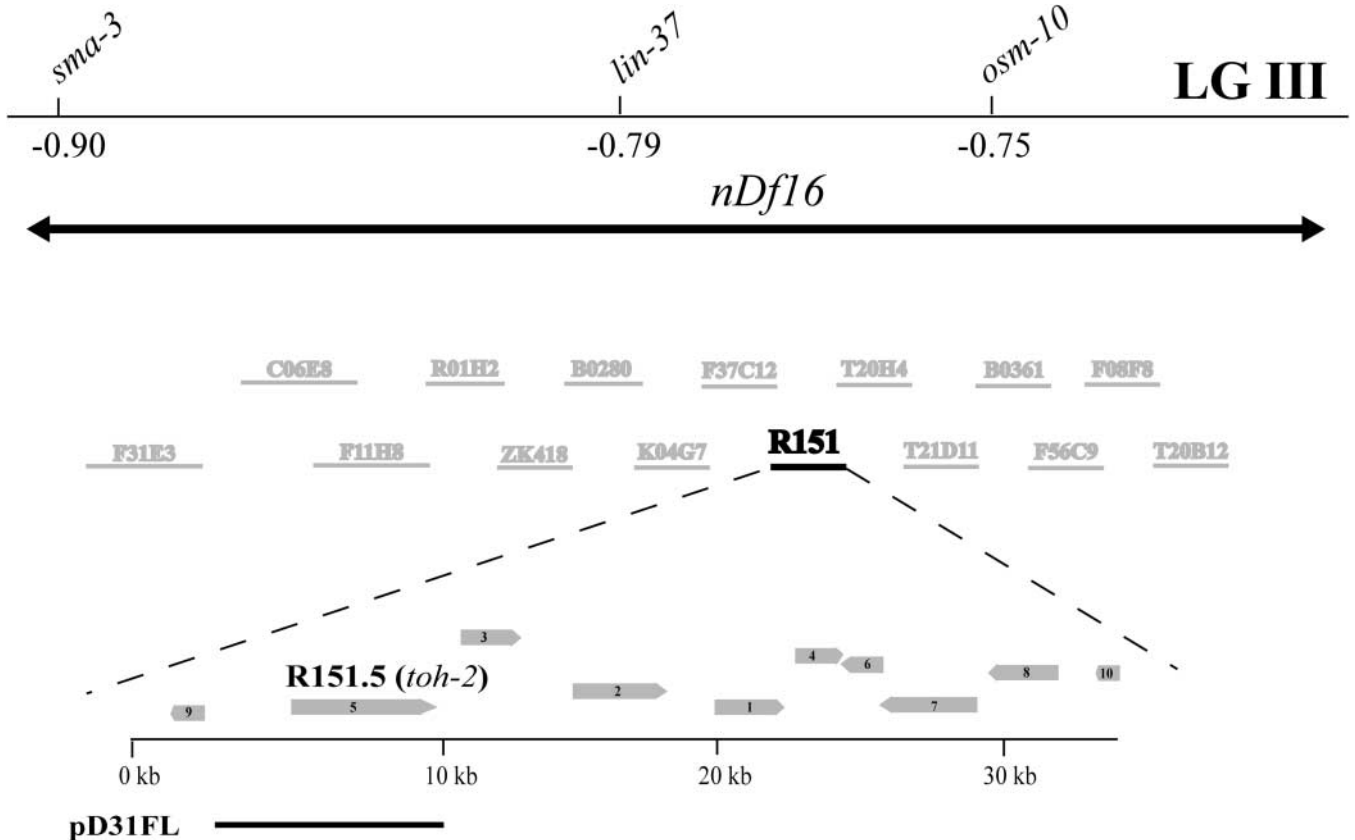
FIGURE 2.—Genetic and physical map of the *dpy-31* locus. The deficiency *nDf16* is depicted by a double arrow. The 15 cosmid clones defined as a physical interval by snip-SNP mapping are indicated by thin bars. The open reading frames in the rescuing cosmid R151 and the rescuing construct pD31FL are depicted by thick bars.

affects the consensus sequence of β-strand 9, thus potentially leading to the disruption of the CUB domain fold.

In the three EMS-induced mutants, *e2920, e2919*, and *ju345*, GC-to-AT transitions were detected. *e2920* maps to the conserved "Met-turn" region of the CAT domain, where it causes the substitution of an invariant His (H280Y; Figure 3, A and B), which is adjacent to a Tyr residue involved in zinc coordination (BOND and BEYNON 1995). This conserved Tyr residue is essential for catalytic activity in BMP-1 (YIALLOUROS *et al.* 2000), indicating that DPY-31 is also likely to function as a protease. *e2919* introduces a stop codon at position 336 of DPY-31 (Q336*Ochre*; Figure 3, A and B). The *e2919* allele is predicted to encode a severely truncated protein lacking the EGF, CUB, and TSP-1 domain. Since the CUB domains are essential for catalytic activity and secretion of BMP-1 (HARTIGAN *et al.* 2003), it seems likely that this allele represents a *dpy-31* null. Finally, the hypomorphic allele *ju345* is a missense mutation causing a Gly-to-Glu substitution downstream of the CAT domain (G339E; Figure 3, A and B). *e2770, e2920*, and *ju345* affect conserved motifs that are also present in BMP-1 (Figure 3).

Following identification of *dpy-31*, we attempted to phenocopy the Dpy-31 phenotype by RNAi. RNAi of *dpy-31* occasionally resulted in an embryonic lethal phenotype in the progeny similar to that exhibited by *dpy-31* mutants. However, we found that RNAi phenocopies the Dpy-31 lethality only at low penetrance [5% of the $F_1$ progeny in WT; $n = 5$; 24.5% of the $F_1$ progeny in *rrf-3* mutants, which are hypersensitive to RNAi ($n = 4$; SIMMER *et al.* 2002)]. Relative insensitivity to RNAi appears to be a common feature of astacin protease genes in *C. elegans* (MÖHRLEN *et al.* 2003).

**DPY-31 is expressed in the hypodermis:** To investigate the expression pattern of *dpy-31*, we generated a reporter gene (pD31P3.4G) containing the *dpy-31* 5′ flanking region fused to the GFP sequence with a NLS. Transgenic worms carrying pD31P3.4G express GFP in most hypodermal nuclei throughout the life cycle (Figure 4). Expression was first detected in embryos at the twofold stage. The GFP signal was particularly intense in embryos (Figure 4A) and early larvae (Figure 4B).

A version of the pD31P3.4G construct devoid of NLS (pD31P3.4G-N) confirmed this expression pattern. In transgenic animals carrying pD31P3.4G-N, GFP was detected in most hypodermal cells, except in the seam cells of larval stages (Figure 4, C and C′). Expression of DPY-31 is particularly pronounced at all life stages in the amphid socket cells at the anterior end of the
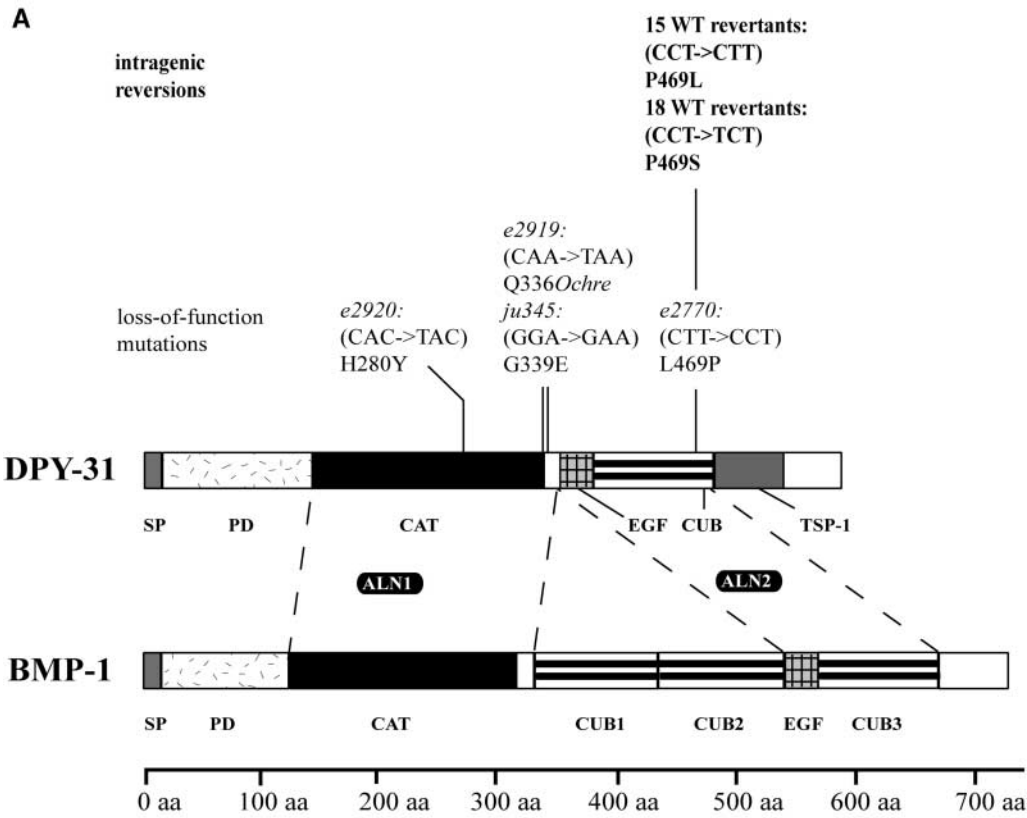
FIGURE 3.—*dpy-31* encodes a BMP-1 homolog. (A) Comparison of *C. elegans* DPY-31 (corresponding to gene R151.5) and human BMP-1 (isoform 1). BMP-1 has two additional CUB domains but lacks the TSP-1 motif. The molecular lesions identified in four *dpy-31* alleles and the intragenic reversion events (boldface type) are shown above the protein structure. The corresponding codon changes are in parentheses. (B) ClustalW alignments of the CAT domain (ALN1) and EGF and CUB domains (ALN2) of DPY-31 and BMP-1. The CUB domain of DPY-31 is most homologous to the third CUB domain of BMP-1 (CUB3), which is shown in the alignment. The positions affected by the mutations identified in *dpy-31* are shown.
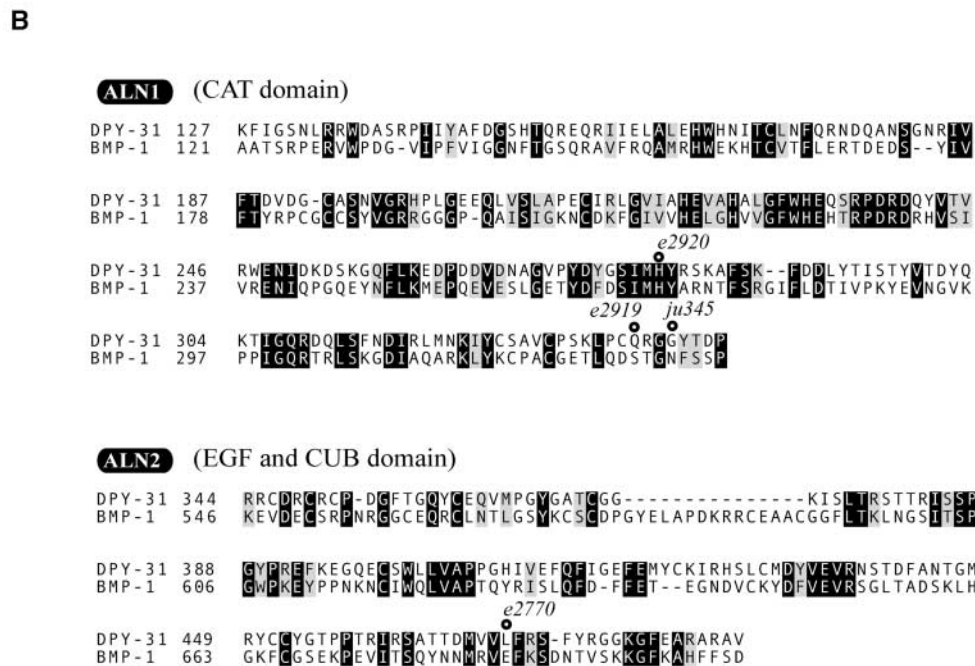
animal and in the rectal epithelial cells at the posterior end (Figure 4, C and C′). In L4 and adults, DPY-31 expression was also seen in the vulval epithelial cells. Fluorescence was also observed in some head neurons, although the identity of such neurons has not been determined.

**Investigation of the *dpy-31* function through reversion analysis:** In vertebrates, the astacin metalloprotease

BMP-1 acts as a procollagen C-proteinase (pCP): BMP-1 converts collagen trimers into a mature form by cleaving their C-terminal propeptides (KESSLER *et al.* 1996). Taken together, our results suggested that DPY-31 could be involved in processing of cuticle collagens. However, the extreme redundancy of the collagen gene family made it arduous to identify potential targets of DPY-31 by biochemical means. We therefore sought to charac-
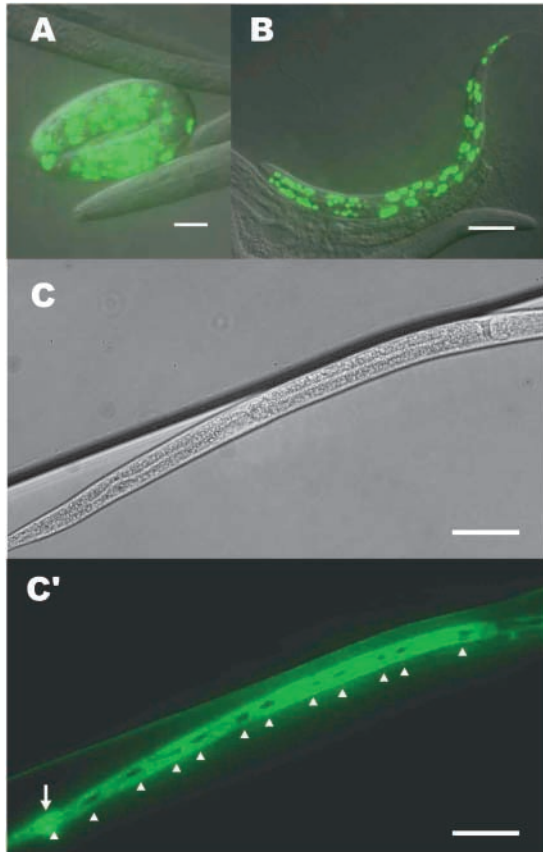
FIGURE 4.—Tissue-specific localization of *dpy-31.* Expression of GFP reporter driven by the *dpy-31* promoter is detected in the hypodermal cells. (A) Fully elongated embryo and (B) L2 larva carrying a multicopy array of pD31P3.4G, which contains the *dpy-31* promoter sequence fused to GFP with a NLS. (C and C′) L4 larva carrying a multicopy array of pD31P3.4G-N, which is a NLS negative version of pD31P3.4G. No expression of DPY-31 is detected in the seam cells (open arrowheads), which have not yet undergone fusion in this animal. The rectal epithelial cells expressing DPY-31 are indicated by an open arrow. Bars: A, 10 μm; B, 20 μm; C and C′, 50 μm.



FIGURE 5.—Revertants of *e2770.* (A) WT revertant animals. This strain is a Ser revertant: the worms appear perfectly WT as far as body size and shape is concerned. (B) Non-*ts* Dpy revertants (*e2770/e2770; sc8/+*) grown at 25°. The strain is viable at restrictive temperature. Bars: A, 0.1 mm; B, 1 mm.

terize the *dpy-31* gene function by genetic means. The fully penetrant *ts* lethality of *dpy-31* mutants lent itself to analysis of genetic interaction via isolation of extragenic suppressors. We first carried out an EMS pilot screen in the course of which five independent revertant strains were isolated. To obtain a wide range of suppressing mutations, most subsequent screens were performed using the mutagen ENU, which generates a wider spectrum of mutations than EMS (DE STASIO *et al.* 1997).

Nine mutagenesis screens yielded >50 independent revertants of *e2770.* Such revertants have different properties but can be grouped into two main classes: WT revertants, *i.e.,* phenotypically WT worms, and non-*ts* Dpy revertants, which are Dpy but no longer lethal at restrictive temperatures.

**The WT revertants carry intragenic mutations restoring the *dpy-31* function:** A total of 33 WT revertants were
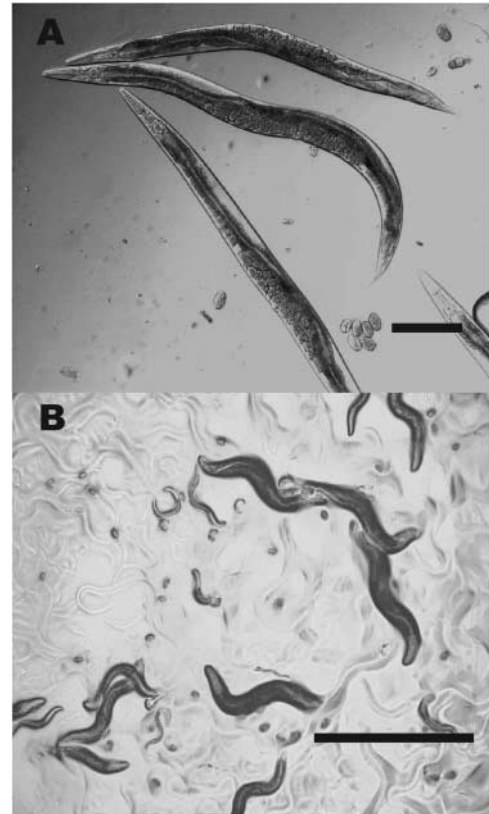
isolated: 19 strains were recovered in EMS screens and 14 in ENU screens. The WT revertants had almost identical properties, resembling normal worms with respect to body size (Figure 5A). All revertants complemented the *e2770* allele and the deficiency *nDf16,* indicating that the suppressor mutations were *cis* dominant and tightly linked to *e2770* and therefore presumably intragenic.

Intragenic reversion events were confirmed by sequence analysis of these alleles. Five WT revertants carried an exact reversion at the codon changed by *e2770,* restoring the WT sequence of DPY-31 (P469L; Figure 3). Five other WT revertants carried a second-site missense mutation on the same codon. This mutation would cause a Pro-to-Ser substitution in the encoded protein (P469S; Figure 3), suggesting that amino acids other than Leu can restore the function of the CUB domain in *dpy-31.* While the "Leu" revertants were indistinguishable from WT, the "Ser" revertants had a residual lethality (9% of the $F_1$ progeny at 20°; $n = 5$).

To rapidly analyze the remaining 23 WT revertants, we used ARMS-PCR (NEWTON *et al.* 1989). This revealed that all the WT revertants carried either a codon for Leu (8/19 EMS; 10/14 ENU) or a codon for Ser (11/

19 EMS; 4/14 ENU) at the codon changed by *e2770*. ENU might be expected to generate additional single-base changes to codons for Ala, Thr, Arg, or His at position 469, but these were not recovered as revertants nor were any compensatory mutations located elsewhere in the coding sequence.

**The non-*ts* Dpy revertants carry extragenic mutations that suppress *dpy-31* lethality:** Twenty-three non-*ts* Dpy revertants of *e2770* were isolated, 6 from EMS screens and 17 from ENU screens. All non-*ts* Dpy revertants are Dpy at all temperatures and viable at 25° (Figure 5B). Adults of these strains are Dpy but often larger than *e2770* viable worms (Figure 5B) while early larvae are markedly less Dpy than *e2770* and sometimes close to WT. Genetic analysis showed that all the non-*ts* Dpy revertants contain the original *e2770* mutation as well as a dominant extragenic suppressor of the lethality.

When separated from *dpy-31*, six of the dominant suppressors (*e2809*, *e2888*, *e2889*, *e2890*, *e2896*, and *e2901*) were found to be associated with a recessive LRol phenotype. Further analysis showed that the six mutations were allelic and linked to LGV. In complementation tests against known genes mapping to LGV, the alleles failed to complement *rol-4*, which also displays a LRol mutant phenotype. Finally, we found that the standard *rol-4* allele *sc8* could also act as a dominant suppressor of the *dpy-31* lethality. This showed that these suppressors are allelic to *rol-4*.

The genetic interaction between *rol-4* and *dpy-31* suggested that this gene was involved at some level in *dpy-31* function. *rol-4* was defined >20 years ago (Cox *et al.* 1980), but its molecular identity has not been hitherto determined. Since seven of eight genes associated with a LRol mutant phenotype have been found to encode collagens, it seemed probable that *rol-4* would encode a collagen target of *dpy-31* or, alternatively, a collagen-modifying enzyme functionally related to *dpy-31*.

**The extragenic suppressor *rol-4* defines a group of mutations affecting the C-terminal region of the SQT-3 collagen:** *rol-4* had been previously mapped to position 4.30 on LGV, but contradictory data have been reported as to whether *rol-4* is uncovered by the deficiency *ctDf1*, spanning map positions 3.4–4.1 (http://www.wormbase.org). We determined that *rol-4* is contained within *ctDf1* but we found that *sc8/ctDf1* worms roll less vigorously than *sc8* homozygotes, suggesting that *ctDf1* contains a hypomorphic allele of *rol-4*.

To clone *rol-4*, a candidate gene approach was taken. We searched the genomic sequences covered by or in proximity to *ctDf1* for genes encoding collagen or modiying enzymes and selected candidates on five cosmid clones. Clone F23H12 rescued the Rol phenotype of *rol-4* worms when carried on an extrachromosomal array (data not shown). F23H12 contains the genomic sequence of the collagen gene *sqt-3*, which has been previously characterized and defined mutationally (Kramer *et al.* 1985; van der Keyl *et al.* 1994). Two lines of

evidence indicated that *sqt-3* and *rol-4* are the same gene: (i) the mutant phenotypes of the two genes are similar, as *sqt-3* mutants display a LRol phenotype and tail abnormalities and both defects are also found in *rol-4* worms and (ii) *sqt-3* alleles exhibit aberrant complementation patterns with *rol-4(sc8)*, which have been interpreted as cases of intergenic noncomplementation (Kusch and Edgar 1986). Our results instead suggest that the complex complementation arises from intragenic interactions.

We then tested the remaining extragenic suppressors for allelism to *sqt-3*. Two other dominant suppressors of *dpy-31* lethality, *e2906* and *e2911*, failed to complement *sqt-3/rol-4* alleles. *e2911* worms display a weak Dpy phenotype but seem otherwise WT; *e2906* homozygotes are also weakly Dpy. In addition, this strain exhibits a strong cold sensitive (*cs*) lethality; *i.e.*, *e2906* worms are viable at 25°, but embryonic or early larval lethal at 15°.

Sequence analysis of the *sqt-3* gene isolated from the nine allelic suppressor strains (*sc8*, *e2809*, *e2888*, *e2889*, *e2890*, *e2896*, *e2901*, *e2906*, and *e2911*) revealed that each of them carried a missense mutation in the last exon of *sqt-3*. In all, five different mutations were obtained, because *sc8*, *e2888*, and *e2890* as well as *e2809*, *e2889*, and *e2896* were found to represent independent isolations of the same events (Figure 6). It should be noted that the use of ENU as a mutagen was instrumental in generating such a diverse range of mutational events. All alleles generated with EMS carry, as expected, GC-to-AT transitions (*sc8*, *e2888*: G to A; *e2809*: C to T). In contrast, three ENU-induced mutations are GC-to-AT transitions (*e2889*, *e2890*, and *e2896*), while *e2901*, *e2906*, and *e2911* are A-to-G transitions.

The five suppressor mutations cause five different amino acid substitutions in the encoded protein. The SQT-3 predicted protein contains a signal peptide (Figure 6; SP: aa 1–37), a N-terminal propeptide (N-PRO: aa 37–105), Gly-X-Y interrupted repeats (GLY-X-Y: aa 105–284), and a C-terminal domain (C-PRO: aa 285–301). The five substitutions map between amino acids 288 and 299 of the SQT-3 C terminus. Significantly, the Tyr and Cys residues occupying positions 286–287 (Figure 6) are likely to be required for tyrosine crosslinking of SQT-3 (Yang and Kramer 1994, 1999).

The nature of the suppressor mutations suggested that structural alterations in the SQT-3 C terminus could cause a partial restoration or replacement of the *dpy-31* function. The most plausible scenario is that DPY-31 is responsible for cleavage of the SQT-3 C terminus. However, an alternative explanation is that the dominant suppression of the *dpy-31* lethality could be a consequence of decreased SQT-3 levels. To discriminate between these two possibilities, we tested mutant alleles of *sqt-3* for suppression of the *dpy-31* lethality. All four mutations defining *sqt-3* are loss of function, but no true genetic null has been hitherto isolated. *sc63*, *e24*, and *e2117* have missense mutations in the triple helical do-
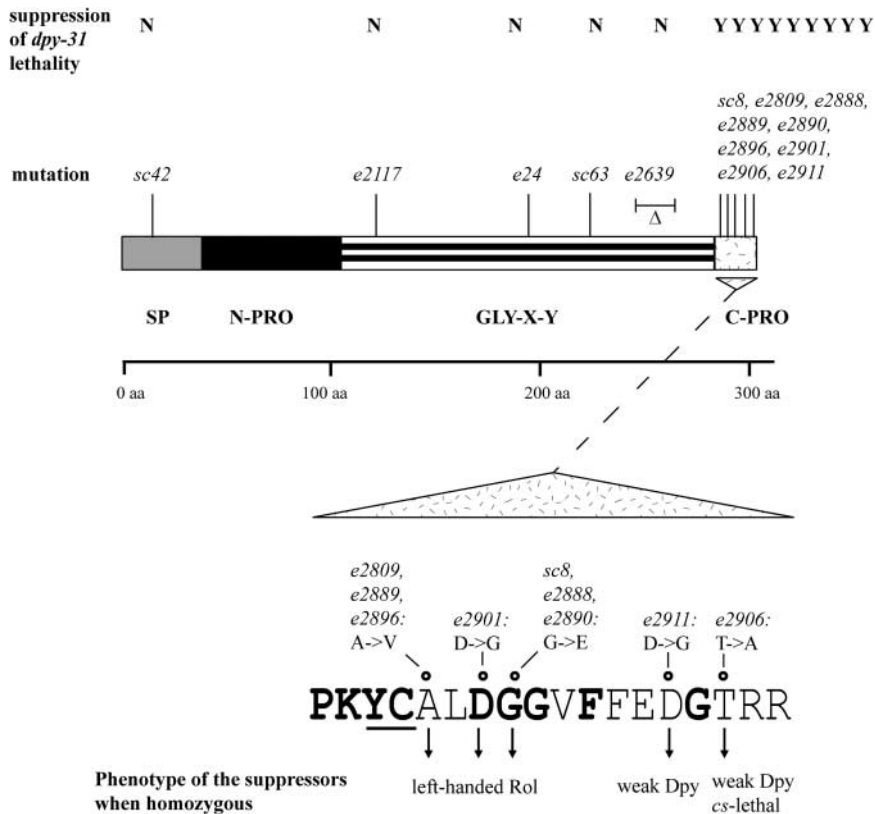
FIGURE 6.—Structure of SQT-3. For clarity, interruptions in the Gly-X-Y repeats are not shown in the diagram. Locations of the molecular lesions corresponding to 14 *sqt-3* alleles are indicated along with their ability to suppress the *dpy-31* lethality. The sequence of the SQT-3 C terminus, the molecular lesions identified in the *sqt-3(sup)* alleles, and the phenotype associated with each suppressor are shown below the protein structure. Residues in boldface type are conserved in all collagens of the SQT-3 subfamily. The YC motif (underlined) may be required for crosslinking of SQT-3.

main (VAN DER KEYL *et al.* 1994). In the *sqt-3* hypomorphic allele *sc42* (COX *et al.* 1980), we identified a missense mutation causing a Gly-to-Asp substitution in the signal peptide of SQT-3 (Figure 6). Moreover, we have recovered a deletion allele of *sqt-3*, *e2639*. This allele contains a 60-bp deletion mapping to the last block of Gly-X-Y repeats (Figure 6). In *2639*, the correct frame is retained and an "in-frame" Gly-X-Y repeat is recreated by rejoining. *e2639* worms display a pronounced Sqt phenotype [*i.e.*, dominant Rol, recessive Dpy, like the canonical *sqt-3* allele *sc63* (COX *et al.* 1980)]. In addition, *e2639* homozygotes are strongly *ts*-lethal. The dominant negative phenotype and the nature of the mutation suggest that *e2639* is not a null allele of *sqt-3*, which would be expected to behave recessively. None of the 5 *sqt-3* alleles was found to alleviate the *dpy-31* lethality in either heterozygous or homozygous combination (Figure 6). No dominant suppression of *dpy-31* lethality was conferred by the deficiency *arDf1*, which entirely deletes *sqt-3* (http://www.wormbase.org). These results demonstrate that partial suppression of the Dpy-31 phenotype by *sqt-3* is specifically associated with this newly identified subset of C-terminal missense mutations. For clarity, in the rest of the article this class of *sqt-3* mutations will be collectively referred to as *sqt-3(sup)*.

**Most *sqt-3* suppressors are exclusively dominant:** In subsequent tests, we found that all four mutant alleles of *dpy-31* are equally suppressible by *sqt-3(sup)*. Since the *e2770*, *e2920*, and *ju345* mutations affect different

domains of DPY-31, and *e2919* is a probable null allele, it seems unlikely that the suppressors could be acting by restoring function to DPY-31 by protein-protein interaction. Rather, the *sqt-3(sup)* mutations might permit action by an alternative processing enzyme, which can substitute for the defective or missing DPY-31. Alternatively, a less likely possibility is that the *sqt-3(sup)* mutations might allow SQT-3 to function in the absence of C-terminal cleavage.

Further analysis revealed that the mechanism of suppression has unusual properties. As stated above, all the *sqt-3(sup)* are dominant; therefore a worm of genotype *dpy-31/dpy-31; sqt-3(sup)/+* is viable at 25°. One might expect that such animals would segregate self-progeny according to the ratio of three viable Dpy to one lethal Dpy. However, for most of the suppressors a 1:1 ratio was observed. These observations suggested that double homozygous worms for *dpy-31* and *sqt-3(sup)* were not viable.

To test the effect of heterozygosity *vs.* homozygosity of the suppressors, we scored the self-progeny of *unc-32/ dpy-31; sqt-3(sup)/+* worms raised at restrictive temperature. We found that all but one of the suppressors of the *dpy-31* lethality act as such only in heterozygous combination with a WT copy of *sqt-3* (Table 1). Therefore, worms of genotype *dpy-31/dpy-31; sqt-3(sup)/+* are fully viable at 25°, whereas double homozygotes for *dpy-31* and *sqt-3(sup)* are invariably lethal at restrictive temperature. The only exception is *e2906*. Double homozygotes for this suppres-

TABLE 1

**Effect of the *sqt-3(sup)* gene dosage on the viability of *dpy-31* mutants**

| *sqt-3(sup)* allele | Genotype of the parental and broods counted | Viable phenotypes | Viable Dpy expected if the *sqt-3(sup)* acts exclusively when heterozygous | Viable Dpy expected if the *sqt-3(sup)* acts both when heterozygous and when homozygous |
|---|---|---|---|---|
| None | *e2770/e189* +/+ (5) | 586 WT 283 Unc 0 Dpy | 0 | 0 |
| *sc8* | *e2770/e189;* *sc8/+* (7) | 841 WT 494 Unc 264 Dpy 319 Rol 109 Rol Unc | 280 | 420 |
| *e2889* | *e2770/e189;* *e2889/+* (5) | 553 WT 301 Unc 181 Dpy 166 Rol 54 Rol Unc | 184 | 265 |
| *e2901* | *e2770/e189;* *e2901/+* (5) | 465 WT 247 Unc 100 Dpy 140 Rol 63 Rol Unc | 155 | 232 |
| *e2906* | *e2770/e189;* *e2906/+* (8) | 1183 WT 608 Unc 433 Dpy | 295 | 443 |
| *e2911* | *e2770/e189;* *e2911/+* (5) | 812 WT 406 Unc 136 Dpy | 203 | 304 |

Note that *sc8*, *e2889*, and *e2901* are associated with a recessive Rol phenotype; therefore, worms homozygous for these suppressors could be scored. *e2906* and *e2911* mutants cannot be easily distinguished from WT; therefore, in this case worms homozygous for the suppressors were scored as WT. For suppression in both heterozygous and homozygous combinations we expected a viable Dpy frequency in the $F_1$ of $\sim$3/16. For suppression exclusively in heterozygous combination, we expected a viable Dpy frequency of $\sim$1/8. The expected figure of viable Dpy was calculated on the basis of the WT scores. The number of observed and expected viable Dpy is underlined. The number of broods counted is in parentheses.

sor and *dpy-31* do not display any sign of lethality at 25° and produce fully viable progeny.

Isolation of double homozygotes for the *sc8* suppressor and *dpy-31* confirmed that *e2770/e2770; sc8/sc8* worms are not suppressed. Moreover, we found that this strain is even less viable than *dpy-31* mutants devoid of suppressor. At permissive temperature, *e2770* homozygotes produce an average of 16 viable self-progeny ($n = 7$) and *e2770/e2770; sc8/+* worms produce an average of 80 ($n = 7$). In contrast, worms of genotype *e2770/e2770; sc8/sc8* produce on average only 3 viable self-progeny ($n = 7$). Enhancement is also seen in homozygotes with the hypomorphic *dpy-31* allele *ju345*. While *ju345* mutants are usually viable at 25°, *ju345* worms carrying the *sc8* suppressor in homozygous combination are inviable at 25° and display very low viability at 15°. Thus, while presence of a *sqt-3(sup) in trans* to a WT copy of the gene leads to a significant alleviation of the

*dpy-31* lethality, homozygosity of the suppressors has deleterious effects on the viability of *dpy-31* mutants.

**Other extragenic suppressors of the *dpy-31* lethality:** Overall, about half of the extragenic suppressors of *dpy-31* were found to be allelic to *sqt-3*. Preliminary mapping of the remaining suppressors demonstrates that they define several distinct loci. We expect that cloning of the other extragenic suppressors will further our understanding of the collagen-processing pathway.

DISCUSSION

We have shown that the *dpy-31* locus of *C. elegans*, which encodes a predicted astacin zinc-metalloprotease, is required for viability and normal morphogenesis. *dpy-31* plays a crucial role in the late stages of embryogenesis following elongation. Loss of DPY-31 causes the embryonic exoskeleton to collapse to a length com-

parable to that of the twofold stage. The marginal viability of *dpy-31* mutants observed at 15° may result either from the activity of a redundant astacin protease or from residual DPY-31 activity, which is sufficient to permit formation of a functional, albeit highly defective, cuticle. Alternatively, the targets of *dpy-31* may be capable of limited function in the absence of cleavage. While the lethality of *dpy-31* mutants indicates that the gene is essential at all temperatures above 15°, the recovery of a *ts* allele has allowed a reversion analysis of the locus. The mutant phenotype and sequence similarities of DPY-31, together with the reversion analysis discussed below, indicate that it acts as an essential pCP for one or more cuticle collagens. Consistent with this proposed role, our expression analysis shows that DPY-31 is widely expressed in the hypodermal cells, which are responsible for cuticle secretion. Our results do not exclude the formal possibility that DPY-31 acts only as a maturase for another protease, which in turn acts as the direct collagen-processing enzyme. However, the structure of DPY-31 and its similarity to a known pCP enzyme favor the simpler model.

In a recent survey, it was determined that the *C. elegans* genome contains as many as 39 actively transcribed astacin protease genes. This is in striking contrast with what has been observed for other metazoan genomes, in which only three to four astacin paralogs can be generally found (Möhrlen *et al.* 2003). The functional significance of the 10-fold expansion undergone by astacins in nematodes is unclear. Since RNAi screens in *C. elegans* have failed to highlight astacin protease genes essential for viability, this class of genes has been hitherto overlooked. Only one astacin gene so far has been defined mutationally. The locus was named *hch-1* (for *hatch* defective). Loss of function of *hch-1* results in a delayed hatching phenotype, indicating that HCH-1 is responsible for degradation of eggshell proteins (Hishida *et al.* 1996). The inefficient and poorly penetrant phenocopy of the *dpy-31* mutant phenotype obtained by RNAi suggests that no conclusion can be drawn from genome-wide RNAi screens on the actual number of astacin protease genes required for viability, and therefore no inference can be made about their actual level of redundancy. In the light of our findings, it seems possible that the expansion of astacin protease genes in the nematode *C. elegans* partially reflects subfunctionalization or specialization in cuticle collagen processing and degradation.

We have investigated the gene function by generating revertants of *dpy-31(e2770)*. Analysis of WT revertants of the mutation showed that full reversion can be obtained only through direct restoration of the *dpy-31* function. In our screens, we have isolated exclusively Leu and Ser WT revertants of the *e2770* mutation in the CUB domain. Failure to recover any other ENU-induced mutations at the codon changed by *e2770* suggests that tight molecular constraints exist on position 469 of
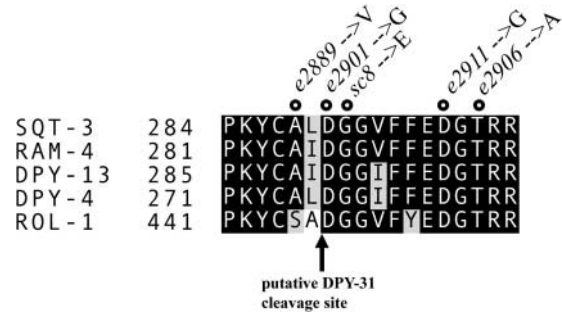


FIGURE 7.—ClustalW alignment of the C-terminal domains of five *C. elegans* collagens belonging to the SQT-3 subfamily. The protein sequences correspond, respectively, to genes F23H12.4 (*sqt-3*), F36A4.10 (*ram-4*), F30B5.1 (*dpy-13*), Y41E3.2 (*dpy-4*), and Y57A10A.11 (*rol-1*). Circles indicate the positions affected by the *sqt-3(sup)* mutations.

DPY-31. Indeed, Ser revertants, albeit WT in size, are partially lethal, showing that even this substitution causes some reduction of function of *dpy-31*. The CUB domain may be essential for catalytic activity or for secretion of the protein, as shown for BMP-1 (Hartigan *et al.* 2003).

Analysis of extragenic suppressors of *dpy-31* has revealed a peculiar genetic interaction between *dpy-31* and the collagen gene *sqt-3*. Five clustered missense mutations, which are associated with recessive hypomorphic phenotypes and alter the structure of the SQT-3 C terminus, possess the property of acting as dominant suppressors of *dpy-31* lethality.

*sqt-3* is a unique cuticle collagen in that it is the only member of this gene family that is indispensable for viability. *sqt-3* is most abundantly expressed in embryos (Kramer *et al.* 1985), where it exerts a crucial structural function in maintenance of the body shape achieved through elongation (Priess and Hirsh 1986). Strong loss-of-function mutations in *sqt-3* result in embryonic lethality and in early larval arrest characterized by the same retracted Dpy phenotype as *dpy-31* mutants (van der Keyl *et al.* 1994), supporting the hypothesis that DPY-31 may be required for assembly of SQT-3. SQT-3-like collagens are evolutionarily conserved, as they have been detected in a vast array of distantly related nematode species (Cox *et al.* 1990) and are almost certainly ubiquitous in the phylum. The fundamental role played by SQT-3 in cuticle structure is also probably conserved: in the parasitic nematode species *Strongyloides stercoralis*, the transcript of the *sqt-3* ortholog gene is found among the 12 most abundant in L1 larvae (Mitreva *et al.* 2004).

In the *C. elegans* genome, we find 49 SQT-3 paralog genes, according to the classification based on the structural homology proposed by Johnstone (2000). In Figure 7, we show an amino acid alignment of the C-terminal regions of the five collagens of the SQT-3 subfamily that are associated with mutant phenotypes, in addition

to SQT-3: DPY-13 (von Mende *et al.* 1988), ROL-1 (L. Molin, personal communication), RAM-4 (R. Y. Yu and K. L. Chow, personal communication), and DPY-4 (Simmer *et al.* 2003). While the average sequence identity of these proteins is ∼60%, the C-terminal regions are almost identical, with only a few conservative substitutions. Also, in the remaining members of the SQT-3 subfamily, the C-terminal sequence appears invariably conserved (minimal amino acid identity to SQT-3: 78%; absolutely conserved residues are highlighted by bold-face type in Figure 6).

Amino acid conservation implies an important role for the C-terminal regions of these collagens. In vertebrates, the C-propeptides of fibrillar procollagens have been shown to perform three distinct functions. First, during assembly, a conserved motif of the collagen C-propeptide is required to bring the three monomers together (Bachinger *et al.* 1980). Second, the C-propeptides of vertebrate procollagen also contain variable sequences, which promote selective oligomerization of structurally compatible monomers (Lees *et al.* 1997). Third, the C-propeptide harbors a consensus sequence for the C-terminal proteinase: upon secretion, cleavage of the C-propeptide occurs at this site, leaving only a short nonhelical region, termed telopeptide, C-terminal to the triple helical domain (Leung *et al.* 1979).

By analogy, the C terminus of collagens of the SQT-3 subfamily could have similar functions. However, given the absence of variable sequences, it seems unlikely that this motif would be able to direct type-specific oligomerization of collagen chains. Presumably, in nematode collagens determinants of chain assembly reside elsewhere. Analysis of the molecular lesions found in the 5 *sqt-3(sup)* alleles suggests that the two other functions of the C-propeptide are probably conserved in members of the SQT-3 subfamily.

The *sqt-3(sup)* mutations can be further divided into two groups. The first includes *e2889*, *e2901*, and *sc8*, which display very similar LRol mutant phenotypes and affect adjacent amino acids in the encoded protein. The second group is represented by *e2911* and *e2906*, which are associated with non-Rol mutant phenotypes and cause the substitution of two close residues at the C-terminal end of SQT-3 (Figures 6 and 7).
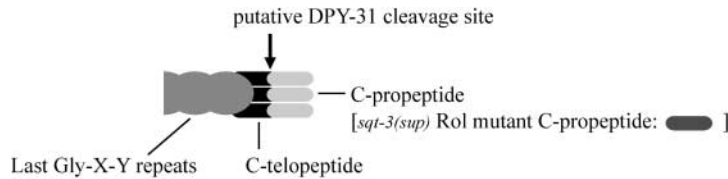
The *e2889*, *e2901*, and *sc8* mutations affect, respectively, residues 288, 290, and 291 of SQT-3. Positions 290 and 291 are strictly conserved in collagens of the SQT-3 subfamily whereas position 288 is conserved in most of the cases (Figures 6 and 7). Two lines of evidence support the hypothesis that the YCALD motif between aa 286 and 290 may represent a site for cleavage by DPY-31. First, this motif closely resembles the Y(X)XXD degenerate consensus site for BMP-1 found in various collagen precursors and in chordin (Scott *et al.* 2000 and references cited therein). Second, use of this site for cleavage is consistent with the fact that SQT-3 C-terminal processing should occur downstream of the YC motif (aa

286–287), which is known to be required for tyrosine crosslinking and therefore likely to be retained in the mature trimers. The Asp residue occupying position 290 may be the actual site of cleavage (YCAL/D). Almost all proven targets of BMP-1 are processed at an Asp residue, which was shown to be indispensable for BMP-1 action (Lee *et al.* 1990).

Likewise, the *e2911* and *e2906* mutations could act to cause indirect alteration or distortion of the DPY-31 cleavage site. However, the nature of the *e2906* mutation suggests that it may have a pleiotropic effect. *e2906* homozygotes are *cs* lethal: the strain is inviable at 15°, while WT-like, or weakly Dpy, at higher temperatures. At 15°, *e2906* arrested embryos display the retracted Dpy phenotype typical of *sqt-3* loss of function. This phenotype is particularly puzzling because it represents the mirror image of what is observed in standard loss-of-function *sqt-3* mutants, such as *e24*, *e2117*, and e*2639*, which are viable only at low temperatures. At the protein level, *e2906* causes a Thr-to-Ala substitution at position 299 of SQT-3. Given that this is the C-terminal-most nucleophilic residue, it seems possible that it may be involved in the initial steps of collagen oligomerization. According to this model, the amino acid substitution could decrease the rate of chain assembly, which would not occur efficiently at low temperatures.

The recovery of five distinct but clustered missense suppressor mutations in *sqt-3* might suggest a direct compensatory interaction between SQT-3 and DPY-31. However, since the *sqt-3(sup)* mutations suppress equally the four *dpy-31* mutant alleles, including *e2919*, which represents a putative molecular null, it is more likely that suppression involves an alternative pCP that acquires the ability to process SQT-3 as a consequence of conformational changes in the collagen C-propeptide. Possibly, this alternative pathway will be defined by one or more of the remaining extragenic suppressors of *dpy-31*.

Finally, the behavior of the *sqt-3(sup)* alleles requires explanation. Most *sqt-3(sup)* alleles were found to act as suppressors of the *dpy-31* lethality only *in trans* to a WT copy of *sqt-3*, while the homozygous combination of the suppressors worsens the Dpy-31 mutant phenotype. This unusual genetic interaction may result from the trimeric nature of functional *sqt-3* gene products. We propose a model to explain suppression by the *sqt-3(sup)* Rol alleles, *i.e.*, *e2889*, *sc8*, and *e2901*. Our hypothesis is based on three assumptions: (i) SQT-3 forms exclusively homotrimers; (ii) the *sqt-3(sup)* mutations allow processing of SQT-3 by the alternative pCP, but DPY-31 retains the ability of cleaving the mutant collagen chains; (iii) cleavage by the alternative pCP interferes with crosslinking of SQT-3 trimers. This could happen in various ways; a reasonable interpretation (illustrated in Figure 8) could be that cleavage of SQT-3 by the alternative protease occurs upstream of the DPY-31 processing site, thus causing loss or inactivation of the YC motif required for crosslinking of SQT-3. A similar type of unspecific

putative DPY-31 cleavage site

C-propeptide

[*sqt-3(sup)* Rol mutant C-propeptide: ]

Last Gly-X-Y repeats

C-telopeptide

| Genotype and phenotype | Predominant types of SQT-3 trimers formed | pCP cleaving | Resulting trimers |
|---|---|---|---|
| *WT* <br> WT | | DPY-31 | |
| *sqt-3(sup)/ sqt-3(sup)* <br> Rol | | DPY-31 and alternative | |
| *sqt-3(sup)/+* <br> WT | | DPY-31 and alternative | |
| *dpy-31/dpy-31* <br> lethal Dpy | | None | |
| *dpy-31/dpy-31 sqt-3(sup)/+* <br> viable Dpy | | Alternative | |
| *dpy-31/dpy-31 sqt-3(sup)/ sqt-3(sup)* <br> lethal Dpy | | Alternative | |

FIGURE 8.—A model proposing an explanation for the exclusively dominant rescue of the *dpy-31* lethality conferred by the *sqt-3(sup)* Rol alleles. The shaded segmented bar represents a SQT-3 trimer. For clarity, only part of the Gly-X-Y repeats, the C-propeptides and C-telopeptides, are shown in the diagram. The C-telopeptides, which contain the sequence required for crosslinking of trimers, are represented as solid. The C-propeptides, which are predicted to be removed by proteolytic cleavage, are shown with light shading in the case of WT monomers and with dark shading in the case of *sqt-3* Rol suppressor mutant monomers.

processing of the C-telopeptide occurs in vertebrate collagens in the absence of cleavage by BMP-1 (BATEMAN *et al.* 1987; LEE *et al.* 1990). This model has the advantage of explaining the recessiveness of the suppressors' LRol phenotype as well as the exclusively dominant suppression of *dpy-31* lethality. On one hand, in a *dpy-31(+)* background, the Rol phenotype of the *sqt-3(sup)* may be recessive because in a heterozygous combination, SQT-3 heterotrimers (composed of a mixture of mutant and WT chains) and homotrimers (composed of purely mutant or purely WT chains) would form. Cleavage by DPY-31 of WT homotrimers and WT chains in the heterotrimers could then be sufficient to ensure normal levels of SQT-3 crosslinking, as a large fraction of collagen trimers would possess C-telopeptides. In *sqt-3(sup)* homozygotes, the alternative protease would compete with DPY-31 for the cleavage of the SQT-3 mutant homotrimers, leading to the partial loss of telopeptide sequences

and, consequently, to a reduction in crosslinking of SQT-3. This hypothesis is consistent with findings by YANG and KRAMER (1994, 1999), who demonstrated that mutations affecting the C-terminal YC crosslinking motif of the collagen SQT-1 and ROL-6 are associated with a recessive LRol phenotype identical to that displayed by the *sqt-3(sup)* Rol alleles. On the other hand, in a *dpy-31* mutant background, the LRol *sqt-3(sup)* alleles may suppress the *dpy-31* lethality because, in the absence of DPY-31, imprecise cleavage of the SQT-3 C terminus by the alternative pCP could allow more efficient packing of the collagen trimers. However, a WT copy of *sqt-3* would be required to ensure formation of heterotrimers containing at least one YC crosslinking motif. The deleterious effects observed when the suppressors are homozygous would then be a consequence of the total loss of SQT-3 C-terminal crosslinking sites, which would preclude SQT-3 deposition. This interpretation provides a

plausible explanation of the suppression by *sqt-3(sup)* alleles. However, the existence of many additional suppressor mutations suggests that there are other ways to ameliorate the lethal effects of the *dpy-31* mutation. Future analysis of these other suppressors, which define at least two distinct loci unlinked to *sqt-3*, will shed further light on collagen processing and astacin proteases in *C. elegans.*

## LITERATURE CITED

AHMED, S., and J. HODGKIN, 2000 MRT-2 checkpoint protein is required for germline immortality and telomere replication in *C. elegans*. Nature **403:** 159–164.

BACHINGER, H. P., P. BRUCKNER, R. TIMPL, D. J. PROCKOP and J. ENGEL, 1980 Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen. Role of disulfide bridges and peptide bond isomerization. Eur. J. Biochem. **106:** 619–632.

BATEMAN, J. F., J. J. PILLOW, T. MASCARA, S. MEDVEDEC, J. RAMSHAW *et al.*, 1987 Cell-layer-associated proteolytic cleavage of the telopeptides of type I collagen in fibroblast culture. Biochem. J. **245:** 677–682.

BOND, J. S., and R. J. BEYNON, 1995 The astacin family of metalloendopeptidases. Protein Sci. **4:** 1247–1261.

BRENNER, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71–94.

COX, G. N., J. S. LAUFER, M. KUSCH and R. S. EDGAR, 1980 Genetic and phenotypic characterization of roller mutants of *Caenorhabditis elegans*. Genetics **95:** 317–339.

COX, G. N., M. KUSCH and R. S. EDGAR, 1981 Cuticle of *Caenorhabditis elegans*: its isolation and partial characterization. J. Cell Biol. **90:** 7–17.

COX, G. N., L. M. SHAMANSKY and R. J. BOISVENUE, 1990 Evidence that the 3A3 collagen gene is a member of an evolutionarily conserved family of nematode cuticle collagens. Exp. Parasitol. **70:** 175–185.

DE STASIO, E. A., and S. DORMAN, 2001 Optimization of ENU mutagenesis of *Caenorhabditis elegans*. Mutat. Res. **495:** 81–88.

DE STASIO, E. A., C. LEPHOTO, L. AZUMA, C. HOLST, D. STANISLAUS *et al.*, 1997 Characterization of revertants of *unc-93(e1500)* in *Caenorhabditis elegans* induced by *N*-ethyl-*N*-nitrosourea. Genetics **147:** 597–608.

EDENS, W. A., L. SHARLING, G. CHENG, R. SHAPIRA, J. M. KINKADE *et al.*, 2001 Tyrosine cross-linking of extracellular matrix is catalyzed by Duox, a multidomain oxidase/peroxidase with homology to the phagocyte oxidase subunit gp91phox. J. Cell Biol. **154:** 879–891.

FIRE, A., S. XU, M. K. MONTGOMERY, S. A. KOSTAS, S. E. DRIVER *et al.*, 1998 Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature **391:** 744–745.

HARTIGAN, N., L. GARRIGUE-ANTAR and K. E. KADLER, 2003 Bone morphogenetic protein-1 (BMP-1). Identification of the minimal domain structure for procollagen C-proteinase activity. J. Biol. Chem. **278:** 18045–18049.

HISHIDA, R., T. ISHIHARA, K. KONDO and I. KATSURA, 1996 *hch-1*, a gene required for normal hatching and normal migration of a neuroblast in *C. elegans*, encodes a protein related to TOLLOID and BMP-1. EMBO J. **15:** 4111–4122.

IMAMURA, Y., B. M. STEIGLITZ and D. S. GREENSPAN, 1998 Bone morphogenetic protein-1 processes the NH2-terminal propeptide, and a furin-like proprotein convertase processes the COOH-terminal propeptide of pro-α1(V) collagen. J. Biol. Chem. **273:** 27511–27517.

JOHNSTONE, I. L., 2000 Cuticle collagen genes. Expression in *Caenorhabditis elegans*. Trends Genet. **16:** 21–27.

KESSLER, E., K. TAKAHARA, L. BINIAMINOV, M. BRUSEL and D. S. GREENSPAN, 1996 Bone morphogenetic protein-1: the type I procollagen C-proteinase. Science **271:** 360–362.

KO, F. C., and K. L. CHOW, 2002 A novel thioredoxin-like protein encoded by the *C. elegans dpy-11* gene is required for body and sensory organ morphogenesis. Development **129:** 1185–1194.

KRAMER, J. M., G. N. COX and D. HIRSH, 1985 Expression of the *Caenorhabditis elegans* collagen genes *col-1* and *col-2* is developmentally regulated. J. Biol. Chem. **260:** 1945–1951.

KUSCH, M., and R. S. EDGAR, 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.

LEE, S. T., E. KESSLER and D. S. GREENSPAN, 1990 Analysis of site-directed mutations in human pro-α2(I) collagen which block cleavage by the C-proteinase. J. Biol. Chem. **265:** 21992–21996.

LEES, J. F., M. TASAB and N. J. BULLEID, 1997 Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. EMBO J. **16:** 908–916.

LEUNG, M. K., L. I. FESSLER, D. B. GREENBERG and J. H. FESSLER, 1979 Separate amino and carboxyl procollagen peptidases in chick embryo tendon. J. Biol. Chem. **254:** 224–232.

MADURO, M., and D. PILGRIM, 1995 Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. Genetics **141:** 977–988.

MARQUÉS, G., M. MUSACCHIO, M. J. SHIMELL, K. WUNNENBERG-STAPLETON, K. W. CHO *et al.*, 1997 Production of a DPP activity gradient in the early *Drosophila* embryo through the opposing actions of the SOG and TLD proteins. Cell **91:** 417–426.

MELLO, C., and A. FIRE, 1995 DNA transformation. Methods Cell Biol. **48:** 451–482.

MITREVA, M., J. P. MCCARTER, J. MARTIN, M. DANTE, T. WYLIE *et al.*, 2004 Comparative genomics of gene expression in the parasitic and free-living nematodes *Strongyloides stercoralis* and *Caenorhabditis elegans*. Genome Res. **14:** 209–220.

MÖHRLEN, F., H. HUTTER and R. ZWILLING, 2003 The astacin protein family in *Caenorhabditis elegans*. Eur. J. Biochem. **270:** 4909–4920.

MYLLYHARJU, J., and K. I. KIVIRIKKO, 2004 Collagen, modifying enzymes and their mutations in humans, flies and worms. Trends Genet. **20:** 33–43.

NEWTON, C. R., A. GRAHAM, L. E. HEPTINSTALL, S. J. POWELL, C. SUMMERS *et al.*, 1989 Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. **17:** 2503–2516.

PAPPANO, W. N., B. M. STEIGLITZ, I. C. SCOTT, D. R. KEENE and D. S. GREENSPAN, 2003 Use of *Bmp1/Tll1* doubly homozygous null mice and proteomics to identify and validate *in vivo* substrates of bone morphogenetic protein 1/tolloid-like metalloproteinases. Mol. Cell. Biol. **23:** 4428–4438.

PRIESS, J. R., and D. I. HIRSH, 1986 *Caenorhabditis elegans* morphogenesis: the role of the cytoskeleton in elongation of the embryo. Dev. Biol. **117:** 156–173.

ROMÃO, M. J., I. KOLLN, J. M. DIAS, A. L. CARVALHO, A. ROMERO *et al.*, 1997 Crystal structure of acidic seminal fluid protein (aSFP) at 1.9 Å resolution: a bovine polypeptide of the spermadhesin family. J. Mol. Biol. **274:** 650–660.

SCOTT, I. C., Y. IMAMURA, W. N. PAPPANO, J. M. TROEDEL, A. D. RECKLIES *et al.*, 2000 Bone morphogenetic protein-1 processes probiglycan. J. Biol. Chem. **275:** 30504–30511.

SHAW, L. M., and B. R. OLSEN, 1991 FACIT collagens: diverse molecular bridges in extracellular matrices. Trends Biochem. Sci. **16:** 191–194.

SIMMER, F., M. TIJSTERMAN, S. PARRISH, S. KOUSHIKA, M. NONET *et al.*, 2002 Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. Curr. Biol. **12:** 1317–1319.

SIMMER, F., C. MOORMAN, A. M. VAN DER LINDEN, E. KUIJK, P. V. VAN DEN BERGHE *et al.*, 2003 Genome-wide RNAi of *C. elegans* using the hypersensitive *rrf-3* strain reveals novel gene functions. PLoS Biol. **1:** E12.

SULSTON, J., and J. HODGKIN, 1987 Methods, pp. 587–606 in *The Nematode Caenorhabditis elegans*, edited by W. B. WOOD. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

THACKER, C., K. PETERS, M. SRAYKO and A. M. ROSE, 1995 The *bli-4* locus of *Caenorhabditis elegans* encodes structurally distinct kex2/subtilisin-like endoproteases essential for early development and adult morphology. Genes Dev. **9:** 956–971.

THEIN, M. C., G. McCORMACK, A. D. WINTER, I. L. JOHNSTONE, C. B. SHOEMAKER *et al.*, 2003 *Caenorhabditis elegans* exoskeleton collagen COL-19: an adult-specific marker for collagen modification and assembly, and the analysis of organismal morphology. Dev. Dyn. **226:** 523–539.

VAN DER KEYL, H., H. KIM, R. ESPEY, C. V. OKE and M. K. EDWARDS, 1994 *Caenorhabditis elegans sqt-3* mutants have mutations in the *col-1* collagen gene. Dev. Dyn. **201:** 86–94.

VON MENDE, N., D. M. BIRD, P. S. ALBERT and D. L. RIDDLE, 1988 *dpy-13*: a nematode collagen gene that affects body shape. Cell **55:** 567–576.

WICKS, S. R., R. T. YEH, W. R. GISH, R. H. WATERSTON and R. H. PLASTERK, 2001 Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. Nat. Genet. **28:** 160–164.

WINTER, A. D., and A. P. PAGE, 2000 Prolyl 4-hydroxylase is an essential procollagen-modifying enzyme required for exoskeleton formation and the maintenance of body shape in the nematode *Caenorhabditis elegans.* Mol. Cell. Biol. **20:** 4084–4093.

YANG, J., and J. M. KRAMER, 1994 In vitro mutagenesis of *Caenorhabditis elegans* cuticle collagens identifies a potential subtilisin-like protease cleavage site and demonstrates that carboxyl domain disulfide bonding is required for normal function but not assembly. Mol. Cell. Biol. **14:** 2722–2730.

YANG, J., and J. M. KRAMER, 1999 Proteolytic processing of *Caenorhabditis elegans* SQT-1 cuticle collagen is inhibited in right roller mutants whereas cross-linking is inhibited in left roller mutants. J. Biol. Chem. **274:** 32744–32749.

YIALLOUROS, I., E. GROSSE BERKHOFF and W. STOCKER, 2000 The roles of Glu93 and Tyr149 in astacin-like zinc peptidases. FEBS Lett. **484:** 224–228.

YOCHEM, J., T. GU and M. HAN, 1998 A new marker for mosaic analysis in *Caenorhabditis elegans* indicates a fusion between *hyp6* and *hyp7*, two major components of the hypodermis. Genetics **149:** 1324–1334.

Communicating editor: K. KEMPHUES