**dpy-13: A Nematode Collagen Gene That Affects Body Shape**

Nicola von Mende, David McK. Bird,* Patrice S. Albert, and Donald L. Riddle
Division of Biological Sciences
Tucker Hall
University of Missouri
Columbia, MO 65211

Summary

Mutations in the Caenorhabditis elegans dpy-13 (dumpy) gene result in a short, chunky body shape. This gene was tagged by insertion of the Tc1 transposon, and the wild-type gene was cloned by chromosomal walking 11 kb from ama-1, a cloned gene encoding the large subunit of RNA polymerase II. Three transposon insertion sites in dpy-13 are located near the 5’ end of a 1.2 kb transcribed region. The EMS-induced reference allele, dpy-13(e184), carries a small deletion near the middle of this gene. The DNA sequence reveals that dpy-13 is a member of the collagen multi-gene family, and it could encode a polypeptide of 302 amino acids. A 146 base pair sequence, encoding amino acids 56–103, is unique in the C. elegans genome, and it hybridizes to a 1 kb mRNA of moderate abundance.

Introduction

In the soil nematode Caenorhabditis elegans, more than 50 genes affecting gross morphology have been identified by mutation and genetically mapped. Mutant types include dumpy and squat (short, chunky), long, small, blistered, and roller (Brenner, 1974; Cox et al., 1980; Kusch and Edgar, 1986). The most common mutant class is dumpy, and 28 such genes have been identified. Because of the obvious visible phenotype, which usually does not interfere with growth and reproduction, many of these mutations have been extremely useful as markers for the genetic mapping of other mutations affecting development and behavior (Brenner, 1974). However, in spite of their important contribution to genetics, the functions of dumpy genes have not been determined. In some cases, the morphological mutants exhibit anatomical alterations in the extracellular cuticle (Cox et al., 1980; Ouazana et al., 1985), but genetic studies have not demonstrated which genes are directly involved in cuticle development and assembly. This question is addressed here by determining the nature of a dumpy gene product.

The C. elegans cuticle is an elastic, multi-layered, extracellular structure, the components of which are secreted by the hypodermis. It is composed primarily of small, covalently cross-linked collagens (Cox et al., 1981; Politz and Edgar, 1984), and it has been proposed that some of the morphological mutants may be defective in genes encoding these structural proteins (Kusch and Edgar, 1986). The cuticle collagens contain the characteristic triple helical structure formed by the repeated (Gly-X-Y), amino acid sequence, in which the X and Y positions are often proline or hydroxyproline (Cox et al., 1988). The gene family encoding these collagens is large, consisting of 50–100 genes for small (30 kd) proteins (Cox et al., 1984), and many of these genes are expressed at the time of cuticle synthesis (Cox and Hirsh, 1985; Kramer et al., 1985). These genes are apparently dispersed throughout the genome, and several have been positioned on the genetic map by means of linked DNA restriction site differences (Cox et al., 1985).

Owing to the overall similarity of the gene family members, it seems likely that many of the gene products might substitute for each other during cuticle assembly. This would make the individual genes difficult to detect genetically, because loss of gene function by mutation would be masked by the activity of related genes. Indeed, putative null alleles of the sqt-1 locus are wild-type in phenotype, whereas other alleles are dominant to wild-type, and interact to produce animals with roller, dumpy, or long phenotypes (Kusch and Edgar, 1986). Other squat and roller mutants exhibit unusual intergenic interactions, suggesting that they may be members of a gene family involved in the formation of the body cuticle.

The dpy-13 locus has been of particular value in the recent analysis of ama-1 (resistance to α-amanitin), which encodes the large subunit of RNA polymerase II, because of the close genetic linkage (Rogalski and Riddle, 1988; Rogalski et al., 1988; Bullerjahn and Riddle, 1988). A short chromosomal walk from the ama-1 gene, which had been cloned by means of a heterologous probe from Drosophila (Bird and Riddle, submitted), permitted the indentification of an array, or “contig,” of overlapping cosmids clones (Coulson et al., 1986) corresponding to the ama-1 region. Location of the physical breakpoint of a small deficiency (mdf4) that deletes ama-1 oriented the cosmid contig with respect to the genetic map, on which the left end of mdf4 (Bird and Riddle, submitted) was identified very close to the left of ama-1 (Rogalski and Riddle, 1988; Bullerjahn and Riddle, submitted). Transcript mapping and DNA hybridization experiments using DNA probes from the ama-1 region revealed that ama-1 was located in a cluster of collagen genes. Hence, the possibility that one of these genes might be dpy-13 was investigated.

We describe here the transposon-tagging, DNA sequence, and transcript analysis of dpy-13. This gene encodes a collagen that is closely related to two other C. elegans collagen genes, col-1 and col-2, and these three genes together constitute one of three distinct collagen superfamilies identified thus far (Cox et al., 1988). We propose that the dpy-13 collagen must have a unique function in cuticle growth because apparent loss-of-function mutations in this gene shorten body length. The presence of
a unique 146 bp coding segment in this gene, and other more subtle differences in deduced structure between the dpy-13 collagen and other collagens, suggest potential structural motifs that may differentiate the functions of these gene family members.

Results

Isolation of Transposon Insertion Mutants

Dumpy mutants are easily recognized by their morphology (Figure 1). In order to tag dumpy genes for molecular analysis, "mutator" strains with a high frequency of Tc1 transposition (Moerman et al., 1986) were screened visually for the spontaneous occurrence of dumpy mutants. Mutant stocks were generated from single true-breeding homozygous hermaphrodites, and crossed with Bristol (low frequency transposition) genetic backgrounds to produce genetically stable strains carrying fewer copies of Tc1 than the parental mutator strain. No more than one mutant was saved from any one population to ensure that each isolate was independent. Three of eighteen dumpy mutants obtained were alleles of dpy-13, as determined by genetic complementation tests. Two of these mutant alleles (m399 and m400) were from a screen of 2.3 x 10^6 animals of strain DR842, and one (m401) was found among 3.2 x 10^6 animals of strain RW7097 (Table 1).

Unusual complementation patterns have occasionally been observed among other morphological mutants (Kusch and Edgar, 1986), so genetic mapping was done to confirm the identity of m399 as a new dpy-13 allele. The results were consistent with the known position of dpy-13 (Edgley and Riddle, 1987), and they also confirmed that m399 behaved like a single recessive point mutation. The m399 allele was first shown to be linked to unc-5 on chromosome IV, then it was positioned approximately midway between unc-17 and unc-5 (a 5 map unit interval) by a three-factor genetic cross. Nine of seventeen Unc-5 recombinants issuing from hermaphrodites of genotype + dpy-13(m399) + unc-17 + unc-5 carried the dpy-13 mutation. Linkage to unc-5 and unc-17 was also shown for both m399 and m400 during the crossing procedures used to place these alleles in a Bristol genetic background (see Experimental Procedures).

Alleles harboring insertions of Tc1 may revert spontaneously by perfect or near perfect excision of the element (Moerman and Waterston, 1984; Eide and Anderson, 1985). Two wild-type revertants of the m399 allele were obtained in a visual screen of 1.7 x 10^6 mutant animals from strain DR896 (Table 1). This genetic instability provided a good indication that m399 carried a transposon insertion.

Location of Insertion Sites

The three putative Tc1 insertion mutants and the oriented cosmid contig containing ama-1 provided the elements for molecular analysis of dpy-13, which is 0.05 map unit to the left of ama-1 on the genetic map. Considering the average of 260 kb per map unit in C. elegans (Greenwald et al., 1987), it seemed likely that dpy-13 should be located within 10–30 kb to the left of ama-1. Hence, the 36 kb cosmid CB80016 (from A. Coulson and J. Sulston) was chosen for analysis, and a restriction map was generated (Figure 2). The right end of this cosmid insert carries the 5′ end of ama-1. The left breakpoint of the deficiency mDf4, which deletes ama-1 (Rogalski and Riddle, 1988), is within a 4.6 kb EcoRI fragment approximately 22 kb from the 5′ end of ama-1 (Bird and Riddle, submitted). In the course of identifying coding regions adjacent to ama-1, two collagen genes (col-33 and col-34) were identified, in addition to another gene very close to the 5′ end of ama-1 that encodes a 0.8 kb mRNA.

Purified EcoRI restriction fragments mapping to the left of ama-1 were hybridized to Southern blots of genomic DNA digested with EcoRI to identify sites of transposon insertion in mutant strains. When the 7.1 kb EcoRI fragment mapping approximately 11 kb to the left of ama-1 was used to probe genomic DNA from two of the dpy-13 insertion mutants, an 8.7 kb band was detected as a result of insertion of the 1.6 kb Tc1 element (Figure 3). The presence of a repetitive sequence within the 7.1 kb probe was detected by its hybridization to other genomic fragments even when high stringency washes were employed.

Figure 3 compares genomic DNA from m399 that had
m400 were from a screen of \(2.3 \times 10^6\) animals of strain RW7097 (Table 1). Segregation patterns have occasionally been crossed six times with the Bristol (low frequency transposition) genetic background (lane 1) with DNA from the initial isolates of \(m399\) (lane 2) and \(m400\) (lane 3) in the mutator (high frequency transposition) background. The presence of a faint 7.1 kb band in the latter two lanes presumably reflects the level of spontaneous excision of the element in somatic cells (Eide and Anderson, 1988). The 8.7 kb band in mutant DNA was not present in DNA from wild-type (lane 4), parental (lane 5), and revertant strains (lanes 6 and 7). Hence, the presence of the 1.6 kb insertion was correlated with the \(dpy-13\) mutation, and loss of the insert occurred in two independent revertants. Once the sites of two transposon insertions were localized to the 7.1 kb EcoRI fragment, additional restriction digests were used to further localize the insertion sites (Figure 4). First, genomic DNA of the parental-type strain and three \(Tc1\)-tagged \(dpy-13\) mutants, \(m399\), \(m400\), and \(m401\), were digested with HindIII, then blotted and hybridized with subclones of the 7.1 kb fragment (Figure 5). The 1.3 kb HindIII-EcoRI probe (containing the shaded region in Figure 4) detected a 2.4 kb HindIII fragment from the parental-type strain (Figure 5a, lane 1) and \(m401\) (lane 4), but the band was shifted to 4.0 kb in DNA from \(m399\) (lane 2) and \(m400\) (lane 3). A 1.6 kb shift in \(m401\) (Figure 5b, lane 2) was detected with the 2.4 kb EcoRV-HindIII probe, which is adjacent to the 1.3 kb fragment (Figure 4). Strong hybridization of the 2.4 kb probe to a mutant 5.7 kb HindIII fragment was observed, along with weak hybridization to the parental-type 4.1 kb fragment that presumably arose from spontaneous excision of the element in somatic cells of the mutator strain.

The 1.3 kb probe contains a repetitive sequence that hybridized with varying intensities to approximately 20 bands in the HindIII digests of genomic DNA (Figure 5a, 5c, and 5e). These bands presumably represent the subset of collagen genes (see below) that is most closely related to \(dpy-13\). With the exception of \(m401\) DNA, the adjacent 2.4 kb EcoRV-HindIII probe hybridized strongly to...
because there are EcoRV sites in the mutants m399, adjacent 2.4 kb HindIII-EcoRI fragments were determined as described in Figure 5, using the 1.3 kb probes to the HindIII site, which is between the HindIII and EcoRI sites, and the adjacent 2.4 kb EcoRV-HindIII fragment as probes.

The combined results map the Tc1 insertion sites approximately 200 bp apart, with the m401 site about 80 bp to the right of the HindIII site, and the m399 and m400 sites about 120 bp to its right (Figure 4). No differences were detected between the two latter strains, and it is possible that the m399 and m400 insertions are at the same site, considering the sequence specificity of the Tc1 insertion (Eide and Anderson, 1988). Examination of the wild-type DNA sequence (see Figure 6) reveals possible Tc1 target sequences at locations corresponding to the approximate sites of insertion in the mutants. The 1.6 kb increase in fragment size, and the introduction of EcoRV sites in the mutant DNAs provide the evidence that the mutations are insertions of Tc1, rather than of some other element. The Southern blots were not probed with Tc1, and it is not certain that new bands arising from insertion in dpy-13 could be distinguished because of the high Tc1 copy number, especially in mutator strains (Moerman et al., 1986).

A Deletion Mutation
DNA from the EMS-induced dpy-13 (e184) was then examined for possible rearrangements where the Tc1 insertions occurred in genomic DNA from the wild-type dpy-13 (e184) (lane 2) strains digested with HindIII. A deletion of approximately 2.4 kb was observed when the DNA was hybridized with the Tc1 probe. This deletion is within the 0.5 kb BamHI fragment that is located to the right of the cluster of Tc1 insertions.

dpy-13 is a Collagen Gene
As a means to determine the nature of the product, the wild-type 3.7 kb EcoRV fragment cloned from cosmid CB#B0016, and the m400 fragment cloned into pUC18 and pUC19 for analysis. Preliminary sequence data extended to the HindIII site, which is between the Tc1 insertions and revealed an open reading frame on both sides (left and right) in Figure 2. A 1.5 kb region including the reading frame and flanking regions of the coding sequence was at least twice in both directions.

Figure 5 shows the nucleotide sequence of the dpy-13 region.
**The dpy-13 Collagen Gene**

571

Figure 6. Nucleotide Sequence of dpy-13 with Encoded Amino Acid Sequence

The nucleotide sequence of dpy-13 is shown from the 5' site (116) to 5' of the proposed initiator methionine to the first possible polyadenylation signal, AATAAA. This 1182 nucleotide region was sequenced at least twice in both directions. Introns are indicated in lower case letters and the deduced amino acid sequence is in italics. Nucleotides are numbered consecutively with the A of the initiator ATG as +1, and the preceding nucleotide as −1. The triple-helix forming regions as glycine rich in proline are indicated in lower case letters and the deduced amino acid sequence is in italics. Nucleotides are numbered from the third amino acid are underlined, and the synonymous triple-helix forming regions (underlined in Figure 6 and boxed in Figure 7) are rich in proline. The four Gly-X-Y regions are characteristic of the triple-helical regions of collagen, both in vertebrates and invertebrates (Adams, 1986).

A Deletion Mutation

DNA from the EMS-induced dpy-13 reference allele, e184, was then examined for possible rearrangements in the region where the Tc1 insertions occurred. Figure 5e compares genomic DNA from the wild-type N2 (lane 1) and dpy-13(e184) (lane 2) strains digested with BamHI and HindIII. A deletion of approximately 30 bp was detected when the DNA was hybridized with the 1.3 kb probe. This deletion is within the 0.5 kb BamHI fragment immediately to the right of the cluster of Tc1 insertions (Figure 4).

**dpy-13 is a Collagen Gene**

As a means to determine the nature of the dpy-13 gene product, the wild-type 3.7 kb EcoRI fragment was cloned from cosmid CB#B0016, and subfragments were cloned into pUC118 and pUC119 for DNA sequence analysis. Preliminary sequence data extending to the right of the HindIII site, which is between the Tc1 insertion sites, revealed an open reading frame oriented 5' to 3' (left to right) in Figure 4. A 1.5 kb region including the entire open reading frame and flanking regions was then sequenced at least twice in both directions.

Figure 6 shows the nucleotide sequence of the wild-type dpy-13 gene, starting at a 5' site in the promoter region and ending with the first possible polyadenylation signal, AATAAA. This 1182 nucleotide region was sequenced at least twice in both directions. Introns are indicated in lower case letters and the deduced amino acid sequence is in italics. Nucleotides are numbered consecutively with the A of the initiator ATG as +1, and the preceding nucleotide as −1. The triple-helix forming regions as glycine rich in proline are indicated in lower case letters and the deduced amino acid sequence is in italics. Nucleotides are numbered from the third amino acid are underlined, and the synonymous triple-helix forming regions (underlined in Figure 6 and boxed in Figure 7) are rich in proline. The four Gly-X-Y regions are characteristic of the triple-helical regions of collagen, both in vertebrates and invertebrates (Adams, 1986).
gen genes col-1 and col-2 (Kramer et al., 1982), not only in the Gly-X-Y coding regions, but also in the flanking non-helical forming domains at the amino and carboxyl termini of the encoded proteins. Alignment of the coding sequences to maximize homology revealed that dpy-13 is 72% similar to col-1 and 67% similar to col-2.

The ATG chosen as the initiator codon is the most 5' methionine codon in the Gly-X-Y reading frame, it is within a CCATGG sequence that matches the eukaryotic consensus for optimal translation initiation signals (Kozak, 1986), and it is positioned at a site very similar to the initiator codon of col-2. Alternative dpy-13 reading frames contain multiple stop codons. The distance from the initiator codon to the stop codon is 1014 bp, with exons totaling 906 bp. Two small introns, 58 bp and 50 bp in length (indicated in lowercase letters in Figure 6), interrupt the amino acid sequence at amino acids 27 and 246, respectively. The short introns, which were identified by their low GC content and conserved boundaries, are typical of C. elegans; three-fourths of the introns sequenced thus far are between 45 and 59 bp long (Emmons, 1988). The GC content of Intron 1 is 21%, and that of Intron 2 is 30%, compared with 60% GC for the deduced coding regions.

It is not known whether C. elegans collagens are initially translated as preprocollagens. Vertebrate preprocollagens contain a signal sequence and amino-terminal and carboxy-terminal propeptides that are cleaved off during formation of the mature collagen (Fessler and Fessler, 1978). The amino-terminal region of the deduced dpy-13 polypeptide includes a hydrophobic region that could function as a signal peptide. Codon usage in dpy-13 is highly biased, as it is in other C. elegans genes that encode relatively abundant proteins. For example, all but two of the 54 prolines are encoded by CCA, and all but two of the 58 glycines are encoded by GGA. In other C. elegans genes, 94% of the proline codons are CCA, and 90% of the glycine codons are GGA (Emmons, 1988).

Comparisons with Other C. elegans Collagens

Overall protein organization and amino acid sequence homology between the dpy-13 collagen and two related C. elegans collagens, col-1 and col-2 (Kramer et al., 1982), are shown in Figure 7. The general organization of the dpy-13 sequence is most similar to the col-1 protein. The arrangement and sizes of the four triple-helix forming regions are identical. Even though col-2 differs by having five such regions, the total number of amino acids in these domains is 150 for all three genes.

To obtain the best alignment among the three sequences, six amino acids in dpy-13 (amino acids 84–90 in Figure 6) and five amino acids in col-2 were looped out in the area of least homology. In addition, a one amino acid gap was introduced between the last two helical regions in col-2.

The lower half of Figure 7 indicates the amino acid sequence similarity between dpy-13 and col-1, and dpy-13 and col-2. Each point represents the number of identical amino acids within a window of seven. Shaded areas indicate homology of at least 57% (four or more identical amino acids). The bar labeled GSP indicates the position and size of the gene-specific region used as a probe in Figure 1.

The Gene-Specific Region

A 142 bp BamHI-FokI fragment (nucleotides 301–443 in Figure 6) was chosen as a potential probe because of its low homology with col-1 and col-2 (Figure 7). Also, a 99 bp fragment from an area of dpy-13 (amino acids 138–176) had been shown to be gene-specific, and specific (Kramer et al., 1985). When hybridized to wild-type total DNA (lane 2), whereas the 1 kb HindIII–EcoRI fragment from the entire dpy-13 coding region, detects a number of bands (Figure 8, lane 1) under similar hybridization conditions.

The dpy-13 collagen is most closely related to col-1 and col-2; the other characterized C. elegans collagens have longer or shorter N-terminal or C-terminal domains, and the amino acid sequences of these regions are more divergent. All six regions of Gly-X-Y sequence blocks and the di­ thionine codon in the Gly-X-Y reading frame, it is within a window of seven. Shaded areas indicate homology of at least 57% (four or more identical amino acids). The bar labeled GSP indicates the position and size of the gene-specific region used as a probe in Figure 1.

Figure 7. Comparison of dpy-13 with col-1 and col-2 Collagens

The three deduced collagen polypeptides are schematically illustrated, drawn as straight lines with the N-terminus on the left and the C-terminus on the right. Triple-helix forming regions, with glycine as every third amino acid, are boxed showing the number of amino acids included. Placement of cysteine residues is indicated by solid vertical lines. To maximize amino acid homology, six amino acids in dpy-13 and five in col-2 were looped out in areas of least similarity. In addition, one amino acid gap was introduced between the last two helical regions in col-2.

The lower half of Figure 7 indicates the amino acid sequence similarity between dpy-13 and col-1, and dpy-13 and col-2. Each point represents the number of identical amino acids within a window of seven. Shaded areas indicate homology of at least 57% (four or more identical amino acids). The bar labeled GSP indicates the position and size of the gene-specific region used as a probe in Figure 1.

The Gene-Specific Region

A 142 bp BamHI–FokI fragment (nucleotides 301–443 in Figure 6) was chosen as a potential probe because of its low homology with col-1 and col-2 (Figure 7). Also, a 99 bp fragment from an area of dpy-13 (amino acids 138–176) had been shown to be gene-specific, and specific (Kramer et al., 1985). When hybridized to wild-type total DNA (lane 2), whereas the 1 kb HindIII–EcoRI fragment from the entire dpy-13 coding region, detects a number of bands (Figure 8, lane 1) under similar hybridization conditions.

The dpy-13 collagen is most closely related to col-1 and col-2; the other characterized C. elegans collagens have longer or shorter N-terminal or C-terminal domains, and the amino acid sequences of these regions are more divergent. All six regions of Gly-X-Y sequence blocks and the di­ thionine codon in the Gly-X-Y reading frame, it is within a window of seven. Shaded areas indicate homology of at least 57% (four or more identical amino acids). The bar labeled GSP indicates the position and size of the gene-specific region used as a probe in Figure 1.

The Gene-Specific Region

A 142 bp BamHI–FokI fragment (nucleotides 301–443 in Figure 6) was chosen as a potential probe because of its low homology with col-1 and col-2 (Figure 7). Also, a 99 bp fragment from an area of dpy-13 (amino acids 138–176) had been shown to be gene-specific, and specific (Kramer et al., 1985). When hybridized to wild-type total DNA (lane 2), whereas the 1 kb HindIII–EcoRI fragment from the entire dpy-13 coding region, detects a number of bands (Figure 8, lane 1) under similar hybridization conditions.
The three deduced collagen polypeptides are schematically illustrated, drawn as straight lines with the N-terminus on the left and the C-terminus on the right. Triple-helix forming regions, with glycine as every third amino acid, are boxed showing the number of amino acids included. Placement of cysteine residues is indicated by solid vertical lines. To maximize amino acid homology, six amino acids in dpy-13 and five in col-1 were looped out in areas of least similarity. In addition, a one amino acid gap was introduced between the last two helical regions in col-2.

The lower half of the Figure is in register with the top half, showing the degree of amino acid sequence similarity between dpy-13 and col-1, and dpy-13 and col-2. Each point represents the number of identical amino acids within a window of seven. Shaded areas indicate homology of at least 57% (four or more identical amino acids). The bar labeled GSP indicates the position and size of the gene-specific region used as a probe in Figure 8.

In Other C. elegans Collagens

Further analysis and amino acid sequence homology of other genes in C. elegans have been reported by Kramer et al. (1986) and Kusch and Edgar (1986). The dpy-13 gene is most closely related to col-1 and col-2, the other characterized C. elegans collagen. Both genes have longer or shorter N-terminal domains, or shorter C-terminal domains, and the amino acid sequences of these regions are more divergent. Also, the organization of Gly-X-Y sequence blocks and the distribution of charged amino acids within those domains differs from the dpy-13, col-1, col-2 subfamily (Cox et al., 1986).

Figure 7 indicates the amino acid between dpy-13 and col-1, as well as between col-1 and col-2. Shaded areas indicate regions of amino acid identity (at least four identical amino acids) within a window of seven. Overall, there is 49% amino acid identity between dpy-13 and col-1, 65% between dpy-13 and col-2, 59% between col-1 and col-2. It is an advantage of C. elegans that the mutant phenotypes are not lethal.

Discussion

Our results, together with those of Kramer et al., (1988) on the sqt-1 gene, associate specific genetic defects in body shape with collagen, the primary constituent of the C. elegans cuticle. The cuticle is the exoskeleton of C. elegans, and therefore determines the nematode's shape. The process by which this is accomplished is not understood, and the discovery that dpy-13 and sqt-1 encode collagens establishes the first link between the molecular analysis of the collagen gene family and developmental consequences of altered gene function. The genetic evidence suggests that the null (complete loss of function) phenotype of sqt-1 may be wild-type (Kusch and Edgar, 1986). Mutants in this gene exhibit dominant phenotypes. By contrast, the only dominant allele of dpy-13 is e184, which contains the small 30 bp deletion near the middle of the gene, and this allele is only semi-dominant. Nine other mutant alleles of dpy-13, including the three transposon insertion mutants reported here, are recessive to wild-type, and they range in phenotype from moderate to severe (short) dumpy. Also, heterozygotes carrying deficiency mutations that delete dpy-13 (Pogalski and Riddle, 1988) are normal in body shape. This is interpreted to mean that the dpy-13 allele produces a product with altered function, whereas the recessive alleles represent various degrees of functional loss. It is not yet known whether the e184 mutation deletes part of the gene-specific region or part of the Gly-X-Y domain within the 0.5 kb BamHI fragment. Sequence analysis will determine if it is an in-frame deletion producing a shortened polypeptide that could assemble with other collagen chains causing dpy-13(e184)/+ animals to be semi-dumpy.

The 142 bp BamHI–FokI fragment (nucleotides 222-362 in Figure 6) was chosen as a potential gene-specific probe because of its low homology with col-1 and col-2 (Figure 7). A 142 bp dpy-13 fragment detected only the single 7.1 kb band that carries dpy-13 (Figure 8, lane 2), whereas the 1.3 kb HindIII–EcoRI probe, containing the entire dpy-13 coding region, detected a large number of bands (Figure 8, lane 1) under the same hybridization conditions. This background of repetitive sequences is also present in Figure 5a, 5c, and 5e. On a Northern blot with total RNA from an asynchronous wild-type population, the gene-specific probe detected a transcript of approximately 1 kb (Figure 8, lane 3), demonstrating that dpy-13 is expressed. The dpy-13-specific transcript was detected in an abundance roughly comparable to the transcripts from col-1 and col-2 (Kramer et al., 1985).
Although dpy-13 is a collagen gene, it seems unlikely that all of the identified dumpy genes encode collagens. In principle, any gene required for synthesis, maturation, or secretion of the dpy-13 collagen or related proteins might be detectable as a dumpy mutant. Other indirect effects on body shape are also possible. For example, it has been proposed that muscle hypercontraction might produce a dumpy phenotype (Lewis et al., 1980), and mutations in certain muscle genes such as unc-105 do result in short, paralyzed animals (Park and Horvitz, 1986). Also, one class of dumpy mutants (dpy-21, 26, 27, and 28) is affected in X chromosome dosage compensation, and XXX animals are also dumpy (Hodgkin, 1988). It is possible that the dumpy phenotype in these cases is an indirect effect of altered collagen ratios resulting from overexpression of one or more X-linked genes.

The function of dpy-13 apparently is not replaced by other collagens of related structure, since loss-of-function mutations are the predominant mutant class. If dpy-13 has a unique function within the collagen gene family, that function is presumably reflected in one or more unique structural motifs. A 50 amino acid region just upstream from the first triple-helix forming domain shows very little similarity among the three proteins. This region is the most divergent among the eight collagen genes sequenced thus far (Cox et al., 1986), and this portion of dpy-13 does not hybridize to any other sequences in the C. elegans genome. Divergent domains in gene family members may have assumed unique functions, but it is equally plausible that structurally divergent domains are those in which precise structure is not required for normal function. Hence, comparison of wild-type sequences from related genes is not sufficient to identify structural motifs for gene-specific function. Clearer insights may be obtained from sequence analysis of the mutant alleles that generate a range of dumpy phenotypes, and this analysis is in progress. The positions of the EcoRV sites introduced in dpy-13 by the Tc1 insertions indicate that the m40I insertion is in the promoter region, near the Smal site used as the start point for the sequence in Figure 6, whereas the m399 and m400 insertions appear to be within Intron 1, where there are several possible target sites for Tc1 insertion. Considering the small size of the gene, it should also be relatively simple to identify single base pair changes in other mutants.

Most cuticle components are synthesized during a 3-4 hr period preceding each molt, accounting for approximately 10% of total protein synthesis during these periods (Cox et al., 1981). Thus, the mRNAs for cuticle collagens are expected to be abundant. The collagen mRNAs in C. elegans are small (1.0-1.4 kb) and produce correspondingly small proteins (Politz et al., 1986). The major collagens produced by in vitro translation of C. elegans RNA are 30-40 kd in apparent molecular mass (Politz and Edgar, 1984). However, the collagens found in C. elegans cuticles range in molecular mass from 50 kd to 200 kd (Cox et al., 1981), much larger than could be encoded by the mRNAs. It seems likely that the discrepancy is due to covalent cross-linking of cysteine and lysine residues between individual collagen chains, as has been demonstrated for other collagens (Seigel, 1979). Studies of the cuticle collagens of C. elegans have demonstrated that they are extensively cross-linked by disulfide bonds as well as by nonreducible, covalent crosslinks (Cox et al., 1981). Other than the gene-specific domain (amino acids 56-101), which may impart gene-specific functions, the primary structural features that distinguish dpy-13 from col-1 or col-2 are the additional cysteine and lysine residues that give the dpy-13 collagen a greater potential for covalent cross-linking.

Collagen is found in all metazoan phyla (Adams, 1978), and is a major protein of the extracellular matrix. The genetic approach to studying collagen secretion, assembly, and function is potentially of great value to the understanding of these biologically important processes in both vertebrates and invertebrates. However, little is known about the molecular genetics of collagen in any animal, and the availability of a large collection of mutants affected in body shape may provide the key to understanding the developmental consequences of specific alterations in collagen structure in this model organism.

Experimental Procedures

Nematode Strains and Culture Conditions

Nematodes were grown at 20°C on NG agar plates seeded with E. coli strain OP50 (Brenner, 1974), or in liquid cultures with a 5% w/v suspension of E. coli strain X1666 in S medium (Suiston and Brenner, 1974).

Nematode strains used in this work include the wild-type Bristol strain N2 (Brenner, 1974) and the following strains: CB904, unc-17(e245) IV; DR821, unc-17(e245) unc-5(e633) IV, and CB1039, unc-5(e633) IV; nuclear(e1392) X. The dpy-13 mutants are listed in Table 1. Mutator strains derived from Bristol/Bergerac hybrids (Moerman and Waterston, 1984) include RW2079, mut-8(a702) IV, and DR842, an unc-22(+) revertant of RW2079, mut-8(a702) I; mut-5(a701) II; unc-22(a136-761) IV; DR948, a derivative of DR842 that had been crossed three times with dpy-13(e194) to remove the parental dpy-13(+) region from the mutator background, was the source of DNA used as a control in some hybridizations. No restriction site differences were detected between any of the dpy-13(+) strains, DR842, RW2079, DR842, or N2. Genetic nomenclature follows guidelines described by Horvitz et al. (1979).

Transposon Mutagenesis

Populations of DR842 or RW2079 were started by placing one hermaphrodite on each of 40-60 100 x 15 mm petri dishes spread with OP50 and incubating the worms at 20°C for 7-10 days, until the food supply was almost depleted. The progeny (F2-F3 generations) were then screened visually for the presence of dumpy mutants. The number of progeny scored (L4 larvae and adults) was estimated by collecting worms from selected plates in a known amount of buffer and extrapolating from the population counts in small aliquots. Worms carrying mutations in numerous other genes were also obtained. Screens for revertants were set up in a similar manner using DR986, dpy-13(m399), as the parent (Table 1).

Genetics

Methods for genetic complementation and mapping have been described by Brenner (1974). Three dumpy mutations failed to complement dpy-13(e184), as indicated by the presence of F1 dumpy males arising from a cross between dpy-13(e184) males and mutant dumpy hermaphrodites. Unlike e184, the insertion mutations are recessive to wild type. A strain of genotype +; dpy-13(m399) / unc-17 +; unc-5 was constructed to position m399 relative to unc-17 and unc-5. Individual unc-5 recombinants among the progeny of these hermaphrodites were selected and allowed to reproduce to determine whether they segregated the dpy-13 mutation.

Selected mutations were crossed into Bristol genetic backgrounds to produce stable strains carrying fewer copies of Tc1. The dpy-13 muta-
The dpy-13 Collagen Gene

575

Collagens (Seigel, 1979). Studies of the C. elegans have demonstrated that y cross-linked by disulfide bonds as collagenous domain (amino acids 15-32), a tropoelastin specific region (amino acids 33-49), and a tropoelastin specific region (amino acids 50-65). These domains are necessary for the formation of the collagen triple helix and for the stability of the collagen molecule. The tropoelastin specific region is responsible for the formation of the junctional collagens. The tropoelastin specific region is responsible for the formation of the junctional collagen triple helix.

The dpy-13 Collagen Gene

575

Collagens (Seigel, 1979). Studies of the C. elegans have demonstrated that y cross-linked by disulfide bonds as collagenous domain (amino acids 15-32), a tropoelastin specific region (amino acids 33-49), and a tropoelastin specific region (amino acids 50-65). These domains are necessary for the formation of the collagen triple helix and for the stability of the collagen molecule. The tropoelastin specific region is responsible for the formation of the junctional collagens. The tropoelastin specific region is responsible for the formation of the junctional collagen triple helix.

The dpy-13 Collagen Gene

575

Collagens (Seigel, 1979). Studies of the C. elegans have demonstrated that y cross-linked by disulfide bonds as collagenous domain (amino acids 15-32), a tropoelastin specific region (amino acids 33-49), and a tropoelastin specific region (amino acids 50-65). These domains are necessary for the formation of the collagen triple helix and for the stability of the collagen molecule. The tropoelastin specific region is responsible for the formation of the junctional collagens. The tropoelastin specific region is responsible for the formation of the junctional collagen triple helix.


A Protein Composed of 10 Subunits Is Encoded by Various Genes with Similarity to ATPase Subunits

Bruce Hay, Lily Yeh Jan, and Yuh Chia Chen

Howard Hughes Medical Institute
Department of Physiology
and Department of Biochemistry
University of California
San Francisco, California 94143

Summary

Determinants of pole cells, which form at the posterior pole of the embryo, are provided maternally to the posterior pole of the Drosophila egg. Polar granules, DNA molecules associated with polar granules, and antiserum to polar granules are necessary for pole cell determination. The DNA probe NaeH4F111 is unique to pole cells. We have cloned the gene for one of these DNA molecules. This gene turns out to be vasa, a gene maternally provided for the formation of pole cells. The gene product is a putative ATPase with a similarity to ATPase subunits.

Introduction

The precursors of the Drosophila germ line, termed pole cells, form at the posterior pole of the embryo. The gene vasa is necessary for pole cell formation. Antisense RNA isolated from late-stage oocytes can restore the ability of UV-irradiated pole cells to form pole cells, although these pole cells are not sufficient to produce germ cells (Okada et al., 1974; Ilimenese and Mahowald, 1975; Ilimensu et al., 1976; Niki, 1986; Horvitz et al., 1986). Several observations associate the function of pole cells with polar granules, which are dense cytoplasmic organelles without limiting membranes (Okada et al., 1972, 1968). First, polar granules appear at the posterior pole of the embryo (Mahowald, 1971) and early embryos (Counce, 1963; Mahowald, 1971). In the dumpy (a mutant of the germ plasm) and the anarchous (a mutant of oogenesis and embryogenesis whose oocytes lack a posterior pole), polar granules are not formed (Okada et al., 1972). Interestingly, polar granules are associated with RNA (Mahowald, 1971) and are lost during oogenesis and embryogenesis when the germ line is capable of inducing germ line development at late stages, e.g., after they are segregated. Finally, the loss of maternally supplied DNA is required for the establishment of the germ line in clonal mutants, such as the grandchildless-knirps class causing a failure of pole cell formation (Boswell and Mahowald, 1985; Schüpbach, 1986; Lehmann and Nüsslein-Volhard, 1986).