Genetic Identification, Sequence, and Alternative Splicing of the *Caenorhabditis elegans* α2(IV) Collagen Gene

Marion H. Sibley, Jeffrey J. Johnson, Craig C. Mello,* and James M. Kramer

Department of Cell, Molecular and Structural Biology, Northwestern University Medical School, Chicago, Illinois 60611; and *Fred Hutchinson Cancer Research Center, Seattle, Washington 98104

Abstract. The nematode *Caenorhabditis elegans* has two type IV collagen genes homologous to the mammalian α1(IV) and α2(IV) collagen genes. We demonstrate by transgenic rescue of mutant animals that the genetic locus encoding the *C. elegans* α2(IV) collagen gene is *let-2* on the X chromosome. The most severe effect of mutations in *let-2* is temperature-sensitive embryonic lethality. The embryonic lethal phenotype is similar to that seen in animals with mutations in the α1(IV) collagen gene, *emb-9*. The sequence of the entire *C. elegans* α2(IV) collagen gene is presented. Comparisons with mammalian type IV collagen sequences show high amino acid sequence conservation in the C-terminal NC1 domain and of crosslinking residues (Cys and Lys) in the N-terminal 7S domain. RT-PCR analysis shows that transcripts of the *C. elegans* α2(IV) collagen gene are alternatively spliced. Transcripts contain one of two mutually exclusive exons, exon 9 or 10. These exons encode very similar products, differing primarily in the sequence of a 9-10 amino acid Gly-X-Y interruption. The expression of these alternatively spliced α2(IV) collagen transcripts is developmentally regulated. In embryos over 90% of the α2(IV) collagen mRNA contains exon 9, while larval and adult RNAs contain 80-90% exon 10. This shift in expression of alternative α2(IV) collagen transcripts suggests that *C. elegans* embryos may require a different form of α2(IV) collagen than do larvae and adults.

BASEMENT membranes are thin sheets of extracellular matrix that underlie epithelial and endothelial cells; they also surround individual muscle cells, nerve cells, and adipocytes. Basement membranes are thought to be involved in many critical biological processes, including cell adhesion, cell migration, morphogenesis, tissue regeneration and repair, and macromolecular filtration (Timpl, 1989; Farquhar, 1991). Type IV collagen is a major structural component of basement membranes (Glanville, 1987). The type IV collagen molecule is a heterotrimer of about 500 kD, most commonly composed of two α1(IV) chains and one α2(IV) chain. Three additional type IV collagen chains (α3, α4, and α5) have recently been identified (Saus et al., 1988; Morrison et al., 1991; Gunwar et al., 1990; Hostikka et al., 1990). In contrast to the ubiquitous α1(IV) and α2(IV) chains, these additional chains appear to be localized primarily, though not exclusively, in kidney glomerular basement membranes. In the type IV collagen molecule, three α chains are arranged in a collagenous $(Gly\text{-}X\text{-}Y)_n$ triple helix that comprises about three-fourths of the molecule. The Gly-X-Y sequence is punctuated with numerous interruptions, ranging from 1 to 24 amino acids in length. These interruptions are thought to produce sites of flexibility in the triple helical domain, and they have also been proposed to act as binding sites for other basement membrane components (Charonis and Tsilibary 1990). Each type IV α chain has two noncollagenous domains, one at the NH_2 terminus, which forms part of the 7S domain, and one at the COOH-terminus, the NC1 domain.

Type IV collagen molecules aggregate to form a complex three-dimensional network of highly branched filaments, which is the major structural support of the basement membrane (Yurchenco and Furthmayr, 1984). The type IV collagen network is formed by a variety of interactions, including formation of tetramers through 7S domains, formation of dimers through NC1 domains, and lateral interactions between molecules along the length of the triple-helical domain (Yuchenco and Ruben, 1987). In addition to these self-interactions, type IV collagen binds cell surface receptors, and also binds other basement membrane components, such as laminin, heparan sulfate proteoglycans, and nidogen (Charonis and Tsilibary, 1990). Any or all of these additional binding activities may be involved in basement membrane network formation, but the detailed structure of this network is as yet unknown.

In an effort to understand the role of type IV collagen in the functioning of basement membranes, we have undertaken a genetic and molecular characterization of type IV collagen in the nematode *Caenorhabditis elegans*. We have shown that *C. elegans* has two type IV collagen genes, homologous to the α1(IV) and α2(IV) genes of mammals (Guo and Kramer, 1989). The NC1 domains of the α1(IV) and

α2(IV) collagen chains of *C. elegans* have 66 and 72% amino acid sequence identity, respectively, to the NC1 domains of human α1(IV) and α2(IV) collagen. The two *C. elegans* type IV collagen genes are on different chromosome (Guo and Kramer, 1989). In contrast, in mouse and human, the α1(IV) and α2(IV) collagen genes are on the same chromosome about 140-bp apart, and are transcribed from a common bidirectional promoter (Burbelo et al., 1988; Kaytes et al., 1988; Poschl et al., 1988; Soininen et al., 1988). We have previously identified the genetic locus for the *C. elegans* α1(IV) collagen gene as *emb-9* on chromosome III (Guo et al., 1991). Mutations in *emb-9* cause temperature-sensitive lethality during late embryogenesis, demonstrating that defects in basement membranes can disrupt normal embryonic development.

In this work, the genetic locus for the α2(IV) collagen chain in *C. elegans* is identified as *let-2* on the X chromosome. Like the α1(IV) collagen gene in *C. elegans*, mutations in the α2(IV) collagen gene also have temperature-sensitive embryonic lethal phenotypes. The sequence of the entire *C. elegans* α2(IV) chain is reported and compared with other α2(IV) collagen chains. We also show that the *C. elegans* α2(IV) chain undergoes alternative splicing to produce two different transcripts whose relative levels change dramatically during development.

Materials and Methods

Transgenic Rescue of let-2 Mutant Animals

The phage clone CH#1, containing the entire wild-type sequence of the α2(IV) collagen gene (Guo and Kramer, 1989) was injected into *let-2* mutant animals as described in Mello et al. (1991). In most cases CH#1 was co-injected with a clone (pRF4) containing a mutant *rol-6* gene (Kramer et al., 1990), which acts as a dominant cotransformation marker (Mello et al., 1991). Adult *let-2(g25)* and *let-2(g30)* hermaphrodites raised at 15°C were injected once in each distal gonad arm. Injected animals were placed at 25°C, and the incidence of viable and fertile offspring produced at 25°C was observed.

DNA Sequence Analysis

Fragments of the original phage clone CH#1 (Guo and Kramer, 1989) were subcloned into Bluescribe. Nucleotide sequences were determined from both strands by the Sanger dideoxynucleotide chain-termination method (Sanger et al., 1977) using Sequenase (United States Biochemical Corporation, Cleveland, OH), using both universal and sequence-specific oligonucleotide primers. In some cases, sequences were obtained from nested deletions of larger clones made using exonuclease III and S1 nuclease (Henikoff, 1984).

Determination of the Start of Transcription by Primer Extension

Primer extension and RNA sequencing were performed essentially as described by Bektesh et al. (1988). An end-labeled 20-mer synthetic oligonucleotide primer located 237-bp downstream of the start codon ATG was annealed to 60 μg *C. elegans* total RNA and extended with avian myeloblastosis virus reverse transcriptase in the presence of individual ddNTPs to determine the RNA sequence, or in the absence of ddNTPs to determine the termination site.

Preparation of Total RNA from Nematodes at Different Developmental Stages

C. elegans (N2, wild-type strain) was grown and maintained on nematode growth medium plates, streaked with *Escherichia coli* strain OP50, according to the standard techniques described by Brenner (1974). For these developmental studies nematodes were grown at 25°C. Embryos at early stages of development were collected by treating gravid adults (~250 μl packed worms) with alkaline hypochlorite solution (30% sodium hypochlorite: 1 M KOH: H_2O 1:2:5). This treatment dissolved adults and released embryos into solution. Embryos were collected by centrifugation at 2,000 *g* for 2 min, washed several times in M9 buffer (Wood, 1988) and total RNA was immediately extracted by resuspending the embryo pellet in 800 μl of TRI phenol-guanidine reagent (Molecular Research Center, Inc., Cincinnati, OH). The suspension was sonicated with a Branson Sonifier 450, and the sonicate was incubated at room temperature for 5 min to dissolve ribonucleoprotein complexes. 160 μl chloroform was added, the sonicate incubated a further 2–3 min at room temperature, and then centrifuged at 12,000 *g* for 15 min at 4°C. RNA was precipitated from the upper aqueous layer by addition of 400 μl isopropanol and incubation at 4°C for 30 min. The total RNA pellet was collected by centrifugation at 10,000 *g* for 10 min at 4°C, washed once with 70% ethanol, and air dried. The RNA pellet was resuspended in 20 μl of diethylpyrocarbonate-treated water and stored at −70°C. Embryos at later stages of development were collected by allowing adults to lay eggs for 2 h, washing the adults off the plate with M9 buffer, and incubating the remaining embryos for a further 4 h. These embryos were collected by alkaline hypochlorite treatment and centrifugation, and extracted immediately for total RNA as described above. Nematodes at the first larval stage (L1) were collected by allowing adults to lay eggs on a large (100 mm) plate overnight, washing off the adults, and incubating the plate for 1 h. Larvae that hatched in this time were collected by centrifugation, and either processed for RNA extraction immediately, or put on a fresh plate and incubated a further 3 h or a further 7 h before RNA extraction. Nematodes at the later larval stages (L2 through L4) and adult stage were collected by allowing adults to lay eggs for 2 h, washing off the adults with M9 buffer, and allowing the embryos to develop for a further 24 h (L2 larvae), 33 h (L3 larvae), 43 h (L4 larvae), or 67 h (adult) before RNA isolation.

RNA from sterile adults was prepared using the *C. elegans* strain BA1, which has the temperature-sensitive mutation *fer-1(hcl)* (Ward and Miwa, 1978). The phenotype of this *fer-1* allele is wild-type at 15° and 20°C, but sterile at 25°C, because of defective sperm. For production of sterile adults, a population of *fer-1* adults grown at 20°C was allowed to lay eggs for several hours at 20°C. The adults were then discarded, eggs were shifted to 25°C and allowed to complete larval development at 25°C until a population of sterile adults was obtained. Total RNA was extracted from these sterile adults as described above.

Reverse Transcriptase and Polymerase Chain Reactions for Determination of the Relative Levels of Exons 9 and 10

Synthesis of cDNA was performed by incubating 1–2 μl stage-specific RNA with 50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 1 mM each dNTP, 20 U RNasin (Promega Biotec, Madison, WI), 200 U MMLV reverse transcriptase (GIBCO-BRL, Gaithersburg, MD), and 0.1 pmole specific downstream primer (see Fig. 4) in a total volume of 20 μl at 37°C for 60 min, followed by 95°C for 10 min. For the PCR, 1–10 μl of this reverse transcriptase reaction was used, along with 200 μM each dNTP, 1 μM each specific upstream and downstream primer (Fig. 4), and 2.5 U AmpliTaq (Perkin Elmer Cetus) in PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin) in a total volume of 100 μl. PCR samples were overlaid with mineral oil and incubated at 94°C for 1 min 15 s, at 59°C for 2 min, and at 72°C for 3 min for 30 cycles in an automated temperature cycler. This PCR reaction was then diluted 1:100 to be used in an additional PCR amplification to incorporate ^{32}P-labeled dATP into the PCR product. This additional amplification was performed by incubating 5 μl diluted PCR product with dTTP, dGTP, and dCTP (200 μM each dNTP), 10 μCi α-[^{32}P]dATP (Amersham Corp., Arlington Heights, IL), 1 μM each specific upstream and downstream primer, and 2.5 U AmpliTaq in PCR buffer in a total volume of 100 μl. The samples were overlaid with mineral oil and three to seven PCR temperature cycles were performed as described above. The reamplified PCR products were digested without further purification with either or both of the restriction enzymes BglII and DdeI. The restriction digestion products were separated on 8% acrylamide gels and the gels were dried under vacuum. The relative amounts of digestion product obtained with each restriction enzyme were quantified by exposing the dried gels to an imaging plate (Fuji) for 15 min to 1 h and scanning the imaging plate in a FUJIX BAS 2000 Bioimaging Analyzer (Fuji Photo Film Co., Ltd., Tokyo, Japan).

Results

Identification of let-2 As the Genetic Locus Encoding the α2(IV) Collagen Chain in C. elegans

We have previously shown that the α1(IV) collagen chain of *C. elegans* is encoded by the *emb-9* gene on chromosome III (Guo et al., 1991). In our search for the genetic locus of the α2(IV) collagen chain of *C. elegans* we assumed that mutations in the α1(IV) and α2(IV) collagen genes should have similar phenotypes. Mutant alleles of *emb-9* display a temperature-sensitive, embryonic lethal phenotype. Lethality occurs late in the morphogenic phase of embryogenesis, and is preceded by gross morphological abnormalities. The *emb-9* locus is highly mutable, in that random screens of the entire *C. elegans* genome for temperature-sensitive embryonic lethal mutations produced five *emb-9* alleles, while most other genes were represented by only one or two alleles (Wood et al., 1980; Miwa et al., 1980; Cassada et al., 1981). High mutability has been observed in other collagen genes, including the human type I collagen gene (Kuivaniemi et al., 1991).

The *C. elegans* α2(IV) gene has been mapped to the right arm of the X chromosome (Guo and Kramer, 1989). The *let-2* gene maps to this region and has properties similar to *emb-9*. Embryonic lethality in *let-2* mutants occurs late in the morphogenic phase of embryogenesis (Isnenghi et al. 1983). By Nomarski optics, arrested *emb-9* and *let-2* mutant embryos show very similar morphological abnormalities preceding lethality. The embryonic temperature-sensitive periods of *let-2* and *emb-9* are similar, occurring relatively late in embryogenesis, between the lima and pretzel stages (Isnenghi et al., 1983; Wood et al., 1980). Fig. 1 shows the similar morphological abnormalities observed in arrested *let-2* and *emb-9* mutant embryos raised at the nonpermissive temperature of 25°C. Elongation is arrested at the twofold stage, when embryos have elongated to about two times their original length. Subsequently, the arrested mutant embryos develop extensive constrictions and herniations (Fig. 1, *A* and *B*). In comparison, a wild-type (N2 strain) embryo at the twofold stage of elongation has a smooth appearance (Fig. 1 *C*). Like *emb-9*, *let-2* is a highly mutable gene, as random screens of the entire *C. elegans* genome for EMS-induced temperature-sensitive embryonic lethal mutations yielded four *let-2* alleles (Wood et al., 1980; Cassada et al., 1981).

Encouraged by these indications that *let-2* may be the gene for the α2(IV) collagen chain, we performed transgenic rescue experiments. When raised at the restrictive temperature of 25°C, 100% of *let-2(g25)* and *let-2(g30)* mutants die during embryogenesis. In the transgenic rescue experiments, the gonads of worms homozygous for the *g25* and *g30* alleles were injected with a phage clone containing the entire wild-type coding sequence of α2(IV) collagen, and the animals placed at 25°C. Viable and fertile offspring were recovered from injected *g25* and *g30* animals. Several of these offspring produced transgenic lines that are continuously viable and fertile at 25°C, demonstrating that the phage DNA contains everything needed for rescue of the temperature-sensitive lethal phenotype of *let-2*. These results provide convincing evidence that *let-2* is the α2(IV) collagen gene of *C. elegans*. In addition, 17 *let-2* mutations have been shown to be single nucleotide alterations within the coding region of the α2(IV) collagen gene (Sibley, M., P. Graham, and J. Kramer, manuscript in preparation).

Structure of the C. elegans α2(IV) Collagen Gene

The *C. elegans* α2(IV) collagen gene is ~9-kb long and contains 20 exons (Fig. 2). Primer extension and direct RNA sequencing experiments confirm that the transcription start site is 510-bp upstream of the ATG start codon (Fig. 2). The *C. elegans* α2(IV) collagen gene is comparable in size to the *C. elegans* α1(IV) collagen gene (Guo et al., 1991), but is much smaller than the human type IV collagen genes, which are both >100-kb long (Soininen et al., 1989; Cutting et al., 1988). The deduced translation product of the *C. elegans* α2(IV) gene is 1,758 amino acids long, somewhat longer than the human α2(IV) chain of 1,712 amino acids (Hostikka and Tryggvason, 1988). The coding region of the *C. elegans* α2(IV) gene (Fig. 2) has a characteristic type IV collagen structure, consisting of a large (1,487 amino acids) central triple helical domain having 15 interruptions in the Gly-X-Y repeating sequence, and two smaller noncollagenous domains, one at the NH_2-terminus and one at the COOH terminus (NC1). The gene has a potential AATAAA polyadenylation signal 405 bp from the end of the coding sequence.

The two *C. elegans* type IV collagen genes have substantially different intron/exon structures. The α2(IV) gene has 20 exons, while the α1(IV) gene has only 12 (Guo et al., 1991). There is no conservation of exon size or intron/exon boundaries between the two *C. elegans* type IV collagen genes. Despite their dissimilarity in intron/exon structure, the coding regions of the major domains (7S, triple helical, and NC1) are all nearly identical in length in the two genes. The two genes also have a considerable degree of amino acid sequence identity (50% overall). The human α1(IV) and α2(IV) collagen genes also have quite different gene structures yet highly conserved amino acid sequence (Hostikka and Tryggvason, 1987).

Amino Acid Sequence of C. elegans α2(IV) Collagen

Fig. 3 shows the complete amino acid sequence of the *C. elegans* α2(IV) collagen. The percentage of amino acid sequence identity between *C. elegans* α2(IV) and other type IV chains is presented in Table I. The α2(IV) chain of the parasitic nematode *Ascaris suum* (Pettitt and Kingston, 1991) has very high amino acid sequence identity with the *C. elegans* α2(IV) chain in all domains, including Gly-X-Y interruption sequences. On the other hand, the human α1(IV) (Brazel et al., 1987) and α2(IV) (Hostikka and Tryggvason, 1988) chains have relatively low amino acid sequence identities with the *C. elegans* α2(IV) chain, except in the COOH-terminal NC1 domain.

The primary transcript of the *C. elegans* α2(IV) gene is alternatively spliced to produce two potential polypeptides, one containing exon 9 and one containing exon 10. The exon 10–containing version (residues 229 to 265) is shown in Fig. 3. The amino acid sequence of exon 9 is shown in Fig. 4. The alternative splicing of the α2(IV) transcript is described in detail below.

A comparison of the *C. elegans* α1(IV) and α2(IV) chains is of functional interest, since these two chains are expected to be in the same heterotrimer ($\alpha1_2\alpha2$). In the *C. elegans*

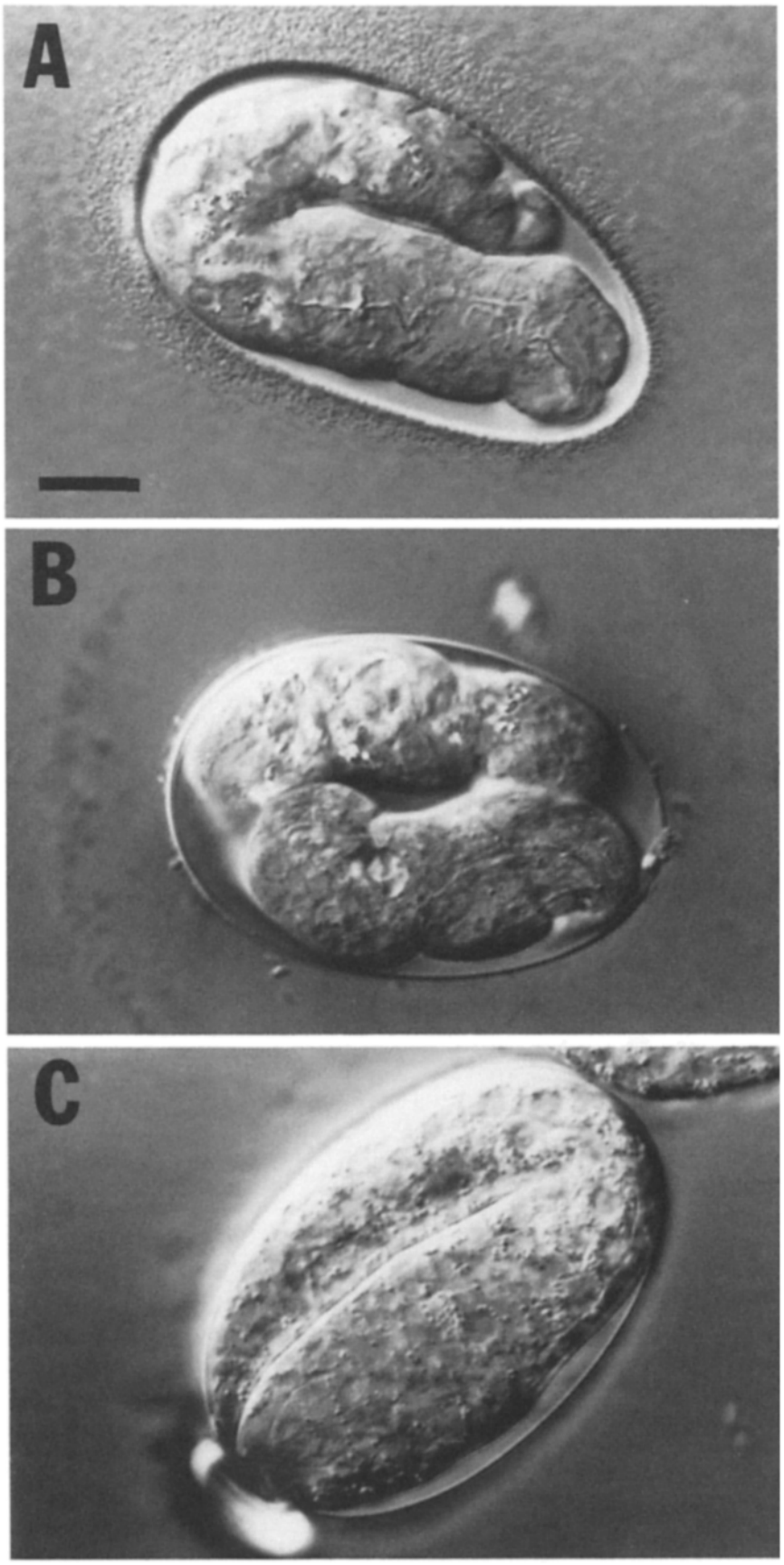

Figure 1. DIC micrographs of *C. elegans* embryos grown at 25°C. (*A*) Arrested *emb-9(g23)* mutant embryo. (*B*) Arrested *let-2(b246)* mutant embryo. (*C*) N2 wild-type embryo at the twofold stage. Bar, 10 μm.

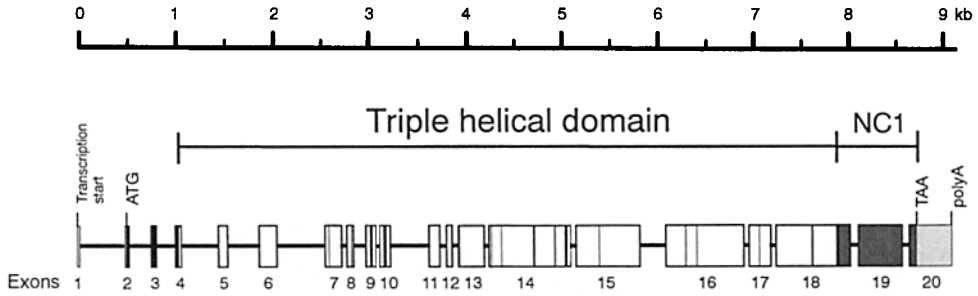

Figure 2. Structure of the *C. elegans* α2(IV) collagen gene. Exons are indicated by boxes, introns by horizontal lines. The Gly-X-Y repeating triple helical domain is unshaded, the NH_2-terminal noncollagenous domain is black, and the COOH-terminal NC1 domain is darkly shaded. 5′- and 3′-untranslated regions are lightly shaded. Interruptions in the Gly-X-Y repeats are shown as solid vertical bars. The transcription start site determined by primer extension is indicated, as is a potential polyadenylation signal 405 bp from the end of the coding sequence. A nucleotide scale is shown in kilobases.

```
<---  Signal peptide ---->              <---  Start of Gly-X-Y domain
MKQRAALGPVLRLAILALLAVSYVQSQATCRDCSNRGCFCVGEKGSMGAPGPQGPPGTQGIRGFPGPEGLAGPKGLKGAQGPPGPVGIKGDRGAVGVPGF     100

PGNDGGNGRPGEPGPPGAPGWDGCNGTDGAPGIPGRPGPPGMPGFPGPPGMDGLKGEPAIGYAGAPGEKGDGGMPGMPGLPGPSGRDGYPGEKGDRGDTG     200

NAGPRGPPGEAGSPGNPGIGSIGPKGDFGDIGAMGPAGPPGPIASTMSKGTIIGPKGDLGEKGEKGEPGEGGQRGYPGNGGLSGQPGLPGMKGEKGLSGP     300

AGPRGKEGRPGNAGPPGFKGDRGLDGLGGIPGLPGQKGEAGYPGRDGPKGNSGPPGPPGGGTFNDGAPGPPGLPGRPGNPGPPGTDGYPGAPGPAGPIGN     400

TGGPGLPGYPGNEGLPGPKGDKGDGGIPGAPGVSGPSGIPGLPGPKGEPGYRGTPGQSIPGLPGKDGKPGLDGAPGRKGENGLPGVRGPPGDSLNGLPGA     500

PGQRGAPGPNGYDGRDGVNGLPGAPGTKGDRGGTCSACAPGTKGEKGLPGYSGQPGPQGDRGLPGMPGPVGDAGDDGLPGPAGRPGSPGPPGQDGFPGLP     600

GQKGEPTQLTLRPGPPGYPGLKGENGFPGQPGVDGLPGPSGPVGPPGAPGYPGEKGDAGLPGLSGKPGQDGLPGLPGNKGEAGYGQPGQPGFPGAKGDGG     700

LPGLPGTPGLQGMPGEPAPENQVNPAPPGQPGLPGLPGTKGEGGYPGRPGEVGQPGFPGLPGMKGDSGLPGPPGLPGHPGVPGDKGFGGVPGLPGIPGPK     800

GDVGNPGLPGLNGQKGEPGVGVPGQPGSPGFPGLKGDAGLPGLPGTPGLEGQRGFPGAPGLKGGDGLPGLSGQPGYPGEKGDAGLPGVPGREGSPGFPGQ     900

DGLPGVPGMKGEDGLPGLPGVTGLKGDLGAPGQSGAPGLPGAPGYPGMKGNAGIPGVPGFKGDGGLPGLPGLNGPKGEPGVPGMPGTPGMKGNGGLPGLP    1000

GRDGLSGVPGMKGDRGFNGLPGEKGEAGPAARDGQKGDAGLPGQPGLRGPQGPSGLPGVPGFKGETGLPGYGQPGQPGEKGLPGIPGKAGRQGAPGSPGQ    1100

DGLPGFPGMKGESGYPGQDGLPGRDGLPGVPGQKGDLGQSGQPGLSGAPGLDGQPGVPGIRGDKGQGGLPGIPGDRGMDGYPGQKGENGYPGQPGLPGLG    1200

GEKGFAGTPGFPGLKGSPGYPGQDGLPGIPGLKGDSGFPGQPGQEGLPGLSGEKGMGGLPGMPGQPGQSIAGPVGPPGAPGLQGKDGFPGLPGQKGESGL    1300

SGLPGAPGLKGESGMPGFPGAKGDLGANGIPGKRGEDGLPGVPGRDGQPGIPGLKGEVGGAGLPGQPGFPGIPGLKGEGGLPGFPGAKGEAGFPGTPGVP    1400

GYAGEKGDGGLPGLPGRDGLPGADGPVGPPGPSGPQNLVEPGEKGLPGLPGAPGLRGEKGMPGLDGPPGNDGPPGLPGQRGNDGYPGAPGLSGEKGMGGL    1500

                                |<---  Start of NC1 domain
PGFPGLDGQPGGPGAPGLPGAPGAAGPAYRDGFVLVKHSQTTEVPRCPEGQTKLWDGYSLLYIEGNEKSHNQDLGHAGSCLQRFSTMPFLFCDFNNVCNY    1600

ASRNDKSYWLSTSEAIPMMPVNEREIEPYISRCAVCEAPANTIAVHSQTIQIPNCPAGWSSLWIGYSFAMHTGAGAEGGGQSLSSPGSCLEDFRATPFIE    1700

CNGARGSCHYFANKFSFWLTTIDNDSEFKVPESQTLKSGNLRTRVSRCQVCVKSTDGRH                                              1759
```

Figure 3. The amino acid sequence of the *C. elegans* α2(IV) collagen chain encoded by *let-2*. Interruptions in Gly-X-Y repeating sequence are underlined. The sequence encoded by the alternatively spliced exon 10 is boxed. The predicted signal peptide is indicated. This predicted signal peptide of 26 residues would result in three amino acids remaining at the NH_2 terminus preceding the most NH_2 terminal of the conserved cysteines in the 7S domain.

```
Exon 8
GGA CCA CGT GGA CCA CCT GGA GAG GCT GGA TCC CCA GGA AAC CCA  2821
Gly Pro Arg Gly Pro Pro Gly Glu Ala Gly Ser Pro Gly Asn Pro

GGA ATC GGA AGC ATT GGA CCA AAA GGA GAT CCT gtg aga aat tgg  2866
Gly Ile Gly Ser Ile Gly Pro Lys Gly Asp Pro

ggc tat ttg ggt ttt ggg ata ttt cta gaa att ttg ttt ctc cct  2911
ttc aga aat gct ttt tca ttc tga act cca atc att tta att att  2956
                    Exon 9 BglII
gta ttt tct aaa ttc tag GGA^GAT CTA GGT TCT GTC GGA CCA CCA  3001
                        Gly Asp Leu Gly Ser Val Gly Pro Pro

GGT CCA CCG GGA CCA CGT GAG TTC ACC GGA TCC GGC TCA ATT GTC  3046
Gly Pro Pro Gly Pro Arg Glu Phe Thr Gly Ser Gly Ser Ile Val

GGA CCT CGC GGA AAC CCT GGA GAA AAG GGA GAC AAG gta ctg ttt  3091
Gly Pro Arg Gly Asn Pro Gly Glu Lys Gly Asp Lys
                        Exon 10
att aat tct ttc ttt ttt tag GGA GAC ATT GGT GCG ATG GGA CCG  3136
                            Gly Asp Ile Gly Ala Met Gly Pro

GCC GGT CCG CCA GGC CCA ATC GCC TCC ACC ATG TCC AAG GGA ACC  3181
Ala Gly Pro Pro Gly Pro Ile Ala Ser Thr Met Ser Lys Gly Thr
                  DdeI
ATT ATC GGT CCT^AAG GGA GAC CTA GGA GAG AAG GGA GAG AAG gca  3226
Ile Ile Gly Pro Lys Gly Asp Leu Gly Glu Lys Gly Glu Lys

aga ata ctt aga gca ggc taa ctg aac gat tcc aac cat tta caa  3271
ttt ttt aca ccc gcc tta ttt acg cta act cta aca gtt ttc gcc  3316
tca gtt tct aaa aga agc ctt ttt ttc aaa tta ctt cta ctt tct  3361
tgc ctt tca agc ttc ttt tgc atg atc ttc cca gag ttc ata tag  3406
act aga ata gct gta tta ttt ggc att gga aac ttc tac tac ttc  3451
tag ctc tgc ctt att ctc ttt tca ttt ttt gaa acg gaa aag aga  3496
aca aaa ttt taa act tac acg ttt act gtt agt acg aat caa taa  3541
ctg cat agc tgg ttt tcc gtg tgt gtt ttg cta tac atc taa ttt  3586
                                                Exon 11
ccc ctc tat gtt tct aat atg ttt tgt gttgcag GGA GAG CCA GGA  3632
                                               Gly Glu Pro Gly

GAG GGA GGT CAA CGC GGT TAC CCA GGA AAT GGA GGA CTC TCA GGA  3677
Glu Gly Gly Gln Arg Gly Tyr Pro Gly Asn Gly Gly Leu Ser Gly
```

Table I. Amino Acid Sequence Identity between Type IV Collagen Chains

	Ascaris α2	human α2	*C. elegans* α1	human α1
C. elegans α2				
7S domain	88%	53%	50%	50%
NC1 domain	97%	72%	60%	66%
GLY-X-Y interruptions	83%	<10%	21%	<10%
Entire sequence	74%	48%	50%	50%

The amino acid sequences of the major domains of *C. elegans* α2(IV) collagen were compared with those of the *C. elegans* α1(IV), *Ascaris suum* α2(IV), and the human α1(IV) and α2(IV) chains. Alignments were performed using the PALIGN program (IntelliGenetics, Mountain View, CA). The number of identical amino acids is expressed as a percentage of the total number of amino acids compared in each domain. The human, *C. elegans*, and *Ascaris* sequences were taken from the sources referenced in the text.

Figure 4. Exon/intron structure of the exon 8 to 11 region of the *C. elegans* α2(IV) collagen gene. The complete sequence of exon 11 is not shown. Numbering of nucleotides designates the transcription start site as +1. Exon sequences are capitalized; intron sequences are lower case. Interruptions in the Gly-X-Y sequence and the unusual splice donors of introns 9 and 10 are shown in bold type. Primers used in the RT-PCR are underlined. Restriction sites for BglII (exon 9) and DdeI (exon 10) are indicated. Note that for splicing to exon 11, exon 9 must use the unusual splice donor GTACT, while exon 10 must use the unusual splice donor GCAAG.

α1(IV) and α2(IV) chains, nearly all the interruptions in the Gly-X-Y sequence are in the same (or very close) positions, and those interruptions in the same position tend to be the same length. In both chains, most interruptions are located in the NH_2-terminal half of the Gly-X-Y repeat domain: 10/15 for the α2(IV) chain, and 12/17 for the α1(IV) chain. Despite the conservation of interruption location and length in the α1(IV) and α2(IV) collagen chains of *C. elegans*, the amino acid sequence identity between corresponding interruptions is very low (Table I).

Alternative Splicing of the C. elegans α2(IV) Collagen Transcript

The *C. elegans* α2(IV) gene has an unusual intron/exon structure in the region spanning exons 8 through 11 (Fig. 2). Fig. 4 shows the complete nucleotide sequence of this region of the gene. The putative exons 9 and 10 are separated by a putative intron that is only 30-bp long. This length is considered to be too short for a functional intron in *C. elegans* (Blumenthal and Thomas, 1988). The 5'-splice donor sequences of introns 9 and 10 are also unusual: the intron 9 donor has a CT in place of the consensus AG and the intron 10 donor has a C in place of the normally invariant T in the second position. Exons 9 and 10 appear to be duplicates, having the general structure $(Gly\text{-}X\text{-}Y)_5-9$ or 10 amino acid interruption$-(Gly\text{-}X\text{-}Y)_4$. The interruption is nine amino acids long in exon 9, and 10 amino acids long in exon 10. This unusual intron/exon structure indicated that these exons could be spliced in some unconventional manner.

To investigate the splicing of exons 9 and 10, we performed reverse transcriptase (RT)[1]-PCR on RNA prepared from mixed populations of *C. elegans*. The primers used were located in exons 8 and 11, flanking the region of interest (Fig. 4). Fig. 5 shows the possible RT-PCR products that could be obtained using these primers. If the spliced transcript contained both exons 9 and 10, an RT-PCR product 312-bp long would be produced. On the other hand, RT-PCR products of 204 or 201 bp would be produced if either exon 9 or exon 10, respectively, were missing from the spliced transcript (Fig. 5). The resulting RT-PCR product was slightly above 200-bp long, large enough to contain either exon 9 or exon 10, but not both exons (Fig. 6, lane 2). To determine which exon(s) were present in the RT-PCR product, we used restriction sites unique to each exon, BglII in exon 9, and DdeI in exon 10. The RT-PCR product is digested to some extent by both enzymes (Fig. 6), indicating that both exons 9 and 10 are present in two separate transcripts in this mixed population RNA. To confirm these results, RT-PCR products were cloned and selected by hybridization to oligonucleotide probes specific for either exon 9 or 10. The sequence of the clones was as predicted in options B and C of Fig. 5, i.e., the clones hybridizing to the exon 9–specific probe contained exons 8, 9, and 11 (option B), while the clones hybridizing to the exon 10-specific probe contained exons 8, 10, and 11 (option C). As a result of this alternative splicing, the triple helical region of the *C. elegans* α2(IV) chain will be exactly the same size as the triple helical region of the *C. elegans* α1(IV) chain (1,487 amino acids) if exon 10 is used, or one amino acid shorter than that of the α1(IV) chain if exon 9 is used.

Developmental Regulation of Alternative Splicing of the C. elegans α2(IV) Collagen Transcript

The ratio of exon 9 to exon 10 in an RNA sample can be estimated by RT-PCR using the primers shown in Fig. 4. These primers flank exons 9 and 10, so both exons are amplified to-

1. *Abbreviation used in this paper*: RT, reverse transcriptase.

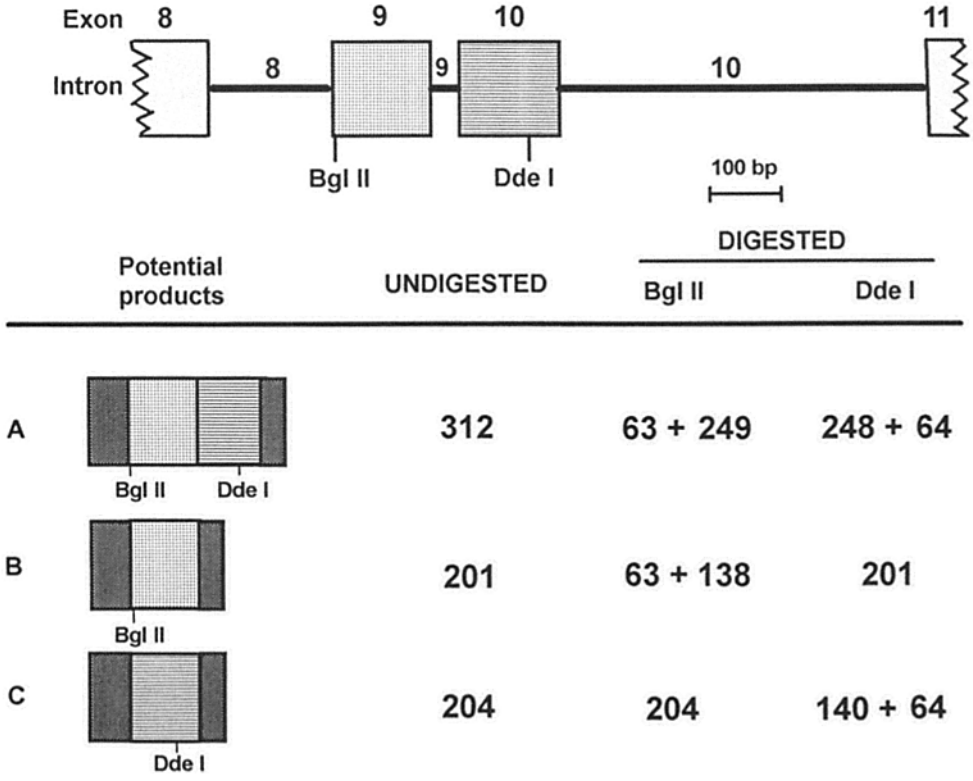

Figure 5. Potential RT-PCR products that could be obtained from the exon 8-11 region of the *C. elegans* α2(IV) collagen gene. Sizes of the potential products in base pairs before and after digestion with BglII and DdeI are indicated. Exons are shown as boxes, introns as horizontal lines.

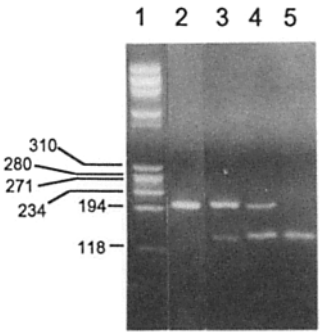

Figure 6. Restriction digests of exon 8-11 region RT-PCR products from mixed population *C. elegans* RNA. Lane *1*, size standards (ϕX174 HaeIII digest); lane *2*, undigested products; lane *3*, digested with BglII (digests exon 9-containing product); lane *4*, digested with DdeI (digests exon 10-containing product); lane *5*, digested with BglII and DdeI.

gether in the same PCR. The relative amounts of exon 9 and exon 10 in the RT-PCR product can then be determined by observing the extent of its digestion by the exon 9-specific enzyme BglII and the exon 10-specific enzyme DdeI. In an RNA sample prepared from a mixed population of animals of all developmental stages (Fig. 6), there are approximately one-half as many exon 9-containing transcripts as exon 10-containing transcripts.

To determine if the proportion of exon 9- to exon 10-containing transcripts changes during development, RNA was isolated from animals at different developmental stages and used as a template for RT-PCR. RNA was isolated from early-stage embryos (0-2-h postfertilization), later-stage embryos (6-8 h postfertilization), L1 larvae (<1-h posthatching, 3-4 h posthatching, and 7-8-h posthatching), L2, L3 and L4 larvae, and adults. Fig. 7 shows autoradiographs of the restriction digests of RT-PCR products from the various developmental stages. It is evident that there is a large change in the exon 9/exon 10 ratio during development in *C. elegans*. Exon 9-containing transcripts appear to predominate in RNA from embryos, while exon 10-containing transcripts predominate in adult RNA. RNA from larval stages appears to have an intermediate ratio of exon 9- to exon 10-containing transcripts. Fig. 8 shows the percentage of the RT-PCR product present in exon 9-containing transcripts at the different developmental stages. Relative levels of exon 9-containing transcripts are highest in embryo RNA (~90% exon 9) and lowest in adult RNA (~10% exon 9). There is a steep decline in the ratio of exon 9 to exon 10 between the late stages of embryogenesis (84% exon 9) and the early part of the first larval (L1) stage (29% exon 9). This decline in the ratio of exon 9 to exon 10 continues more gradually throughout larval development to the adult stage (Fig. 8).

It is possible that adult tissues have no exon 9-containing transcripts, and that the 10% exon 9 detected in adult RNA is contributed by the embryos they contain. We investigated this possibility by extracting RNA from the temperature-sensitive *fer-1* mutant strain, which at 25°C fails to produce embryos due to a sperm defect (Ward and Miwa, 1978). It would be expected that if embryos contribute substantially to the proportion of exon 9 observed in N2 adults, RNA extracted from *fer-1* adults grown at 25°C would have an extremely low or undetectable proportion of exon 9-containing transcripts. However, *fer-1* adults grown at 25°C have an exon 9/exon 10 ratio of ~19% (Fig. 8), as great or greater than that observed in N2 adults. This indicates that embryo RNA does not contribute significantly to the proportion of exon 9-containing transcripts detected in N2 adults, and, therefore, exon 9-containing transcripts are expressed in adult tissues.

Discussion

We have shown that the genetic locus encoding the *C. elegans* α2(IV) collagen chain is *let-2* on the X chromosome. Mutations in *let-2* and in the *C. elegans* α1(IV) collagen gene *emb-9* cause similar phenotypes. Their most severe effect is embryonic lethality at the twofold stage of embryogenesis, with similar morphological abnormalities preceding lethality. This similarity in phenotype between *let-2* and *emb-9* is consistent with the interaction of their products in type IV collagen molecules and supramolecular aggregates in the basement membrane. A total of 24 mutant alleles of *emb-9* and *let-2* have been isolated (Wood et al., 1980; Miwa et al., 1980; Meneely and Herman, 1979; Meneely and Herman, 1981; Cassada et al., 1981), providing a foundation for genetic analysis of basement membrane function and assembly in *C. elegans*.

With the completion of the amino acid sequence of the *C. elegans* α2(IV) chain, the first non-mammalian α1/α2(IV) collagen pair can now be examined. At present, the only other α1/α2(IV) pairs that have been completely sequenced are those of human and mouse (Brazel et al., 1987; Hostikka and Tryggvason, 1988; Muthukumaran et al., 1989; Saus et

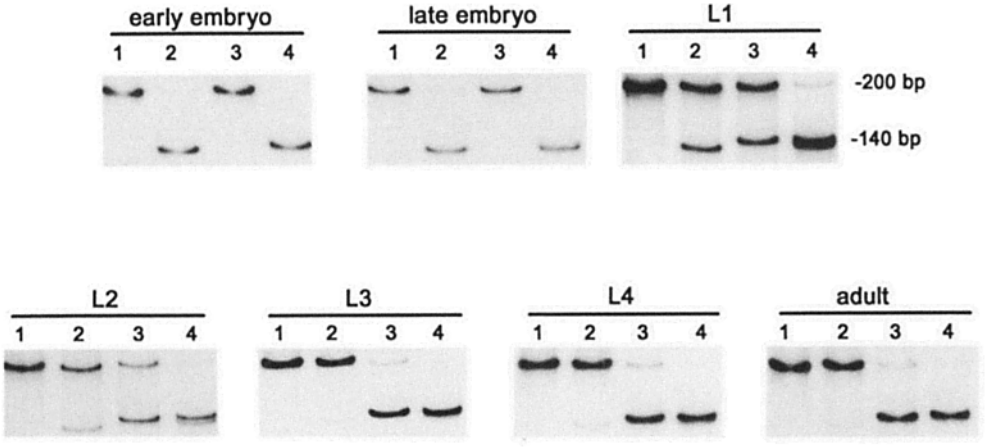

Figure 7. Autoradiographs of exon 8-11 region RT-PCR products from *C. elegans* RNA prepared at different developmental stages. Lane *1*, undigested; lane *2*, digested with BglII (digests exon 9-containing product); lane *3*, digested with DdeI (digests exon 10-containing product); lane *4*, digested with BglII and DdeI. Developmental stage is indicated above each panel. The L1 sample shown is from 7-8 h L1 larval RNA.

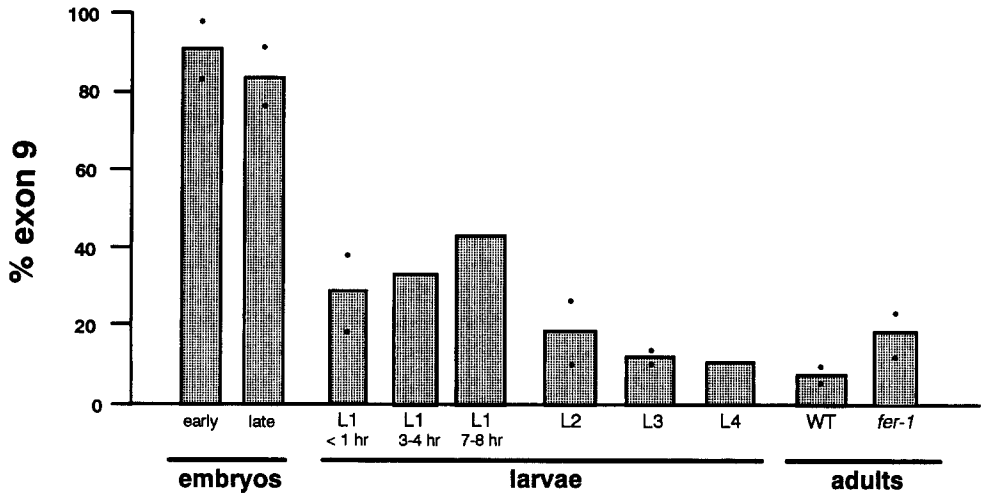

Figure 8. Determination of the relative amount of exon 9 in exon 8–11 region RT-PCR products from developmentally staged *C. elegans* RNA. RNA isolations were generally performed twice for each developmental stage, and these separately determined values are indicated by dots. *WT*, wild-type.

al., 1989). In addition, complete sequences of the following type IV chains are known: human α5 (Zhou et al., 1992), *Drosophila* α1 (Blumberg et al., 1988), sea urchin α1 (Exposito et al., 1993), and *Ascaris suum* α2 (Pettitt and Kingston, 1991). In both the *C. elegans* and mammalian α1/α2(IV) collagen pairs nearly all interruptions in the Gly-X-Y are in the same positions. In both pairs, interruptions in the same positions tend to be the same length, but have very low amino acid sequence identity. Conservation of the location and length of interruptions between the α1(IV) and α2(IV) chains indicates that areas of potential flexibility would correspond. Presumably, differences in amino acid sequence would not significantly affect flexibility of the chain, so there would be little pressure to conserve these sequences between α1(IV) and α2(IV) chains.

The sequences of interruptions in a given α chain, however, tend to be highly conserved between species in the same phylum. In α2(IV) interruptions, for example, there is 70% amino acid sequence identity between human and mouse and 83% identity between *C. elegans* and *Ascaris*. This degree of sequence identity is greater than that for the X and Y positions of the Gly-X-Y repeat domain, suggesting that there are greater functional constraints on interruption sequences than on X-Y sequences. Certain Gly-X-Y interruptions have especially low amino acid sequence identity between α2(IV) chains of species in the same phylum. These interruptions reveal regions of relatively rapid change in the α2(IV) chain. For example, interruption #10 in the *C. elegans* chain has only 36% sequence identity with the *Ascaris* α2(IV) chain. This interruption corresponds in position to interruption #13 in the human and mouse α2(IV) chain, which is a 24-amino acid interruption having a disulfide-linked loop that is thought to protrude out of the molecule (Brazel et al., 1988). Such a loop is completely absent from both nematode α2(IV) chains, indicating that this may be a region of rapid change in α2(IV) collagen chains.

We have shown that the *C. elegans* α2(IV) collagen mRNA is alternatively spliced, and that embryos express predominantly exon 9–containing transcripts while larvae and adults express predominantly exon 10–containing transcripts. The greatest shift in the ratio of exon 9– to exon 10–containing transcripts occurs between late embryogenesis and early in the first larval (L1) stage, when the proportion of exon 9–containing transcripts drops from 84 to 29%. This shift in the exon 9/exon 10 ratio coincides with a dramatic shift in *C. elegans* development, from a phase of rapid and extensive morphological change (embryogenesis) to a phase that primarily entails symmetrical growth (early larval through adult stages). During embryogenesis in *C. elegans*, the embryo is transformed in only a 5 to 6 h period from a spherical ball of cells to an elongated cylindrical worm, with an increase in length of about fourfold. After this period of rapid morphogenesis, the larval and adult stages of *C. elegans* development are essentially growth stages, characterized by symmetrical enlargement of tissues previously established during embryogenesis. The morphological changes of embryogenesis may require basement membranes with different properties than those present in larvae or adults. In embryos, modifications in type IV collagen cross-linking, glycosylation or in interactions with other basement membrane components may be needed to allow rapid cell movement and changes in cell shape. In the mouse, embryonic basement membranes initially lack type IV collagen and have polymerized laminin as their major structural component (Lievo et al., 1980). Thus, mammals can also have embryonic basement membranes that appear to be structurally and functionally quite different from adult basement membranes. During the shift from embryonic to larval development in *C. elegans*, a more fluid or flexible form of α2(IV) collagen (exon 9–containing) may be replaced by a more permanent form (exon 10–containing).

It is possible that exons 9 and 10 contain different binding or cross-linking sites that could change basement membrane characteristics during development. When the EMBL protein data base was searched using the interruptions of exons 9 and 10, no significant homology to any known sequence emerged. However, the interruption of exon 9 contains a potential glycosaminoglycan addition site Ser-Gly (Hardingham and Fosang, 1992); this site is lacking in the exon 10 interruption. The sequence just NH_2-terminal to the Ser-Gly in exon 9 is Glu-Phe-Thr-Gly, which conforms to previous observations that Ser-Gly attachment sites for glycosaminoglycans are in close association with hydrophobic and acidic residues (Doege et al., 1987; Noonan et al., 1988).

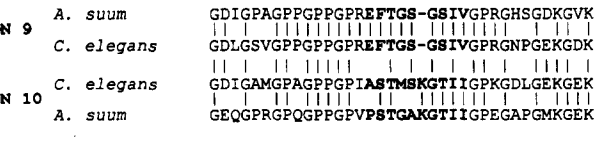

Figure 9. Alignment of the amino acid sequences of the alternatively spliced exons 9 and 10 of *Ascaris suum* and *C. elegans* α2(IV) collagen. Sequence of the *A. suum* exons is from Pettitt and Kingston (1991). Interruptions in the Gly-X-Y repeating sequence are shown in bold type. Identical amino acids are indicated by vertical bars.

Exon 10 contains a lysine residue in the central portion of its interruption which may provide an additional glycosylation site, or may be used in cross-linking.

Transcripts of the *Ascaris suum* α2(IV) collagen gene also appear to be alternatively spliced at exactly the same sites as *C. elegans* α2(IV) collagen (Pettitt and Kingston, 1991; I. B. Kingston, personal communication). An alignment of the amino acid sequences of exons 9 and 10 of these two nematodes is shown in Fig. 9. Exon 9, the embryonic exon in *C. elegans*, is more highly conserved between the two nematodes (81% amino acid identity) than is exon 10 (65% identity). The interruption sequence of exon 9, including the potential glycosaminoglycan attachment sequence, is 100% conserved between the two nematodes. In contrast, the interruption sequence of exon 10 is only 70% conserved. It is possible that the expression of the two exons is also developmentally regulated in *Ascaris*. The alternative splicing of exons 9 and 10 is yet another common feature between these highly similar nematode α2(IV) collagen chains; alternative splicing has not been observed in type IV collagen from any other organism. The fact that alternative splicing of α2(IV) collagen has been maintained between these two distantly related nematodes suggests that it may have an important role in basement membrane function.

The unusual splice donor sites seen in both exons 9 and 10 may be involved in the regulation of alternative splicing of the α2(IV) collagen transcript. It is remarkable that the same unusual GCAAG splice donor site seen at the 5′ end of intron 10 is found in the human α1(IV) collagen gene at the 5′ end of intron 34 (Soininen et al., 1989). Exon 34 is near the COOH-terminal end of the triple helix, about 700 amino acids away from the alternatively spliced region in the *C. elegans* α2(IV) chain. Although the locations of the GCAAG splice donors do not correspond between the human α1(IV) and the *C. elegans* α2(IV) chains, the occurrence of this unusual splice donor in two type IV collagen genes is, if nothing more, an extraordinary coincidence. Shapiro and Senapathy (1987) report that this sequence was seen only five times in 1,893 published non-immunoglobulin gene sequences. They propose that genes containing a GCAAG splice donor may have a role in regulating genetic pathways of cell growth and differentiation. Our observations are consistent with this proposal: we find that the transcript of a collagen gene containing this splice site is alternatively spliced in a developmentally regulated manner.

Our identification of *let-2* as the genetic locus of the *C. elegans* α2(IV) collagen gene allows a systematic study of type IV collagen function in vivo. Through the efforts of several laboratories, 19 mutant alleles of *let-2* have been isolated (Meneely and Herman, 1979; Meneely and Herman, 1981; Cassada et al., 1981). These 19 alleles display a broad range of phenotypes. The *let-2* alleles with the mildest mutant phenotype are viable and fertile at 15° and 20°C, and embryonic lethal at 25°C. There are several *let-2* alleles having phenotypes of intermediate severity, including some alleles that are viable and fertile at 15°C and embryonic lethal at both 20° and 25°C. Some *let-2* alleles are only slightly fertile at 15°C, while the most severe alleles are embryonic lethal at 15°, 20°, and 25°C. The diversity of *let-2* phenotypes contrasts with the similar phenotypes seen in mutants of *emb-9*, the gene encoding the *C. elegans* α1(IV) collagen chain. All five *emb-9* alleles are wild-type at 15°C, larval lethal at 20°C, and embryonic lethal at 25°C. A broader spectrum of phenotypes may be revealed when additional alleles of *emb-9* are isolated. We have identified the precise mutations in the α2(IV) collagen chain in seventeen *let-2* alleles (M. Sibley, P. Graham, and J. Kramer, manuscript in preparation), and are currently examining the effects of these mutations on type IV collagen synthesis and assembly. Unfortunately, none of these mutations are within the alternatively spliced region of the gene. It will be necessary to generate specific mutations in the alternatively spliced exons to genetically analyze their function.

The sequence data reported here are available from EMBL/GenBank/DDJB under accession number Z22964.

This work was supported by a grant from the National Institutes of Health (HD27211 to J. M. K.). The *fer-1* strain BA1 was provided by the *Caenorhabditis* Genetic Center, which is supported by the NIH Division of Research Resources.

Received for publication 17 April 1993 and in revised form 16 June 1993.

References

Bektesh, S., K. Van Doren, and D. Hirsh. 1988. Presence of the *Caenorhabditis elegans* spliced leader on different mRNAs and in different genera of nematodes. *Genes Dev.* 2:1277–1283.

Blumberg, B., A. J. MacKrell, and J. H. Fessler. 1988. *Drosophila* basement membrane procollagen α1(IV). II. Complete cDNA sequence, genomic structure, and general implications for supramolecular assemblies. *J. Biol. Chem.* 263:18328–18337.

Blumenthal, T., and J. Thomas. 1988. *Cis* and *trans* splicing in *C. elegans*. *Trends Genet.* 4:305–308.

Brazel, D., I. Oberbaumer, H. Dieringer, W. Bable, R. W. Glanville, R. Deutzmann, and K. Kuhn. 1987. Completion of the amino acid sequence of the α1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain. *Eur. J. Biochem.* 168:529–536.

Brazel, D., R. Pollner, I. Oberbaumer, and K. Kuhn. 1988. Human basement membrane collagen (type IV). The amino acid sequence of the α2(IV) chain and its comparison with the α1(IV) chain reveals deletions in the α1(IV) chain. *Eur. J. Biochem.* 172:35–42.

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

Burbelo, P. D., G. R. Martin, and Y. Yamada. 1988. α1(IV) and α2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer. *Proc. Natl. Acad. Sci. USA.* 85:9679–9682.

Cassada, R., E. Isnenghi, M. Culotti, and G. von Ehrenstein. 1981. Genetic analysis of temperature-sensitive embryogenesis mutants in *Caenorhabditis elegans*. *Dev. Biol.* 84:193–205.

Charonis, A. S., and E. C. Tsilibary. 1990. Assembly of basement membrane proteins. *In* Organization and Assembly of Plant and Animal Extracellular Matrix, W. S. Adair and R. P. Mecham, editors. Academic Press, New York. 85–117.

Cutting, G. R., H. H. Kazazian, S. E. Antonorakis, P. D. Killen, Y. Yamada, and C. A. Francomano. 1988. Macrorestriction mapping of COL4A1 and COL4A2 collagen genes on human chromosome 13q34. *Genomics.* 3:256–263.

Doege, K., M. Sasaki, E. Horigan, J. R. Hassell, and Y. Yamada. 1987. Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones. *J. Biol. Chem.* 262:17757–17767.

Exposito, J.-Y., M. D'Alessio, M. DiLiberto, and F. Ramirez. 1993. Complete primary structure of a sea urchin type IV collagen α chain and analysis of

the 5'-end of its gene. *J. Biol. Chem.* 268:5249–5254.
Farquhar, M. G. 1991. The glomerular basement membrane: A selective macromolecular filter. *In* Cell Biology of Extracellular Matrix. E. D. Hay, editor. Plenum Press, New York. 365–418.
Glanville, R. W. 1987. Type IV collagen. *In* Structure and function of collagen types, R. Mayne and R. E. Burgeson, editors. Academic Press, New York. 43–79.
Gunwar, S., J. Saus, M. E. Noelken, and B. G. Hudson. 1990. Glomerular basement membrane. Identification of a fourth chain, α4, of type IV collagen. *J. Biol. Chem.* 265:5466–5469.
Guo, X., and J. M. Kramer. 1989. The two *Caenorhabditis elegans* basement membrane (type IV) collagen genes are located on separate chromosomes. *J. Biol. Chem.* 264:17574–17582.
Guo, X., J. J. Johnson, and J. M. Kramer. 1991. Embryonic lethality caused by mutations in basement membrane collagen of *C. elegans*. *Nature (Lond.).* 349:707–709.
Hardingham, T. E., and A. J. Fosang. 1992. Proteoglycans: many forms and many functions. *FASEB (Fed. Am. Soc. Exp. Biol.) J.* 6:861–870.
Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted break points for DNA sequencing. *Gene.* 28:351–361.
Hostikka, S. L., and K. Tryggvason. 1987. Extensive structural differences between genes for the α1 and α2 chains of type IV collagen despite conservation of coding sequences. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 224: 297–305.
Hostikka, S. L., and K. Tryggvason. 1988. The complete primary structure of the α2 chain of human type IV collagen and comparison with the α1(IV) chain. *J. Biol. Chem.* 263:19488–19493.
Hostikka, S. L., R. L. Eddy, M. G. Byers, M. Hoyhyta, T. B. Shows, and K. Tryggvason. 1990. Identification of a distinct type IV collagen α chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome. *Proc. Natl. Acad. Sci. USA.* 87:1606–1610.
Isnenghi, E., K. Cassada, C. Smith, K. Denitch, A. Radnia, and G. von Ehrenstein. 1983. Maternal effects and temperature-sensitive period of mutations affecting embryogenesis in *Caenorhabditis elegans*. *Dev. Biol.* 98:465–480.
Kaytes, P., L. Wood, N. Theriault, M. Kurkinen, and G. Voegli. 1988. Head-to-head arrangement of murine type IV collagen genes. *J. Biol. Chem.* 263:19274–19277.
Kramer, J. M., R. P. French, E.-C. Park, and J. J. Johnson. 1990. The *Caenorhabditis elegans rol-6* gene, which interacts with the *sqt-1* collagen gene to determine organismal morphology, encodes a collagen. *Mol. Cell Biol.* 10:2081–2089.
Kuivaniemei, H., G. Tromp, and D. Prockop. 1991. Mutations in collagen genes: causes of rare and some common diseases of humans. *FASEB (Fed. Am. Soc. Exp. Biol.) J.* 5:2052–2060.
Lievo, I., A. Vaheri, R. Timpl, and J. Wartiovaara. 1980. Appearance and disappearance of collagens and laminin in the early mouse embryo. *Dev. Biol.* 76:100–114.
Mello, C. C., J. M. Kramer, D. T. Stinchcomb, and V. Ambros. 1992. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO (Eur. Mol. Biol. Organ.) J.* 10: 3959–3970.
Meneely, P. M., and R. K. Herman. 1979. Lethals, steriles, and deficiencies in a region of the X chromosome of *Caenorhabditis elegans*. *Genetics.* 92:99–115.
Meneely, P. M., and R. K. Herman. 1981. Suppression and function of X-linked lethal and sterile mutations in *Caenorhabditis elegans*. *Genetics.* 97:65–84.
Miwa, J., E. Schierenberg, S. Miwa, and G. von Ehrenstein. 1980. Genetics and mode of expression of temperature-sensitive mutations arresting embryonic development in *Caenorhabditis elegans*. *Dev. Biol.* 76:160–174.
Morrison, K. E., G. G Germino, and S. T. Reeders. 1991. Use of the polymerase chain reaction to clone and sequence a cDNA encoding the bovine α3 chain of type IV collagen. *J. Biol Chem.* 266:34–39.
Muthukumaran, G., B. Blumberg, and M. Kurkinen. 1989. The complete primary structure for the α1 chain of mouse collagen IV. Differential evolution of collagen IV domains. *J. Biol. Chem.* 264:6310–6317.
Noonan, D. M., E. A. Horigan, S. R. Ledbetter, G. Vogeli, M. Sasaki, Y. Yamada, and J. R. Hassell. 1988. Identification of cDNA clones encoding different domains of the basement membrane heparan sulfate proteoglycan. *J. Biol. Chem.* 263:16379–16387.
Pettitt, J., and I. B. Kingston. 1991. The complete primary structure of a nematode α2(IV) collagen and the partial structural organization of its gene. *J. Biol. Chem.* 266:16149–16156.
Poschl, E., R. Pollner, and K. Kuhn. 1988. The genes for the α1(IV) and α2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:2687–2695.
Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad Sci. USA.* 12:5463–5467.
Saus, J., J. Wieslander, J. P. M. Langeveld, S. Quinones, and B. G. Hudson. 1988. Identification of the Goodpasture antigen as the α3(IV) chain of collagen IV. *J. Biol. Chem.* 263:13374–13380.
Saus, J., S. Quinones, A. MacKrell, B. Blumberg, G. Muthukumaran, T. Pihlajaniemi, and M. Kurkinen. 1989. The complete primary structure of mouse α2(IV) collagen. Alignment with mouse α1(IV) collagen. *J. Biol. Chem.* 264:6318–6324.
Shapiro, M. B., and P. Senapathy. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15:7155–7174.
Soininen, R., M. Huotari, S. L. Hostikka, D. J. Prockop, and K. Tryggvason. 1988. The structural genes for α1 and α2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region. *J. Biol. Chem.* 263:17217–17220.
Soininen, R., M. Huotari, A. Ganguly, D. J. Prockop, and K. Tryggvason. 1989. Structural organization of the gene for the α1 chain of human type IV collagen. *J. Biol Chem.* 264:13565–13571.
Timpl, R. 1989. Structure and biological activity of basement membrane proteins. *Eur. J. Biochem.* 180:487–502.
Ward, S., and J. Miwa. 1978. Characterization of temperature-sensitive, fertilization-defective mutants of the nematode *Caenorhabditis elegans*. *Genetics.* 88:285–302.
Wood, W. B. 1988. The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 589 pp.
Wood, W. B., R. Hecht, S. Carr, R. Vanderslice, N. Wolf, and D. Hirsh. 1980. Parental effects and phenotypic characterization of mutants that affect early development in *Caenorhabditis elegans*. *Dev. Biol.* 74:446–469.
Yurchenco, P. D., and H. Furthmayr. 1984. Assembly of basement membrane collagen. *Biochemistry.* 23:1839–1850.
Yurchenco, P. D., and G. C. Ruben. 1987. Basement membrane structure *in situ*: evidence for lateral associations in the type IV collagen network. *J. Cell Biol.* 105:2559–2568.
Zhou, J., J. M. Hertz, A. Leinonen, and K. Tryggvason. 1992. Complete amino acid sequence of the human α5(IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient. *J. Biol. Chem.* 267:12475–12481.