MOLECULAR AND CELLULAR BIOLOGY, Apr. 1997, p. 2301–2311 Vol. 17, No. 4
0270-7306/97/$04.00+0

cis Regulatory Requirements for Hypodermal Cell-Specific Expression of the *Caenorhabditis elegans* Cuticle Collagen Gene *dpy-7*

JOHN S. GILLEARD,* J. DAVID BARRY, AND IAIN L. JOHNSTONE

Wellcome Unit of Molecular Parasitology, Anderson College, University of Glasgow, Glasgow G11 6NU, United Kingdom

Received 23 September 1996/Returned for modification 22 November 1996/Accepted 7 January 1997

The *Caenorhabditis elegans* cuticle collagens are encoded by a multigene family of between 50 and 100 members and are the major component of the nematode cuticular exoskeleton. They are synthesized in the hypodermis prior to secretion and incorporation into the cuticle and exhibit complex patterns of spatial and temporal expression. We have investigated the *cis* regulatory requirements for tissue- and stage-specific expression of the cuticle collagen gene *dpy-7* and have identified a compact regulatory element which is sufficient to specify hypodermal cell reporter gene expression. This element appears to be a true tissue-specific promoter element, since it encompasses the *dpy-7* transcription initiation sites and functions in an orientation-dependent manner. We have also shown, by interspecies transformation experiments, that the *dpy-7 cis* regulatory elements are functionally conserved between *C. elegans* and *C. briggsae*, and comparative sequence analysis supports the importance of the regulatory sequence that we have identified by reporter gene analysis. All of our data suggest that the spatial expression of the *dpy-7* cuticle collagen gene is established essentially by a small tissue-specific promoter element and does not require upstream activator or repressor elements. In addition, we have found the DPY-7 polypeptide is very highly conserved between the two species and that the *C. briggsae* polypeptide can function appropriately within the *C. elegans* cuticle. This finding suggests a remarkably high level of conservation of individual cuticle components, and their interactions, between these two nematode species.

Caenorhabditis elegans produces a series of five distinct cuticles, the first of which is secreted by the underlying hypodermis during late embryogenesis and is subsequently shed and replaced at each of the four larval molts. The nematode hypodermis is a complex tissue comprising the external epithelial cells of the organism, and many of these cells are multinucleate syncytia arising from cell fusions which occur during development. The largest of these is hyp-7, which forms the outer surface of the main body region. In the newly hatched L1 larva, hyp-7 contains 23 nuclei, which increase in number throughout development to 133 in the adult as a result of fusions with the progeny of lateral (seam cells [H1, H2, and V cells]) and ventral (P cells) hypodermal blast cells (38, 39). Other hypodermal cells form the outer epithelium of the head and tail (38, 39).

The major components of the cuticle are the cuticle collagens which are encoded by a multigene family of between 50 and 100 members (12, 17). Considerable developmental complexity exists in the formation of nematode cuticles. First, although the cuticles of each developmental stage are basically similar, some morphological and biochemical differences exist, possibly reflecting functional specialization (5). Second, there is complexity in the expression patterns of different cuticle collagen genes. Examples exist of differential expression of individual cuticle collagen genes between different developmental stages (4, 20, 24). Also, recently we have shown an additional level of complexity of cuticle collagen gene expression. Within each developmental stage, individual collagen genes are expressed at different times, resulting in temporally distinct waves of expression during cuticle synthesis. The precise pattern of these waves of gene expression is repeated during each larval stage (13). We are interested in understanding how the spatial and temporal regulation of this large multigene family is achieved.

Detailed analysis of *cis* regulatory elements has been performed for only a small number of *C. elegans* genes. However, as with other multicellular organisms, it appears that a wide spectrum of regulatory mechanisms exists. Some genes, such as the myosin (31) and vitellogenin (26) genes, are regulated in a relatively simple fashion by a tissue-specific basal promoter whose activity is enhanced by separate activator elements which can lie either 5′ or 3′ to the promoter or within an intronic sequence. Other genes such as *ges-1* (1), *mec-3* (41), and *hlh-1* (21) are regulated in a more complex fashion in which a series of both activator and repressor elements is required to establish properly regulated expression. We have examined the *cis* regulatory requirements for spatial and stage-specific expression of the cuticle collagen gene *dpy-7*. We have chosen to study *dpy-7* since it is one of a number of cuticle collagen genes which produce severe alterations in organismal morphology when mutated (14). This observation suggests that *dpy-7* performs an important structural role during *C. elegans* development, and since recessive mutant alleles exist, rescue of the mutant phenotype by transformation with cloned copies of the gene can be used as an assay of expression. In addition, of the genes that we have examined to date, *dpy-7* is expressed earliest within the reiterated temporal series of waves of cuticle collagen gene expression (13).

In this report, we present a detailed description of the tissue- and stage-specific pattern of reporter gene expression conferred by *dpy-7 cis* regulatory sequences. We have located the

* Corresponding author. Mailing address: Wellcome Unit of Molecular Parasitology, University of Glasgow, Anderson College, 56 Dumbarton Rd., Glasgow G11 6NU, United Kingdom. Phone: 0141 339 8866, ext. 2218. Fax: 0141 330 5422. E-mail: gbga48@udcf.gla.ac.uk.

transcription initiation sites of *dpy-7*, which is a *trans*-spliced gene, and found that sequences immediately surrounding these sites are sufficient to direct hypodermal cell-specific expression. We have also used interspecies transformation experiments to show that the regulation of the *dpy-7* gene and function of the DPY-7 polypeptide are conserved between *C. elegans* and *C. briggsae*, a related nematode of estimated 40 million years' divergence (15). Comparative sequence analysis, reporter gene experiments, and repair of phenotype experiments all suggest that the tissue- and stage-specific regulation of *dpy-7* is achieved essentially by a compact regulatory region encompassing the basal promoter.

MATERIALS AND METHODS

Nematode culture. *C. briggsae* and *C. elegans* strains were cultured as described elsewhere (37). *C. briggsae* wild-type strain G16 was used for preparation of genomic DNA and transformation experiments. The *C. elegans* strains used were the wild-type Bristol strain N2 and the *dpy-7* mutant strain CB88 [*dpy-7*(*e88*)X] (14). Nematode genomic DNA was prepared by using standard methods (37). L1 larvae were obtained for RNA preparation by hatching embryos, previously collected by hypochlorite treatment of gravid hermaphrodites, in M9 buffer (37).

RT-PCR and 5′ RACE. RNA preparation and reverse transcription (RT)-PCR were performed as previously described (13, 23). The sequences of the PCR primers used in these experiments are as follows: 569, 5′GGGCATCGTAGTT GAACACAAGATGTGC3′; 256765, 5′GCCCAGTTCGGGGGCCAACCACG T3′; 204014, 5′TTGGTTGTCGGATTGAGCACTATC3′; 87390, 5′GTTGCGA GCAACGAACTTGAAGGAGCCC3′; 28090, 5′GGCTTCTCCATTTATCTG GAACAAAATGTAAG3′; SL1, 5′GGTTTAATTACCCAAGTTTGAG3′; SL2, 5′GGTTTTAACCCAGTTACTCAAG3′; cyp9RT, 5′GCTCTAGATCATTTCA TGCTTCTTCTCATTG3′; cyp9REVT, 5′GCTCTAGATCACACGAGTTCTC CACAATTGGAG3′; 247499, 5′CTGTACACATTTCAGAAGGACGCTCG3′; anchor primer, 5′CUACUACUACUAGGCCACGCGTCGACTAGTACGGG IIGGGIIGGGIIG3′; and universal amplification primer, 5′CUACUACUACU AGGCCACGCGTCGACTAGTAC3′.

The 5′RACE (rapid amplification of 5′ cDNA ends) system (GIBCO-BRL) was used to identify the 5′ ends of *dpy-7* primary transcripts. First-strand cDNA was synthesized by using antisense *dpy-7*-specific primer 569, which anneals within the second exon. Following addition of a homopolymer tail, a 1:100 dilution of first-strand cDNA was used as the template for a first-round PCR designed to amplify both prespliced and mature RNA. Thirty amplification cycles were performed with a second antisense *dpy-7*-specific primer, 204014 (anneals within the second exon but is nested with respect to primer 569), and the anchor primer (anneals to homopolymer tail of first-strand cDNA). A 1:100 dilution of the resulting first-round PCR was used as the template for a number of separate 30-cycle second-round PCRs. Reaction 1 used primers SL1 (sense sequence of *C. elegans* SL1 spliced leader) and 87390 (antisense of cDNA sequence spanning the *dpy-7* intron/exon junctions), and reaction 2 used the universal amplification primer (shorter version of the anchor primer) and 87390. These two reactions were designed to be specific for the mature transcript. Reaction 3 used the universal amplification primer and 28090 (antisense primer spanning the *dpy-7* SL1 splice acceptor site) and was designed to be specific for the primary transcript. All sets of PCRs were performed in parallel with identical reactions using a mock cDNA synthesis reaction, from which reverse transcriptase had been omitted, as the template for the first round of PCR. Products from the second-round reaction 1, performed with ^{32}P end-labelled primers 87390 and 28090, were also analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. The sizes of the products were estimated by comparison to products of a dideoxynucleotide chain termination sequencing reaction, using primer 28090 on pdpy7 as the template, which were run in adjacent tracks. End labelling of primers was performed by phosphorylating with T4 kinase in the presence of [α-^{32}P]ATP (32).

The 230- to 250-bp products from reaction 3 were gel purified from two independently performed first- and second-round PCRs and, after end filling with Klenow polymerase and restriction digestion with *Sal*I (the universal amplification primer contains a *Sal*I site), were independently subcloned into a *Sal*I/*Sma*I-double-digested pBluescript KS+ vector. Ten clones derived from each of the two independently cloned PCR products were sequenced to identify the 5′ ends.

Cloning and sequencing of the *C. briggsae dpy-7* homolog. (i) Southern blotting. Restriction enzyme digestion, agarose gel electrophoresis, and Southern blotting to Hybond-N nylon membrane (Amersham) were performed by using standard methods (32). A *dpy-7*-specific fragment (−459 to +474 relative to the initiator ATG) encompassing 933 bp of genomic sequence immediately upstream of the *dpy-7* Gly-X-Y repeats was generated by PCR from plasmid pdpy7 (pBluescript containing the *C. elegans dpy-7* gene), using oligonucleotides 247499 and 569. For use as a probe, the fragment was isolated from an agarose gel and random primer labelled with [γ-^{32}P]dCTP by using the Prime-it II system (Stratagene) and then separated from unincorporated radiolabel by using a Nunc-trap push column (Stratagene). Hybridizations were performed at 50°C in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–5× Denhardt's solution–0.1% sodium dodecyl sulfate with 100 mg of denatured, sonicated herring sperm DNA per ml. Washes were performed to a final stringency of 2× SSC–0.1% sodium dodecyl sulfate at 55°C.

(ii) Isolation of the *C. briggsae dpy-7* homolog. A *C. briggsae* genomic library (35) was screened by using the probe and hybridization conditions described above for the genomic Southern blot. Three positively hybridizing phage clones were identified and found to be indistinguishable following purification, DNA extraction, and restriction enzyme analysis using standard methods (32). A hybridizing 3.1-kb *Eco*RI fragment from one of the clones, CB3, was subcloned into pBluescript KS+ (plasmid pJB1). Exonuclease III deletions were prepared for use as sequencing templates as described previously (32). Sequencing was performed by the dideoxynucleotide chain termination method (33) with Sequenase 2.0 (United States Biochemical) on double-stranded DNA templates. Second-strand sequencing was performed by primer walking using oligonucleotides designed from the first-strand sequence. Sequence manipulation and analysis were performed with the Genetics Computer Group sequence analysis software package (6).

Construction of reporter gene fusions. Details of the construction of plasmid p20.6.90/7, in which *lacZ* was cloned in frame within the *dpy-7* coding sequence, are available on request. Briefly, *lacZ* was excised from vector pPd16-43 (9) and engineered, in frame, into the *Xho*I site present in the *dpy-7* coding sequence of the 3.6-kb *Hin*dIII genomic fragment from plasmid pdpy7 (14). To prevent secretion of DPY-7–β-galactosidase fusion protein from the cell, a synthetic transmembrane domain from vector pPd34-110 (9) was inserted into the *Kpn*I site immediately upstream of the pPd16-43-derived *lacZ*. All other *dpy-7* reporter gene constructs were made by using the *C. elegans lacZ* and green fluorescent protein (GFP) promoterless reporter gene expression vectors, pPd21-28 and pPd95-67, respectively (8a). These plasmids encode a nuclear localization signal at the N terminus of β-galactosidase or GFP, which, by conferring nuclear localization on the reporter molecule, aids in the identification of individual cells (9). Appropriate regions of the *dpy-7* 5′ flanking sequence were amplified from plasmid pdpy7 (14) by PCR using *Pfu* polymerase (Stratagene). For translational fusions, upstream and downstream PCR primers, designed to include *Hin*dIII and *Bam*HI sites, allowed simple in-frame subcloning of *dpy-7* 5′ sequence into the multiple cloning site of the reporter gene plasmids. To avoid inclusion of the *dpy-7* secretory signal peptide in the encoded fusion protein, only the first six amino acids of the *dpy-7* coding sequence were included. The insert for transcriptional fusions pdp7.6 and pdp7.7 was amplified by PCR from pdpy7 and, following the ligation of *Hin*dIII adapters, subcloned into the *Hin*dIII site of pPd21-28. The inserts for the remaining transcriptional fusions were blunt cloned into the *Sma*I site of the 5′ polylinker pPd21-28 following phosphorylation of the appropriate PCR fragment. Plasmid DNA was prepared for transformation of nematodes either by CsCl gradient centrifugation (32) or by using Qiagen-tip 100 (Qiagen). The inserts of all *lacZ* reporter gene constructs were sequenced in order to ensure the absence of PCR-induced mutations.

DNA transformation of nematodes. Transformation of *C. elegans* was performed by microinjection of plasmid DNA into the distal arms of the hermaphrodite gonad (28). The *lacZ* reporter gene constructs, at a final concentration of 20 μg/ml, were coinjected with plasmid pRF4 (18, 28) at a final concentration of 100 μg/ml. Plasmid pRF4, which contains the dominant mutant allele of the gene *rol-6*(*su1006*), produces a visible right roller phenotype in transformed animals, allowing them to be cloned individually. Transmitting lines were established from cloned F_1 rollers and maintained for three generations before staining for β-galactosidase expression. The same method of transformation was used for *C. briggsae* in which pRF4 produces a right-roll phenotype similar to that seen in *C. elegans*. The efficiency of transformation as determined by the number of F_1 rollers was similar to that obtained with *C. elegans*, although the proportion of F_1 rollers which gave rise to stable lines was lower (approximately 1 to 2% for *C. briggsae*, compared with approximately 5 to 10% for *C. elegans*). Plasmid pJB1 containing the wild-type *C. briggsae dpy-7* genomic fragment was microinjected into the distal gonad of hermaphrodites of the *C. elegans* mutant strain CB88 [*dpy-7*(*e88*)X] at a final concentration of 100 μg/ml, and F_1 animals were examined for repair of phenotype. Integration of extrachromosomal arrays was achieved by irradiation of late L4 larvae carrying free arrays by using a gamma ray source (3,800 rads). F_2 animals were cloned, and integrants were identified by scoring for 100% transmission of the roller phenotype.

Fixation and staining of nematodes. The histochemical staining procedure was modified from the method described by Fire et al. (9). Worms were washed off agar plates with M9 buffer, pelleted by centrifugation, resuspended in 100 μl of distilled H_2O, and pipetted onto a slide which had been previously coated with 0.1% poly-L-lysine. A coverslip was overlaid, and the worms were then frozen by placing the slide on an aluminum block packed in dry ice. After 5 min, the coverslip was flipped off with a scalpel blade, and the slide was placed in methanol at −20°C for 5 min followed by acetone at −20°C for 5 min. The preparation was allowed to air dry at room temperature, after which 100 μl of staining solution (9) containing 0.01% 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) was applied, a coverslip was overlaid, and the edges were sealed with nail varnish. The slides were incubated at 25°C until sufficiently stained. Animals were examined for GFP expression by being mounted on 2% agarose pads

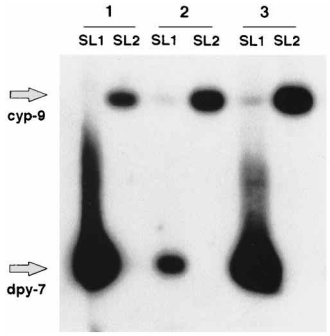

FIG. 1. SL1 *trans* splicing of the *dpy-7* transcript. Gene-specific oligonucleotides for *dpy-7* (*569*) and a control SL2 *trans*-spliced gene, *cyp-9* (cyp9REVT), were used in a single PCR with either an SL1-specific or an SL2-specific oligonucleotide. These reactions were performed on *C. elegans* cDNA prepared from the following developmental stages: 1, L1 larvae 8 h post-L1 block; 2, L1 larvae 12 h post-L1 block; and 3, mixed-stage animals. The number of hours post-L1 block refers to the time elapsed after larvae, previously hatched in M9 buffer, were placed on a bacterial lawn. cDNA synthesis and PCR methodology were as described in reference 13. The products were separated on an agarose gel, Southern blotted, and probed with a nested radiolabelled gene specific oligonucleotide for *dpy-7* (256765). Following autoradiography, the blot was stripped and reprobed with a nested radiolabelled gene-specific oligonucleotide for *cyp-9* (cyp9RT), and autoradiography was again performed. The autoradiographs are shown superimposed, and the *dpy-7* and *cyp-9* products are indicated.

containing 0.01% sodium azide (37) and observed under UV illumination, using a Zeiss Axioplan with filter set 15 (BP546/12, FT580, and LP590).

Immunofluorescence. The monoclonal antibody MH27 specifically recognizes belt desmosomes, allowing the boundaries of hypodermal cells to be visualized (10). If transgenic animals are costained with MH27 and an anti-β-galactosidase antibody, it is possible to distinguish between *lacZ* expression in hyp-7, P-cell, and V-cell nuclei. Animals were fixed by the same method as for X-Gal staining except that following the acetone incubation, samples were rehydrated through an acetone series (80, 60, 40, and 20%) for 60 s each step. Slides were then washed in 1× phosphate-buffered saline (PBS)–0.2% Tween 20 and blocked with 1% nonfat dried milk in PBS–0.2% Tween 20 for 45 min. Primary antibody (MH27 or mouse anti-β-galactosidase diluted 1:300 or 1:2,000, respectively, in the same block buffer) was applied and incubated at room temperature overnight. Following three 10-min washes in 1× PBS–0.2% Tween 20, secondary antibody (goat anti-mouse immunoglobulin G [whole molecule] diluted 1:100 in block buffer) was applied and incubated at 37°C for 3 h. Following three 10-min washes in 1× PBS–0.2% Tween 20, samples were mounted in Aquamount (BDH), a coverslip was overlaid, and the edges were sealed with nail varnish. Specimens were observed under UV illumination as described above.

RESULTS

***dpy-7 trans* splicing and transcription initiation.** Application of the Genefinder program to the *dpy-7* gene predicts another gene just 346 bp upstream of, and on the same strand as, *dpy-7* (6a). Most genes examined in *C. elegans* which are within 400 bp of each other have been found to belong to a polycistronic transcription unit (43). The downstream genes of such units are *trans* spliced either exclusively with SL2 or with both SL1 and SL2 (43). To determine the nature of any *dpy-7 trans* splicing, we performed RT-PCR separately with SL1 or SL2 upstream primers in conjunction with *dpy-7* gene-specific downstream primers. A primer specific for the cyclophilin gene *cyp-9*, which is predominantly SL2 *trans* spliced (31a), was included in each reaction to act as an internal control for amplification of SL2-spliced transcripts. Reaction products were separated on an agarose gel, blotted, and sequentially probed with nested *dpy-7*- and *cyp-9*-specific radiolabelled oligonucleotides. *dpy-7* appears to be exclusively *trans* spliced with SL1 and not SL2, as *dpy-7*-specific SL2 spliced transcripts could not be detected even after prolonged exposure of the autoradiograph (Fig. 1). This finding suggests that *dpy-7* is not cotranscribed with the upstream gene but is transcribed from its own promoter.

The identification of the transcriptional start site is an important first step in locating the promoter region. However, *trans*-spliced genes, which account for up to 70% of *C. elegans* genes, present a problem in this respect since it is necessary to map the 5′ end of the prespliced transcript, which is of low abundance relative to the mature transcript. Consequently, an approach involving 5′RACE was designed to provide a specific and sensitive method for identifying the 5′ ends of *dpy-7* prespliced transcripts (Fig. 2). cDNA synthesis was primed with a gene-specific primer, and following homopolymer tailing, first-strand cDNA was used as the template in a first round of PCR using primers designed to amplify both prespliced and mature *dpy-7* transcripts. Three separate second-round PCRs were then performed (Fig. 2A). Reactions 1 and 2 were designed to be specific for the mature transcript, and these amplified single fragments of the sizes predicted from the position of the known SL1 splice acceptor site (Fig. 2). Reaction 3 was designed to be specific for prespliced transcripts, and this amplified two major products of between approximately 230 and 260 bp (Fig. 2B). Reaction 3 was also performed with radiolabelled primers, and the products were resolved on a polyacrylamide sequencing gel in order to maximize resolution (Fig. 2, track 4). This yielded two major bands of approximately 230 and 250 bp. The relatively poor resolution of products may be due, at least in part, to variation in the length of the homopolymer tail on the first-strand cDNA.

To define the 5′ ends of the primary transcripts precisely, the products of two independently performed PCR amplifications corresponding to reaction 3 were subcloned, and the sequences of inserts of 10 plasmids resulting from each subcloning were determined. A number of different 5′ ends were identified (Fig. 2C). The presence of products with identical 5′ ends from independent reactions strongly suggests that these products, at least, represent genuine cDNAs and are not simply degradation products. Therefore, it appears that the *dpy-7* gene initiates transcription from a cluster of different sites with two major sites at −164 and −186 bp upstream of the initiator ATG (−160 and −182 relative to the *trans*-splice acceptor site).

Expression pattern of *dpy-7–lacZ* and *dpy-7*–GFP reporter gene fusions. Following injection of plasmid DNA into the syncytial gonads of *C. elegans* hermaphrodites, injected circular DNA molecules recombine into large multimers termed extrachromosomal arrays (28). These are inherited in a non-Mendelian fashion, being transmitted to between 10 and 90% of the progeny. The frequency of transmission is believed to depend on the size of the arrays, which can include up to several hundred linked copies of the injected plasmid molecules. Plasmids containing different cloned sequences can be coinjected, and the different molecules are efficiently incorporated into the same extrachromosomal arrays. Thus, a plasmid containing a marker gene, which when expressed in transgenic animals produces a visible phenotype, e.g., plasmid pRF4 (28), can be coinjected into the hermaphrodite gonad with the plasmid of interest. This allows easy identification of transformed animals in the F_1 progeny. Transgenic lines are then established by clonally picking transformed F_1 animals, a proportion of which produce transformants in the F_2 generation which are picked to establish lines. In subsequent generations of such a line, individual transgenic animals transmit the multicopy arrays to a proportion of their progeny. Since individual animals show a degree of somatic mosaicism due to random mitotic loss of the arrays, several independent lines and a large number of individual animals are examined to establish precisely

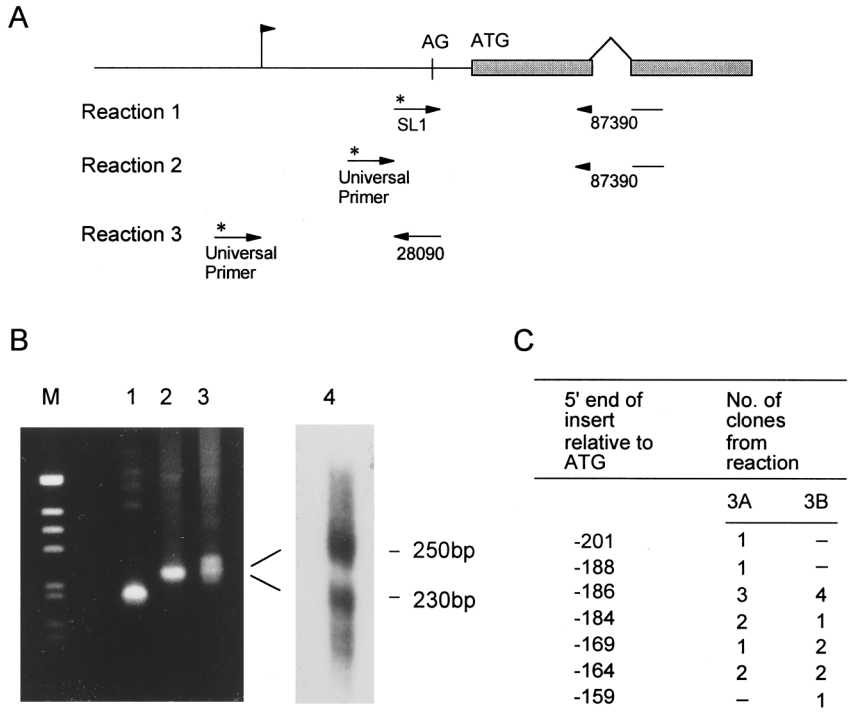

FIG. 2. Mapping of *dpy-7* transcription initiation. (A) Schematic representation of the second-round PCRs of the 5′RACE experiment. The flag represents the site of transcription initiation, and the *trans*-splice acceptor site is marked AG. Reactions 1 and 2 were designed to amplify product corresponding to the mature transcript, and reaction 3 was designed to amplify product corresponding to the primary transcript. Primers 87390 and 28090 are complementary to the *dpy-7* sequence. The SL1 and universal primers (marked with asterisks) do not anneal to the *dpy-7*-specific sequence but anneal to the spliced leader sequence and to the anchor primer sequence, respectively. (B) Ethidium bromide-stained agarose gel showing second-round PCR products of reactions 1, 2, and 3. Lane M, 1-kb ladder (GIBCO-BRL). The sizes of these molecular weight marker fragments, from the largest to the smallest, are 508, 396, 344, 298, 220, 201, 154, and 134 bp. None of the products were present in PCRs performed in parallel on the mock cDNA synthesis reaction (data not shown). Lane 4 shows the radiolabelled products of reaction 3 separated on a 6% polyacrylamide sequencing gel. The sizes of these products were estimated by comparison to products of a dideoxynucleotide chain termination sequencing reaction, using primer 28090 on pdpy7 as the template, which were run in adjacent tracks. (C) Table showing 5′ end points of subcloned PCR products. The products of two independent PCR amplifications corresponding to reaction 3 (3A and 3B) were subcloned, and the inserts of 10 plasmids resulting from each subcloning were sequenced.

which cells express a given reporter gene construct. Transgenic lines in which free arrays have been integrated into a chromosome at a random site can also be established. This is most commonly achieved by gamma irradiating a population of animals carrying an extrachromosomal array and selecting animals in the F_2 population which are homozygous for the transforming sequences. These integrated lines exhibit 100% inheritance of the transforming DNA between generations and exhibit significantly less somatic mosaicism than lines carrying free arrays (21). A detailed comparison of transgenic lines carrying multicopy extrachromosomal arrays with those carrying randomly integrated transgenes has been performed during an analysis of the *cis* regulatory sequences of the *vit-2* gene (27). No differences were found between the two approaches, validating the use of extrachromosomal arrays for this purpose. Although rare homologous insertions of transgenic DNA into chromosomal loci have been reported (2), an efficient method for homologous transformation of *C. elegans* has not yet been developed.

Plasmid pdpy7, containing a 3.6-kb genomic fragment encompassing the *dpy-7* gene, has previously been shown, by transformation experiments, to rescue the phenotype of animals which are homozygous for *dpy-7* recessive mutant alleles (14). This genomic fragment comprises 1,262 bp of *dpy-7* 5′ flanking sequence, the entire *dpy-7* coding sequence and intron, and 1,267 bp of 3′ flanking sequence. Hence, these sequences must be sufficient to direct expression of *dpy-7* from extrachromosomal arrays in a manner which basically reflects true *dpy-7* expression. As an initial approach to determine the *dpy-7* expression pattern, *lacZ* was inserted into the *dpy-7* coding region of a plasmid containing precisely the same *dpy-7* sequences as pdpy7. A synthetic transmembrane domain was included at the 5′ end of the *lacZ* gene to prevent the DPY-7–β-galactosidase fusion protein from being secreted from the cell (9). Following cotransformation of this reporter construct, p20.6.90/7, with the marker plasmid pRF4, two independent transgenic lines carrying extrachromosomal arrays were established, and the resulting β-galactosidase expression pattern was examined. In both transgenic lines, β-galactosidase expression was seen in hypodermal cells of the body, head, and tail of embryos from the late comma stage onward, in each of the four larval stages and in the adult. This expression pattern was entirely consistent with that expected for a cuticle collagen gene, since the components of the cuticle are synthesized and secreted by the hypodermis. However, this approach was not suitable for a more detailed analysis of the expression pattern, since the distribution of β-galactosidase throughout the cytoplasm does not allow hypodermal cells to be individually resolved. Consequently, we have used a set of promoterless *C. elegans lacZ* and GFP reporter gene vectors (kindly supplied by A. Fire, A. Xu, J. Ahnn, and G. Seydoux) to perform a detailed analysis of the expression pattern conferred by *dpy-7* 5′ *cis* regulatory sequences. These vectors encode a nuclear localization signal at the N terminus of β-galactosidase or GFP which,

by conferring a nuclear localization signal on the reporter molecule, causes its localization to the nucleus. This aids in the identification of individual cells which express the reporter molecule.

pdp7.1 contains 432 bp of *dpy-7* 5′ flanking sequence (relative to the initiator ATG) in translational fusion with *lacZ* in expression vector pPd21-28 (9). This construct was cotransformed with the marker plasmid pRF4, and the β-galactosidase expression pattern produced by this construct was studied in three independent transgenic lines carrying extrachromosomal arrays and two independent chromosomally integrated lines. No significant differences were detected between the lines except that, as expected, the integrated lines produced a much higher proportion of animals which stained and exhibited significantly less somatic mosaicism. In addition, the expression pattern was examined in F_1 progeny from hermaphrodites injected with pdp7.1 alone. The β-galactosidase expression pattern observed in these animals was not discernibly different from that observed in the transgenic lines, suggesting that the expression pattern was not dependent on, or significantly influenced by, the presence of the cotransforming marker plasmid pRF4. Also, four independent transgenic lines carrying the parental reporter gene vector pPd21-28 and pRF4 were examined, and in no case was β-galactosidase expression detected in hypodermal or other cells.

Expression was first seen in comma-stage embryos in nuclei corresponding to the positions of hypodermal cells (Fig. 3B). The positions of all hypodermal cell nuclei are most accurately determined in the L1 larvae (Fig. 3A, C, and D), and the identity of nuclei was confirmed by double-labelling immunofluorescence using MH27 and anti-LacZ antibodies to visualize simultaneously the hypodermal cell boundaries and β-galactosidase-expressing nuclei (Fig. 3F). In the majority (at least 80%) of those animals showing expression, β-galactosidase was detected in nuclei corresponding to the positions of hyp-7 and P-cell nuclei and not in V-cell or other seam cell nuclei (Fig. 3C, D, and F). These animals also showed similar levels of expression in hyp-5, hyp-6, and hyp-7 nuclei in the head and hyp-7 and hyp-10 nuclei in the tail (Fig. 3C). Staining was seen in other hypodermal cells of the head and tail but generally was more mosaic, was less intense, and took a longer time to develop. Similarly, a minority of the animals expressing β-galactosidase (less than 20%) showed some staining in seam cells. This was very mosaic, even in the integrated lines, and of variable intensity. It has previously been reported that in *C. elegans*, the level of *lacZ* reporter gene expression is reflected by the number of animals and cells expressing β-galactosidase as well as the intensity of X-Gal staining (21, 30, 31). Hence, these results suggest that the *dpy-7 cis* regulatory elements present in the reporter construct specify a high level of expression in hyp-7, P cells, and particular hypodermal cells of the head and tail but specify a much lower level of expression in hypodermal seam cells. β-Galactosidase expression was seen in an increasing number of nuclei at each developmental stage, including the adult, corresponding to the increase in hypodermal syncytial cell nuclei (Fig. 3E).

pdp7.GFP contains the same 432 bp of *dpy-7* 5′ flanking sequence as pdp7.1 but in a translational fusion with GFP in expression vector pPd95-67 (8a). Two independent lines carrying extrachromosomal arrays and two independent integrated transgenic lines were examined for GFP expression. As with the *lac-Z* constructs, expression was first seen at the early comma stage, and all fluorescing embryos showed high levels of expression in hyp-7 and P-cell nuclei and nuclei corresponding to the positions of hypodermal cell nuclei in the embryonic head and tail (Fig. 3G). Also, much fainter expression could be seen in the seam cell nuclei of some embryos (approximately 20 to 30% of fluorescing embryos). Expression was maintained throughout development, and in the postembryonic stages, nuclear localization was imperfect, leading to staining of both nuclei and cytoplasm. This allowed expression in the hyp-7 syncytium, P cells, and seam cells to be differentiated more easily. In postembryonic stages, high levels of GFP expression were seen in the nuclei and cytoplasm of hyp-7 and P cells (Fig. 3H and I) as well as head and tail hypodermal cells. In contrast, the seam cells appear as dark shadows, suggesting that GFP is either absent or expressed at much lower levels in the seam cells than in the rest of the hypodermis (Fig. 3H and I). Therefore, the GFP expression pattern is entirely consistent with the LacZ expression pattern.

dpy-7 transcripts can be detected by RT-PCR in all developmental stages except the adult (the adult cuticle is synthesized during the L4 stage), suggesting that the *dpy-7* polypeptide is present in all five cuticles synthesized during *C. elegans* development (13). In transgenic animals, β-galactosidase and GFP are detected in the hypodermis of each developmental stage, including the adult. We believe that the stability of the β-galactosidase and GFP polypeptides is responsible for their presence in the adult hypodermis and that our reporter gene analysis is consistent with the temporal patterns of wild type *dpy-7* mRNA abundance.

Identification of *dpy-7* 5′ flanking sequences necessary and sufficient to specify the hypodermal cell expression pattern. A number of translational *dpy-7–lacZ* reporter gene fusions were constructed (Fig. 4) and tested for the ability to produce the expression pattern defined above. Multiple independent transgenic lines carrying extrachromosomal arrays were tested for each construct. Plasmids pdp7.2 and pdp7.3, containing, respectively, 1,265 and 305 bp of 5′ flanking sequence, produced a pattern of β-galactosidase expression indistinguishable from that obtained with pdp7.1. Similarly, pdp7.5, identical to pdp7.1 except for an internal deletion between positions −131 and −12 relative to ATG, produced the same pattern. However, plasmid pdp7.4, which contained only 158 bp of 5′ flanking sequence, produced no β-galactosidase expression in any cells. Therefore, a region between −305 and −158 appears to be necessary for the hypodermal cell expression pattern, and conversely, sequences 5′ to −305 and 3′ to −131 appear not to be necessary. Ectopic expression in cells other than the pattern previously described was not detected with any of the deletion constructs. These results are entirely consistent with an alternative assay of transgenic expression, transformation rescue of phenotype. It was found that 305 bp of 5′ flanking sequence was sufficient to enable a cloned copy of the wild-type *dpy-7* gene to rescue the mutant phenotype in transgenic animals, whereas 158 bp was not (14).

A number of transcriptional *dpy-7–lacZ* fusions were constructed to determine the minimal 5′ element which is capable of specifying the hypodermal cell expression pattern (Fig. 4). The pattern of expression produced by the transcriptional fusions was essentially the same as for the translational fusions except for the following. The time taken for the staining to develop was increased (4 to 5 h, as opposed to less than 1 h), the staining in developmental stages after the L1 was weaker and much more mosaic, and no expression was seen in seam cells. We believe that these differences are due to the transcriptional fusions being less sensitive than the translational fusions, since the differences were seen with all of the transcriptional fusions and did not correlate with the absence of any particular region of *dpy-7* flanking sequence.

The results suggest that *dpy-7* 5′ flanking sequence present in the transcriptional fusion constructs functions as a true

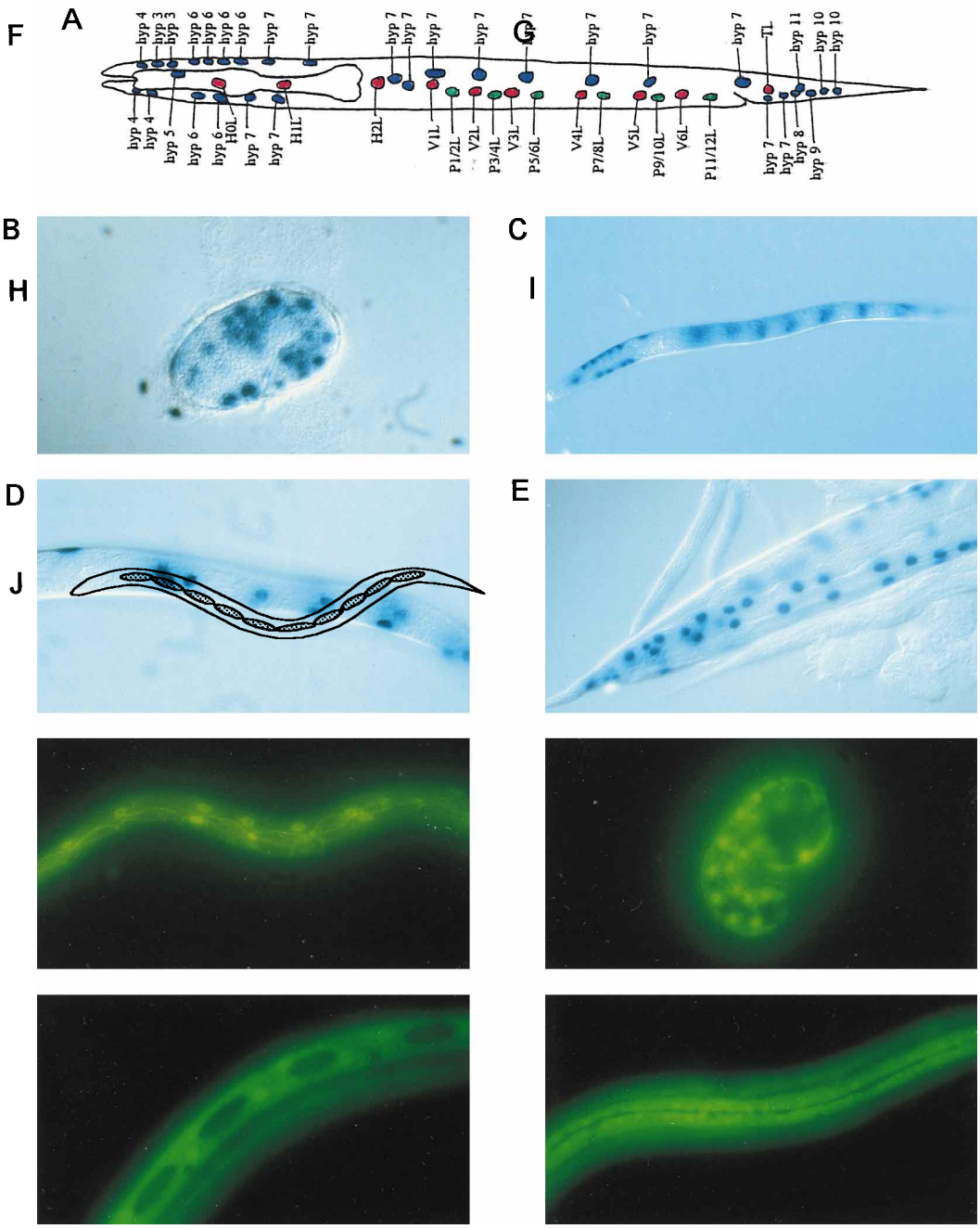

FIG. 3. *dpy-7* reporter gene expression. (A) Diagrammatic representation of the positions of hypodermal cell nuclei on the left lateral aspect of a newly hatched L1 larva (adapted from reference 39). The right lateral aspect is identical except for an extra hyp-7 nucleus which lies dorsal to H2R. The hypodermal cell nuclei of hyp-7 and of the head and tail are shown in blue. The P-cell nuclei are shown in green, and the seam cell nuclei are shown in red. (B) X-Gal-stained 1 1/2 fold embryo from an integrated line carrying pdp7.1. A lateral plane of focus is shown. Nuclei which stain correspond to the position of hypodermal cell nuclei of the main body region, head, and tail. An identical pattern is seen in the opposite lateral plane of focus. The anterior is to the left. (C) X-Gal-stained L1 larva from an integrated line carrying pdp7.1. A mid-line plane of focus is shown. A series of six pairs of nuclei down each lateral side of the animal are seen out of the focal plane. Nuclei in the head corresponding to hyp-5, hyp-6, and hyp-7 can be seen in the focal plane (other nuclei of the head and tail stain on more prolonged incubation). The anterior is to the left. (D) Same animal as panel C, at higher magnification and lateral plane of focus (head to left). A cluster of three nuclei which correspond to the position of hyp-7 nuclei is followed by a series of six pairs of nuclei (the anterior four pairs are shown). The same pattern is seen on the opposite lateral plain of focus except that only two nuclei, as opposed to three, are seen anterior to the set of six pairs. This pattern is consistent with β-galactosidase expression in all hyp-7 nuclei and either all V (seam cell) or all P hypodermal cell nuclei. Double-labelling immunofluorescence, an example of which is shown in panel F, was used to show that it is the P-cell and not the V-cell nuclei which stain in the majority of animals. The anterior is to the left. (E) Posterior end of an adult animal from an integrated line carrying pdp7.1. A lateral plane of focus is shown. Hypodermal syncytial cell nuclei can be seen in focus, while nuclei on the opposite side of the animal can be seen out of focus. The

anterior is to the right. (F) Double-labelling immunofluorescence. Shown is an L1 larva from an integrated line carrying pdpy7.1 doubly labelled with anti-β-galactosidase antibody and monoclonal antibody MH27. The nuclei expressing LacZ can be seen to be positioned above and below the seam cell boundaries outlined by MH27 and hence correspond to the position of hyp-7 and P-cell nuclei and not seam cell (V-cell) nuclei. The anterior is to the right. (G) Comma-stage embryo from an integrated line carrying the GFP reporter gene construct pdp7.GFP. A lateral plane of focus is shown. Fluorescence is seen in nuclei corresponding to the positions of hyp-7 and P-cell nuclei. An identical pattern is seen in the opposite lateral plane of focus. Nuclei corresponding to the positions of hypodermal cells in the head and tail are seen in the mid-line plane of focus (not shown). The anterior is to the top. (H) L1 larva from integrated line carrying the GFP reporter gene construct pdp7.GFP. A lateral plane of focus is shown. Fluorescence is seen in the cytoplasm surrounding the seam cells, which appear as dark shadows, and in nuclei corresponding to the positions of hyp-7 and P-cell nuclei. An identical pattern was seen in the opposite lateral plane of focus. The anterior is to the left. (I) L4 larva from integrated line carrying the GFP reporter gene construct pdp7.GFP. A lateral plane of focus is shown. Fluorescence can be seen in the cytoplasm and nuclei of the hypodermal syncytium, and the seam cells are seen as a thin line of dark shadows. The anterior is to the left. (J) Schematic representation of the boundaries of the seam cells in an L1 larvae. These are outlined by MH27 antibody staining in panel F and appear as dark shadows in the GFP-expressing transgenic animals. The anterior is to the left.

tissue-specific promoter and not simply as an enhancer element activating a cryptic promoter in the reporter plasmid. First, the inserts are functional only in forward orientation (pdp7.6 and pdp7.8 produce expression, but pdp7.7 and pdp7.9 do not). Second, a deletion between −197 and −157 which removes the *dpy-7* transcription initiation sites ablates all reporter gene expression (compare plasmid pdp7.8 with pdp7.10).

A series of 15-bp 5′ deletions was produced in order to identify the 5′ limit of the tissue-specific promoter element. Plasmids pdp7.11 and pdp7.12 both produced the full hypodermal cell expression pattern, whereas plasmids pdp7.13, pdp7.14, and pdp7.15 did not. Even overnight staining of animals at 37°C failed to detect any β-galactosidase activity in these lines. Therefore, the deletion of 15 bp of sequence between −267 and −252 appears to ablate β-galactosidase expression completely.

As with the translational fusions, none of the deletion derivatives of the transcriptional fusions produced expression in nonhypodermal cells.

Comparative sequence analysis of the *C. elegans* and *C. briggsae dpy-7* genes. As a complementary approach to the reporter gene studies, we have compared the *C. elegans dpy-7* gene with its homolog in the nematode *C. briggsae*, which is estimated to have diverged from *C. elegans* at least 40 million years ago (15). Intergenic sequences have diverged considerably between *C. elegans* and *C. briggsae*, but conservation of both sequence and function of regulatory elements has now been shown for a number of genes (15, 21, 42). Since the *C. elegans* cuticle collagen gene family consists of between 50 and 100 highly related members, it was important to identify unambiguously the true *dpy-7* homolog of *C. briggsae* as opposed to some other closely related cuticle collagen family member. Hybridization of a probe consisting of 933 bp of genomic sequence immediately upstream of the *C. elegans dpy-7* Gly-X-Y repeats detected a single band on both *C. elegans* and *C. briggsae* genomic Southern blots at a final wash stringency of 2× SSC at 55°C (data not shown). The same probe was used to identify hybridizing phage clones from a *C. briggsae* genomic library from which a hybridizing 3.1-kb *Eco*RI fragment was subcloned and sequenced. Virtual translation of the sequence revealed an open reading frame, interrupted by a single intron, encoding a predicted polypeptide which is 92.5% identical to the predicted *C. elegans* DPY-7 polypeptide (Fig. 5A). The

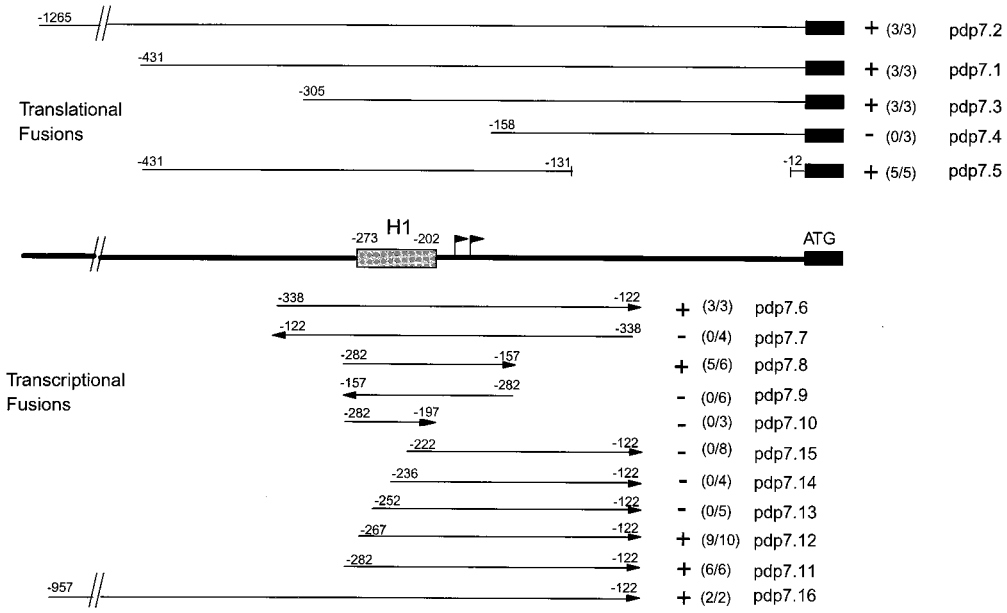

FIG. 4. *dpy-7–lacZ* fusion constructs. All fragments were subcloned into the *C. elegans* promoterless *lacZ* expression vector pPd21-28. Inserts ending in blocks, shown in the upper portion, represent in-frame translational fusions, whereas inserts ending in arrows, shown in the lower portion, represent transcriptional fusions where an arrow indicates the orientation of the insert. Numbers indicating the endpoints of inserts are relative to A of ATG (+1). Constructs producing β-galactosidase expression are indicated by +, whereas those failing to produce expression are indicated by −. The numbers in parentheses (*x*/*y*) indicate the proportion of transgenic lines expressing β-galactosidase, where *x* is the number of independent transgenic lines expressing β-galactosidase and *y* is the total number of transgenic lines tested. The transcription initiation sites are indicated by flags, and the H1 *C. elegans-C. briggsae* block of homology is indicated by the hatched box marked H1.

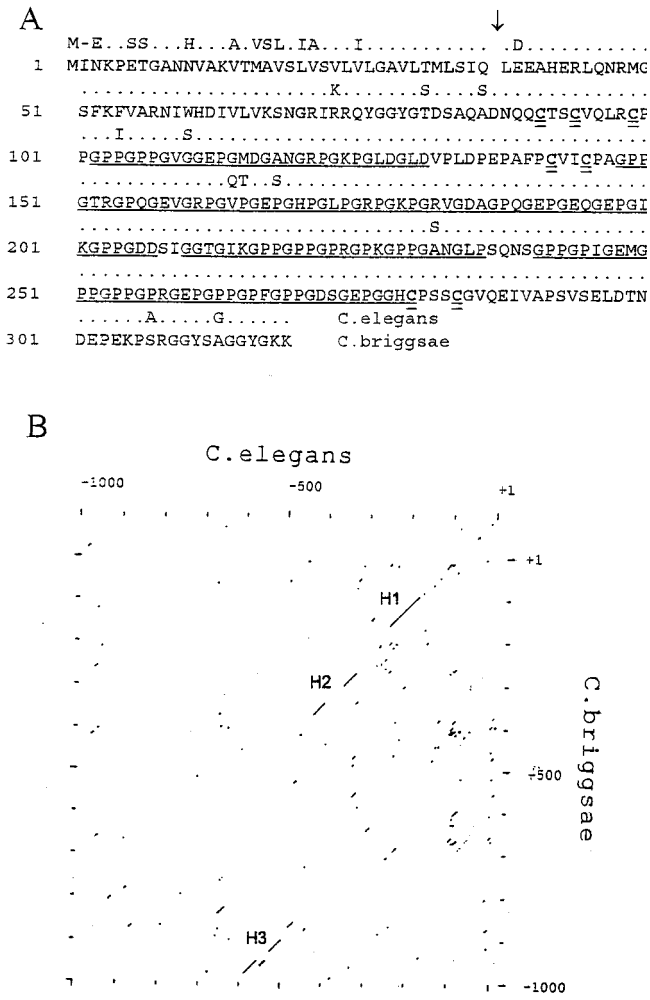

FIG. 5. *C. elegans-C. briggsae dpy-7* sequence comparison. (A) Comparison of amino acid sequences of the *C. elegans* and *C. briggsae* DPY-7 polypeptides. The entire *C. briggsae* sequence is shown with *C. elegans* residues above, identical residues being indicated by dots. Gly-X-Y domains are singly underlined, and cysteine residues are doubly underlined. The arrow indicates the putative site of signal peptide cleavage predicted as described by von Heijne (40). (B) Dot matrix comparison of the 5′ flanking sequences of the *C. elegans* and *C. briggsae dpy-7* genes. Three distinct regions of homology are indicated: H1, H2, and H3. The numbers on the dot matrix scales and the ends of the sequences indicate base positions relative to the A of ATG (+1).

important structural features of size and spacing of the Gly-X-Y repeats and number and spacing of the cysteine residues are precisely conserved. The signal peptide cleavage site, predicted by the method of von Heijne (40), is at precisely the same site as in the *C. elegans* DPY-7 polypeptide. The portion of polypeptide N terminal to the predicted cleavage site is significantly less conserved (70.3% similarity and 67.5% identity) than the putative secreted portion of the molecule (96.1% similarity and 95.7% identity). The position of the predicted intron is precisely conserved but consists of only 138 bp in *C. briggsae*, compared to 178 bp in *C. elegans*, and has relatively poor sequence conservation (57.2% identity). The predicted 3′ splice acceptor site sequence is precisely conserved (ATTCC AG/TT), and just a single nucleotide difference occurs in the predicted 5′ splice donor site (AG/GTTGGT in *C. elegans* and AG/GTAGGT in *C. briggsae*).

The overall identity of 1 kb of sequence immediately upstream of the *dpy-7* initiation codon between *C. elegans* and *C. briggsae* was only 46.8% when compared by GAP alignment (6). However, dot matrix analysis revealed three major blocks of homology, designated H1, H2, and H3, which have identities of 90.1, 67.3, and 76.3%, respectively (Fig. 5B). The Genefinder sequence analysis program predicts the presence of another gene upstream of, and on the same strand as, the *dpy-7* gene in both *C. elegans* and *C. briggsae* (6a). Homology blocks H2 and H3 correspond to the last two exons of this gene, with the positions of consensus splice donor and acceptor sites and stop codons being precisely conserved (data not shown). The predicted polypeptide is 65.6% identical between the two species over the last two exons, and no significant homologies have been found in BlastX or FastA database searches. The presence of the same predicted upstream gene in both the *C. elegans* and *C. briggsae* sequences confirms the identity of the *C. briggsae dpy-7* homolog.

The H1 homology region, which shows the highest degree of homology of the three blocks, is located immediately upstream of the transcription initiation sites and lies within the minimal functional promoter element defined by the reporter gene studies (Fig. 4 and 6). Hence, we consider H1 to likely represent the major regulatory site for the *dpy-7* gene, particularly since it is the only extensive region of homology in the sequence between *dpy-7* and the predicted upstream gene.

Functional conservation of *dpy-7 cis* regulatory elements and the DPY-7 polypeptide between *C. elegans* and *C. briggsae*. An interspecies phenotypic rescue experiment was performed to determine whether the regulation and function of the *dpy-7*

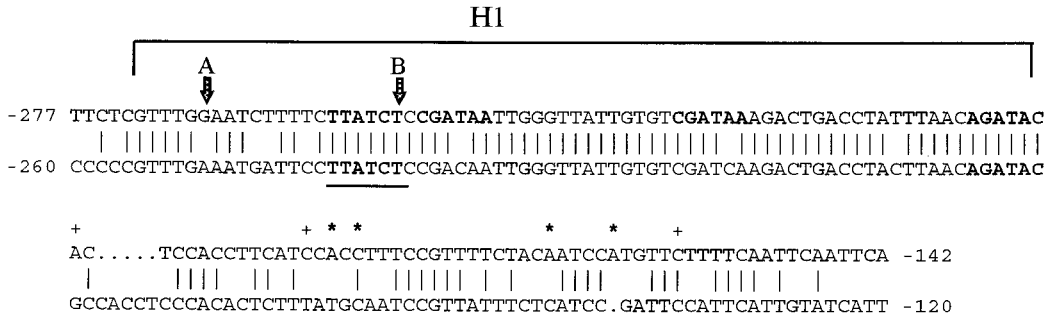

FIG. 6. *C. elegans* (upper) and *C. briggsae* (lower) sequences in the region encompassing homology H1 and the mapped *dpy-7* transcription initiation sites. Major initiation sites (bases at the 5′ ends of more than one PCR product cloned from two independent 5′RACE reactions) are marked with asterisks. Minor initiation sites (bases at the 5′ end of just a single PCR product cloned from only one PCR) are marked with plus signs. Arrow A marks the 5′ end of the insert of transcriptional *lacZ* fusion pdpy7.12, which produced the full β-galactosidase hypodermal cell expression pattern. Arrow B marks the 5′ end of the insert of the transcriptional *lacZ* fusion pdpy7.13, which failed to produce any β-galactosidase expression. The GATA-like motifs are shown in boldface, and the inverted consensus (W)GATA(R) motif is underlined.

FIG. 7. (A) Adult *C. elegans dpy-7* mutant strain CB88 [*dpy-7*(*e88*)X]. (B) Repaired adult F_1 produced by *C. elegans dpy-7* mutant strain CB88 [*dpy-7*(*e88*)X] following microinjection with the *C. briggsae* wild-type *dpy-7* gene (plasmid pJB1). (C) β-Galactosidase expression pattern of a *C. briggsae* L1 larva carrying the *C. elegans dpy-7–lacZ* fusion plasmid pdp7.3 on extrachromosomal arrays. (D) β-Galactosidase expression pattern of a *C. briggsae* L4 larva carrying the *C. elegans dpy-7–lacZ* fusion plasmid pdp7.3 on extrachromosomal arrays.

gene were conserved between *C. elegans* and *C. briggsae*. The *C. elegans dpy-7* mutant allele *e88* has an ethyl methanesulfonate-induced point mutation resulting in a glycine-to-arginine substitution within the second triple helical domain. This allele is recessive to the wild type, and the cloned wild-type allele can rescue the mutant phenotype in transgenic animals (14). Plasmid pJB1 contains a 3.1-kb genomic fragment which encompasses the *C. briggsae dpy-7* homolog and includes 1,091 bp of 5′ flanking sequence. This plasmid was microinjected into the syncytial gonads of 10 adult hermaphrodites of the *C. elegans* mutant *dpy-7* strain CB88 [*dpy-7*(*e88*)X] at a final concentration of 100 μg/ml. We observed 32 F_1 individuals which appeared to be fully rescued to the wild-type phenotype, as well as a number of partially rescued animals (Fig. 7A and B). In a second experiment, 10 adult hermaphrodites were coinjected with pJB1 at 10 μg/ml and pBluescript at 90 μg/ml; this yielded 15 apparently fully rescued F_1 animals. Ten hermaphrodites injected with 100 μg of pBluescript per ml alone yielded no rescued F_1 animals.

This interspecies phenotypic rescue experiment leads to a number of conclusions. First, it suggests that the *cis* regulatory elements of the *C. briggsae dpy-7* wild-type gene are sufficiently conserved for the gene to be expressed in the correct tissues and developmental stages in *C. elegans*. Second, it supports the sequence analysis in suggesting that this *C. briggsae* collagen gene is the true *dpy-7* homolog and also demonstrates that the polypeptide is sufficiently conserved to fulfill its structural role within the *C. elegans* cuticle.

To examine further the degree to which *dpy-7* regulatory elements are conserved between the two species, *C. elegans dpy-7–lacZ* reporter gene constructs were transformed into *C. briggsae*. Two independent transgenic lines carrying the translational fusion plasmid pdp7.3 produced a β-galactosidase expression pattern indistinguishable from that seen in *C. elegans* (Fig. 7C and D). Furthermore, three independent *C. briggsae* transgenic lines, transformed with plasmid pdp7.4, produced no detectable β-galactosidase expression, as was the case for *C. elegans*. Therefore, the *C. elegans dpy-7 cis* regulatory sequences are sufficiently conserved to specify a hypodermal cell-specific expression pattern in *C. briggsae*. Also, the 147-bp region which encompasses homology region H1 and is necessary for reporter gene expression in *C. elegans* appears to be necessary for expression in *C. briggsae*. Two stably transformed *C. briggsae* lines carrying plasmid pdpy7.6, a transcriptional *dpy-7–lacZ* fusion containing *C. elegans dpy-7* 5′ flanking sequence from −338 to −122, express β-galactosidase in a pattern indistinguishable from that produced in *C. elegans*. Hence, 216 bp of sequence, which encompasses both homology region H1 and the *C. elegans dpy-7* transcription initiation sites, is sufficient to direct a hypodermal cell expression pattern in *C. briggsae.*

DISCUSSION

We have set out to determine the *cis* sequences responsible for regulating the expression of the cuticle collagen gene *dpy-7*. The results of our experiments indicate that regulated tissue- and stage-specific expression of *dpy-7* is achieved essentially by a compact tissue-specific promoter element close to the 5′ end of the gene.

A gene is independently predicted by Genefinder only 346 and 358 bp upstream of the *dpy-7* ATG in *C. elegans* and *C. briggsae*, respectively. The conservation of the exons, stop codon, and splice sites between the two species strongly supports the accuracy of the Genefinder prediction. To our knowledge, all *C. elegans* genes examined to date which are within 400 bp of an upstream gene are downstream genes of polycistronic transcription units. Such genes are *trans* spliced either exclusively with SL2 or with both SL1 and SL2 (43). We have shown that *dpy-7* is *trans* spliced with SL1 but not with SL2, suggesting that it is not cotranscribed with the upstream gene. This finding, together with our repair-of-phenotype experiments, reporter gene experiments, transcription initiation site mapping, and *C. elegans-C. briggsae* comparative sequence analysis, places the *dpy-7* promoter within the 346 bp of intergenic sequence. This appears to be the smallest intergenic sequence in *C. elegans*, reported to date, which contains a promoter.

We have shown that a small region of sequence around the *dpy-7* transcription initiation sites is capable, in an orientation-specific manner, of driving reporter gene expression exclusively in hypodermal cells. Reporter gene expression was not detected in nonhypodermal cells, even with the smallest deletion constructs, suggesting that repressor elements outside the minimal promoter element are not required to establish a properly regulated pattern of expression. We cannot rule out the existence of additional enhancer elements, outside the region we have examined, which serve to modulate the level of *dpy-7* expression. However, if these exist, they are clearly not necessary to establish tissue- and stage-specific expression. This suggests a model of cuticle collagen gene expression that is similar to that proposed for the *C. elegans* myosin and vitellogenin gene families (31, 26) and involves a tissue-specific promoter element controlled fundamentally by activators rather than repressors of transcription. This contrasts markedly with a number of other *C. elegans* genes such as *ges-1* (1), *mec-3* (41), and *hlh-1* (21), which appear to be regulated in a more complex modular fashion involving both positive and negative regulatory elements. In these cases, minimal promoter elements, when fused to reporter genes, result in ectopic expression due to the absence of upstream repressor elements. For example, reporter gene constructs which contain less than 1 kb of *ges-1* 5′ flanking sequence are ectopically expressed in cells of the pharynx and tail but not in the gut, which is the site of expression of the wild-type gene (1, 7). Although the *cis* regulatory requirements of only a small number of *C. elegans* genes have been studied in any detail, it has been suggested that this latter form of complex modular regulation may be a general feature of those genes that are expressed prior to cellular differentiation (21). In contrast, genes that encode abundant structural proteins, expressed after cell fate has been determined, appear to be regulated in a simpler manner. Our work on *dpy-7* supports this hypothesis and shows that the simpler form of transcriptional regulation applies to a member of a structural gene family as large and complex as that of the cuticle collagen genes. Indeed, it is possible that simplicity may be an important regulatory feature of large, dispersed multigene families, since complex control of each and every family member, requiring multiple distal regulatory elements, might impose severe logistical problems. Regulation of expression by a compact tissue-specific promoter element may well be a general feature of the cuticle collagen gene family, since *lacZ* reporter gene constructs including only 235 bp of *col-19* (25) and less than 300 bp of *rol-6* 5′ flanking sequence (16a) are expressed in the hypodermis.

Interestingly, a 15-bp 5′ deletion (between −267 and −252 relative to the initiator ATG) within the H1 homology region which completely ablates hypodermal cell reporter gene expression disrupts an inverted AGATAA motif which is precisely conserved in *C. briggsae*. This motif fits the consensus binding site (WGATAR) of the GATA transcription factor family, whose members are important regulators of both lineage-specific gene expression and cellular differentiation in vertebrate and invertebrate species. Such motifs have already been shown to be important in the regulation of other *C. elegans* genes such as *ges-1* (7) and *vit-2* (26, 27). It is also worth noting that three more GATA-like motifs are also present within the *dpy-7* minimal functional promoter element. These could be of functional significance, since although they do not fit the perfect consensus, vertebrate GATA factors can bind nonconsensus motifs in vitro (16, 29). Indeed, the *C. elegans* GATA transcription factor, *elt-1*, has recently been shown to activate reporter gene expression in a yeast one-hybrid system via a GATC core sequence as efficiently as via a GATA core sequence (34). We have recently identified a new member of the *C. elegans* GATA transcription factor family, *elt-3*, which appears to be expressed in the hypodermis. Expression is first seen in pre-comma-stage embryos shortly prior to the time when *dpy-7* expression is first seen, and hypodermal expression is maintained throughout postembryonic development (10a). We are currently investigating the possible role of this transcription factor in the regulation of *dpy-7* and other cuticle collagen genes.

A comparison of the *dpy-7* 5′ flanking sequence with those other cuticle collagen genes, either published or predicted by the *C. elegans* genome project, did not reveal any particularly striking regions of conservation. Our analysis, in common with that of others (3), failed to find any extensive regions of upstream flanking sequence common to groups of cuticle collagen genes. Similarly, a comparison of the H1 region with the flanking sequence of other cuticle collagens, using the FastA program (6), revealed only small regions of imperfect homology, the significance of which is difficult to assess. WGATAR consensus motifs are present upstream of many of these collagen genes, often more frequently than predicted by chance, but any conclusions regarding their significance must await functional analysis.

The comparison of *dpy-7* between *C. elegans* and *C. briggsae* also has a number of important implications for cuticle collagen structure and function. The ability of the *C. briggsae dpy-7* gene to repair the *C. elegans* mutant phenotype is of particular interest, since cuticle assembly is thought to require multiple interactions between different collagen monomers, both for triple helix formation and for polymerization of collagen trimers (12, 17). Hence, the interspecies repair of phenotype suggests that not only are all of the important structural features of DPY-7 conserved between the species, but so too are those of other interacting molecules, implying a high degree of conservation of cuticle architecture and assembly between *C. elegans* and *C. briggsae*. This conclusion is strongly supported by the fact that transformation with the dominant mutant allele of the *C. elegans* cuticle collagen gene *rol-6*(*su1006*) produces the same strong right roller phenotype (sRRol) in *C. briggsae* as in *C. elegans*, allowing it to be used as a visible transformation marker in *C. briggsae*. In *C. elegans*, *rol-6* function is dependent on the presence of another cuticle collagen, SQT-1, since none of the *rol-6* mutations [including *rol-6(su1006)*] is visible in a *sqt-1* null mutant background (19). This finding has led to the suggestion that these two monomers are present in the same collagen heterotrimer. Indeed, the phenotype produced by *rol-6(su1006)* has been examined in a variety of *sqt-1* mutant backgrounds, and in the majority of cases it produces a phenotype markedly different from sRRol (19). Hence, not only does *sqt-1* have to be present for *rol-6* mutant phenotypes to be visible, but a variety of single substitution mutations in the SQT-1 polypeptide dramatically change the phenotype produced by *rol-6* alleles such as *su1006*. Consequently, the fact that *rol-6(su1006)* produces an sRRol phenotype in *C. briggsae*, identical to that seen in *C. elegans*, strongly suggests that *sqt-1* as well as *rol-6* is highly conserved. This evidence, taken together with that for *dpy-7*, suggests that the structural details of the *C. elegans* and *C. briggsae* cuticles are remarkably similar.

The high degree of conservation of the entire portion of the polypeptide predicted to be secreted into the cuticle is also of interest since the degree of divergence between *C. elegans* and *C. briggsae* is such that conservation is thought to be confined to functionally important residues (8, 15, 36). Genes with low codon bias are saturated for silent substitutions, and intergenic sequences and introns are generally poorly conserved (36). Genes encoding polypeptides which are expected to have spe-

cific functional domains such as *ced-9* (66.5% overall identity), *tra-2* (60% overall identity), and *hlh-2* (68% overall identity) exhibit clear regions of greater and lesser homology (11, 21, 22). In the case of *dpy-7*, all of the residues generally believed to be important for cuticle collagen function are conserved. However, there is also a high degree of conservation throughout the non-Gly-X-Y domains, and most of the X-Y residues in the Gly-X-Y domains are precisely conserved. This was a somewhat surprising observation and may suggest that these residues are more critical to cuticle collagen function than previously indicated by genetic studies.

ACKNOWLEDGMENTS

We thank Richard Durbin for the Genefinder analysis, Andy Fire, S. Xu, J. Ahnn, and G. Seydoux for supplying the reporter gene expression vectors, David Baillie for the *C. briggsae* genomic library, Bob Waterson for the MH27 monoclonal antibody, and Tony Page for the *cyp-9*-specific primers.

Some of the strains used in this work were supplied by the Caenorhabditis Genetics Centre, which is funded by the NIH National Centre for Research Sources. J.D.B. is a Wellcome Trust Senior Lecturer, and I.L.J. is a recipient of an MRC career development award. This work was supported by the Wellcome Trust.

REFERENCES

1. **Aamodt, E. J., M. A. Chung, and J. D. Mcghee.** 1991. Spatial control of gut-specific gene expression during *Caenorhabditis elegans* development. Science **252:**579–582.
2. **Broverman, S., M. MacMorris, and T. Blumenthal.** 1993. Alteration of *Caenorhabditis elegans* gene expression by targeted transformation. Proc. Natl. Acad. Sci. USA **90:**4359–4363.
3. **Cox, G. N., C. Fields, J. M. Kramer, B. Rosenzweig, and D. Hirsh.** 1989. Sequence comparisons of developmentally regulated collagen genes of *Caenorhabditis elegans*. Gene **76:**331–344.
4. **Cox, G. N., and D. Hirsh.** 1985. Stage-specific patterns of collagen gene expression during development of *Caenorhabditis elegans*. Mol. Cell. Biol. **5:**363–372.
5. **Cox, G. N., S. Staprans, and R. S. Edgar.** 1981. The cuticle of *Caenorhabditis elegans*. II. Stage-specific changes in ultrastructure and protein composition during postembryonic development. Dev. Biol. **86:**456–470.
6. **Devereux, J., P. Haerberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.

6a. **Durbin, R.** Personal communication.

7. **Egan, C. R., M. A. Chung, F. L. Allen, M. F. P. Heschl, C. L. Buskirk, and J. D. McGhee.** 1995. A gut-to-pharynx/tail switch in embryonic expression of the *Caenorhabditis elegans ges-1* gene centers on two GATA sequences. Dev. Biol. **170:**397–419.
8. **Emmons, S. W., M. R. Klass, and D. Hirsh.** 1979. Analysis of the constancy of DNA sequences during development and evolution of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **76:**1333–1337.

8a. **Fire, A., A. Xu, J. Ahnn, and G. Seydoux.** Personal communication.

9. **Fire, A., S. White Harrison, and D. Dixon.** 1990. A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. Gene **93:**189–198.
10. **Francis, G. R., and R. H. Waterson.** 1985. Muscle organisation in *C. elegans*: localisation of proteins involved in thin filaments attachment and I-band localisation. J. Cell Biol. **101:**1532–1549.

10a. **Gilleard, J.** Unpublished data.

11. **Hengartner, M. O., and R. H. Horvitz.** 1994. The *C. elegans* cell survival gene *ced-9* encodes a functional homolog of the mammalian proto-oncogene *bcl-2*. Cell **76:**665–676.
12. **Johnstone, I. L.** 1994. The cuticle of the nematode *Caenorhabditis elegans*: a complex collagen structure. Bioessays **16:**171–178.
13. **Johnstone, I. L., and J. D. Barry.** 1996. Temporal reiteration of a precise gene expression pattern during nematode development. EMBO J. **15:**3633–3639.
14. **Johnstone, I. L., Y. Shafi, and J. D. Barry.** 1992. Molecular analysis of mutations in the *Caenorhabditis elegans* collagen gene *dpy-7*. EMBO J. **11:**3857–3863.
15. **Kennedy, B. P., E. J. Aamodt, F. L. Allen, M. A. Chung, M. F. P. Heschl, and J. D. McGhee.** 1993. The gut esterase gene (*ges-1*) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. J. Mol. Biol. **229:**890–908.
16. **Ko, L. J., and J. D. Engel.** 1993. DNA-binding specificities of the GATA transcription factor family. Mol. Cell. Biol. **13:**4011–4022.

16a. **Kramer, J.** Personal communication.

17. **Kramer, J. M.** 1994. Structures and functions of collagens in *Caenorhabditis elegans*. FASEB J. **8:**329–336.
18. **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.
19. **Kramer, J. M., and J. J. Johnson.** 1993. Analysis of mutations in the *sqt-1* and *rol-6* collagen genes of *Caenorhabditis elegans*. Genetics **135:**1035–1045.
20. **Kramer, J. M., G. N. Cox, and D. Hirsh.** 1985. Expression of the *Caenorhabditis elegans* collagen genes *col-1* and *col-2* is developmentally regulated. J. Biol. Chem. **260:**1945–1951.
21. **Krause, M., S. W. Harrison, S. Q. Xu, L. S. Chen, and A. Fire.** 1994. Elements regulating cell-specific and stage-specific expression of the *C. elegans myo-d* family homolog *hlh-1*. Dev. Biol. **166:**133–148.
22. **Kuwabara, P. E., and S. Shah.** 1994. Cloning by synteny—identifying *C. briggsae* homologs of *C. elegans* genes. Nucleic Acids Res. **22:**4414–4418.
23. **Larminie, C. G. C., and I. L. Johnstone.** 1996. Isolation and characterisation of four devolpmentally regulated cathepsin B-like cysteine protease genes from the nematode *Caenorhabditis elegans*. DNA Cell Biol. **15:**75–82.
24. **Levy, A. D., and J. M. Kramer.** 1993. Identification, sequence and expression patterns of the *Caenorhabditis elegans col-36* and *col-40* collagen-encoding genes. Gene **137:**281–285.
25. **Liu, Z., S. Kirch, and V. Ambros.** 1995. The *Caenorhabditis elegans* heterochronic gene pathway controls stage-specific transcription of collagen genes. Development **121:**2471–2478.
26. **MacMorris, M., S. Broverman, S. Greenspoon, K. Lea, C. Madej, T. Blumenthal, and J. Speith.** 1992. Regulation of vitellogenin gene expression in transgenic *Caenorhabditis elegans*: short sequences required for activation of the *vit-2* promoter. Mol. Cell. Biol. **12:**1652–1662.
27. **MacMorris, M., J. Speith, C. Madej, K. Lea, and T. Blumenthal.** 1994. Analysis of the VPE sequences in the *Caenorhabditis elegans vit-2* promoter with extrachromosomal tandem array-containing transgenic strains. Mol. Cell. Biol. **14:**484–491.
28. **Mello, C. C., J. M. Kramer, D. Stinchcomb, and V. Ambros.** 1991. Efficient gene-transfer in *C. elegans*—extrachromosomal maintenance and integration of transforming sequences. EMBO J. **10:**3959–3970.
29. **Merika, M., and S. H. Orkin.** 1993. DNA-binding specificity of GATA family transcription factors. Mol. Cell. Biol. **13:**3999–4010.
30. **Okkema, P. G., and A. Fire.** 1994. The *Caenorhabditis elegans* NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in pharyngeal muscle. Development **120:**2175–2186.
31. **Okkema, P. G., S. Harrison, V. Plunger, A. Aryana, and A. Fire.** 1993. Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. Genetics **135:**385–404.

31a. **Page, A. P.** Personal communication.

32. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
34. **Shim, Y., J. J. Bonner, and T. Blumenthal.** 1995. Activity of a *C. elegans* GATA transcription factor, ELT-1, expressed in yeast. J. Mol. Biol. **253:**665–676.
35. **Snutch, T. P., and D. L. Baillie.** 1984. A high degree of DNA strain polymorphism associated with the major heat shock gene in *Caenorhabditis elegans*. Mol. Gen. Genet. **195:**329–335.
36. **Stenico, M., A. T. Lloyd, and P. M. Sharp.** 1994. Codon usage in *Caenorhabditis elegans*—delineation of translational selection and mutational biases. Nucleic Acids Res. **22:**2437–2446.
37. **Sulston, J., and J. Hodgkin.** 1988. Methods, p. 587–606. *In* W. B. Wood (ed.), The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
38. **Sulston, J. E., and R. H. Horvitz.** 1977. Post-embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev. Biol. **56:**110–156.
39. **Sulston, J. E., E. Schierenberg, J. G. White, and J. N. Thompson.** 1983. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev. Biol. **100:**64–119.
40. **von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14:**4683–4690.
41. **Way, J. C., L. Wang, J. Run, and A. Wang.** 1991. The *mec-3* gene contains *cis*-acting elements mediating positive and negative regulation in cells produced by asymmetric cell division in *Caenorhabditis elegans*. Genes Dev. **5:**2199–2211.
42. **Xue, D., M. Finney, G. Ruvkin, and M. Chalfie.** 1992. Regulation of the *mec-3* gene by the *C. elegans* homeoproteins UNC-86 and MEC-3. EMBO J. **11:**4969–4979.
43. **Zorio, D. A. R., N. N. N. Cheng, T. Blumenthal, and J. Spieth.** 1994. Operons as a common form of chromosomal organisation in *C. elegans*. Nature **372:**270–272.