ELSEVIER

Gene 183 (1996) 77–85

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Conservation of function and expression of *unc-119* from two *Caenorhabditis* species despite divergence of non-coding DNA

Morris Maduro, David Pilgrim *

Department of Biological Sciences, University of Alberta, Edmonton, Alta. T6G 2E9, Canada

Received 17 January 1996; revised 10 June 1996; accepted 13 June 1996

Abstract

The *Caenorhabditis briggsae* homologue of the *Caenorhabditis elegans* neuronal gene *unc-119* has been cloned by low-stringency hybridization. Genomic clones containing the *C. briggsae* gene are able to completely rescue the *unc-119* phenotype in transgenic *C. elegans* mutants. The open reading frame (ORF) of the predicted *C. briggsae* cDNA is 90% identical to that of *C. elegans*. Although the splice donor and acceptor sites are conserved, the untranslated regions, and the introns, differ greatly. For this gene, the average intron size in *C. elegans* is over 600 base pairs (bp); in *C. briggsae* it is only 113 bp. Their upstream control regions share limited sequence similarities; however, reporter gene fusions of the two species show strongly similar expression in *C. elegans*. These results are consistent with the maintenance not only of the function of the *unc-119* gene but also the transcriptional control of the gene through tens of millions of years of evolution.

Keywords: *C. briggsae*; Homologous gene; Sequence comparison; Neuronal promoter

1. Introduction

Sequence comparisons of protein and gene sequences between species are frequently used to establish conserved elements which may have importance for protein function or gene expression (e.g. Pilgrim et al., 1995). In the nematode, *C. elegans*, a project to sequence the entire genome is proceeding rapidly (reviewed in Hodgkin et al., 1995), and the data are analyzed concurrently for sequences predicted to be genes. In over 20 Mbp of contiguous sequence, 4000 predicted protein coding regions have been found, of which 45% show sequence similarity to known genes (Berks, 1995). As in the 'genome projects' of other organisms, there is a large proportion of predicted genes whose function and expression are unknown. It may be possible to assign a biological role to these sequences by examination of their mutant phenotypes following targeted mutagenesis (Hodgkin et al., 1995; Burns et al., 1994). Alternatively, comparison of conserved regions between 'sibling species' can be used to identify, by conservation, those regions likely to be important for function and regulation.

Comparison of *C. elegans* genes to those of the related species *C. briggsae* has demonstrated that while coding sequences are often highly conserved (allowing for degeneracy), intronic and flanking sequences have completely diverged (Snutch, 1984; Prasad and Baillie, 1989). The *C. briggsae* homologue of a *C. elegans* gene can often be detected by low stringency hybridization (Zucker-Aprison and Blumenthal, 1989; Kuwabara and Shah, 1994), and in some cases similar sequences in the 5′ flanking regions have been seen, suggesting that the control mechanisms for tissue- or cell-specific gene expression have been maintained (Zucker-Aprison and Blumenthal, 1989).

We have previously described a novel protein encoded by the *C. elegans unc-119* gene, which is expressed throughout the nervous system (Maduro and Pilgrim, 1995). The phenotype of *unc-119* mutants consists of defects in nervous system function; however, no biochemical role for the UNC-119 protein has yet been assigned. Since all known alleles of *unc-119* are molecu-

* Corresponding author. Tel. +1 403 4922792; Fax +1 403 4929234; e-mail: dave_pilgrim@biology.ualberta.ca

Abbreviations: aa, amino acid(s); bp, base pair(s); βgal, β-galactosidase; dCTP, 2′-deoxycytosine triphosphate; DIC, differential interference contrast; GFP, green fluorescent protein; kb, kilobase(s) or 1000 bp; Mbp, megabase(s) or 10^6 bp; Myr, 10^6 years; NLS, nuclear localization signal; ORF, open reading frame; RACE, rapid amplification of cDNA ends; *UTR*, untranslated region(s).

PII S0378-1119(96)00491-X

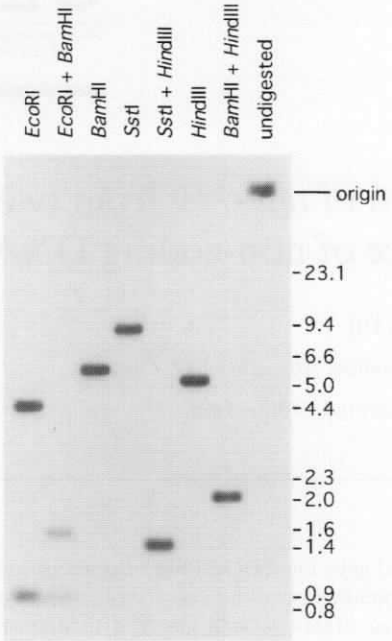


Fig. 1. Southern (1975) analysis of restriction-digested pDP#MMCb1. **Methods:** A *C. briggsae* λCharon4 genomic library (a gift from T. Snutch and D. Baillie) was screened with purified insert DNA from pDP#MM008 (Maduro and Pilgrim, 1995) labeled with ^{32}P-dCTP using the T7 Quickprime kit (Pharmacia). The library was grown on *E. coli* LE392 and lifted to Hybond N (Amersham). Hybridization was performed at 62°C for 12 h, followed by three 10 min rinses at 55°C in 2×SSC containing 0.1% SDS, prior to autoradiography. Purification of λ and manipulation of DNA were performed as described (Sambrook et al., 1989) using the vector pBluescript KS− (Stratagene). Restriction enzymes were obtained from Gibco BRL, except for *Hin*dIII (Pharmacia). Fragments of the positive λ clone pDP#MMCb1 were separated on a 0.7% agarose gel prior to blotting. Hybridization with purified *unc-119* cDNA was carried out as for the library screen.

lar nulls, and the predicted UNC-119 protein is not significantly similar to other known proteins, there are no clues as to which regions are important for function.

The homologous *unc-119* gene in *C. briggsae* was analyzed in order to identify which parts of the gene are likely to be necessary for activity or regulation. We show that the *C. elegans* gene can be functionally replaced by the *C. briggsae* counterpart, and that despite

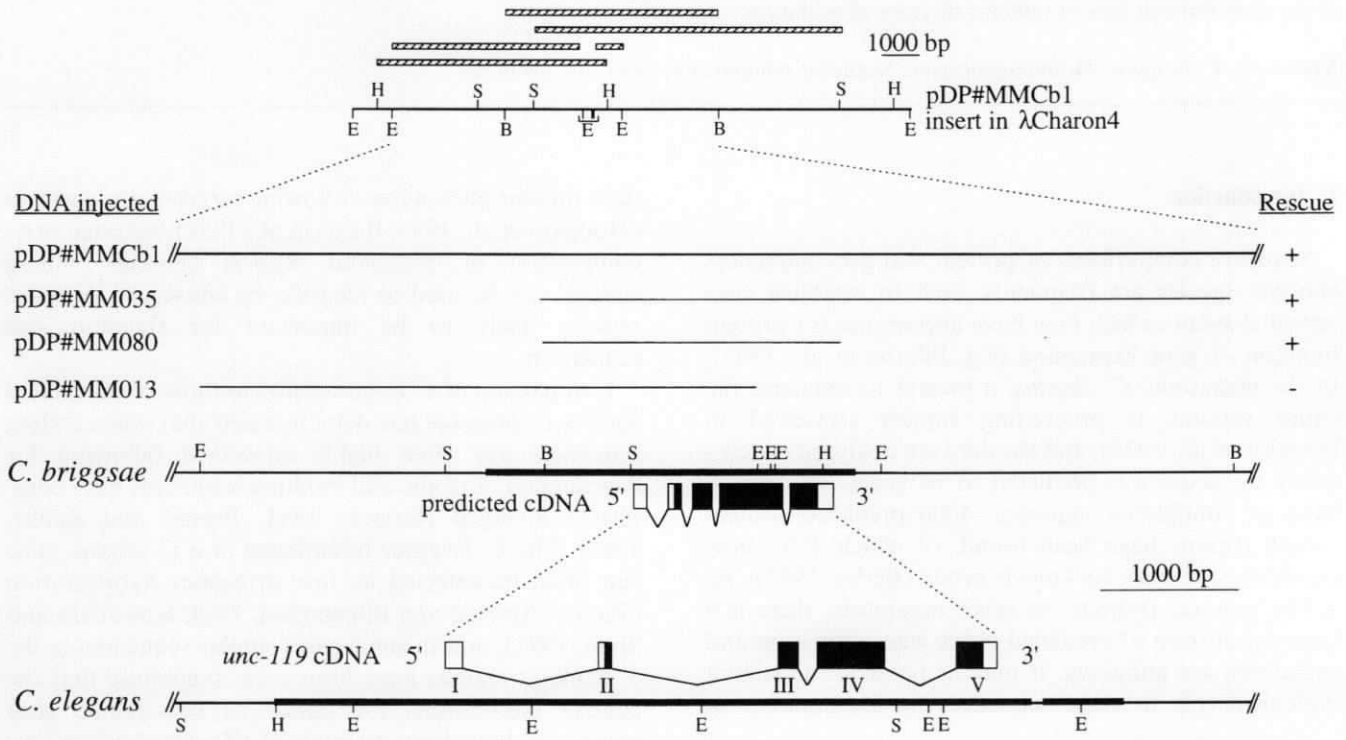


Fig. 2. Restriction map of the *C. briggsae* insertion from the λCharon4 clone pDP#MMCb1, rescuing ability of subclones, and the predicted cDNA aligned with the *C. elegans* counterpart. Fragments that cross-hybridize with the *unc-119* cDNA (from Fig. 1) are shown above the map as shaded boxes. Transgenic animals carrying the various clones were constructed and phenotypically analyzed as previously described (Maduro and Pilgrim, 1995). Rescue is denoted as '+', while failure to rescue is shown as '−'. The clone pDP#MM080 is a fusion of *C. briggsae unc-119* at an *Nde*I site in the 3′ end of the coding region to a *lacZ* reporter gene containing the *unc-54* 3′-untranslated region (UTR) (Fire et al., 1990) which excludes only the last amino acid in the ORF. Rescue by pDP#MM035 was only ascertained on the basis of rescue of the locomotory defect. The thicker regions on the bottom maps respresent those areas that have been sequenced. The *C. elegans unc-119* sequence, which has been published (Maduro and Pilgrim, 1995), is available through GenBank (accession No. U32854). B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sst*I.

only small sequence similarities, the control regions from each species drive βgal reporter gene expression in *C. elegans* indistinguishably.

2. Results and discussion

2.1. Identification of the C. briggsae homologue of unc-119

A genomic phage library of *C. briggsae* was screened at low stringency using a genomic fragment containing the majority of *unc-119* from *C. elegans*. One positive clone, pDP#MMCb1, was chosen for further study. The *C. elegans unc-119* cDNA cross-hybridizes to specific restriction fragments of this clone (Fig. 1). In order to verify that this clone contained the entire *unc-119* homologue, we injected *unc-119*(e2498) animals with pDP#MMCb1 and the plasmid pRF4, which contains the dominant *rol-6* marker as a control for transformation (Mello et al., 1991). The e2498 allele results from a transposon insertion in *unc-119* and has a null phenotype (Maduro and Pilgrim, 1995). Among the F_1 progeny, several Rol non-Unc animals were seen, characteristic of rescue of the *unc-119* mutant phenotype, while injection of pRF4 alone did not confer rescue.

A restriction map of the *unc-119* region of *C. briggsae* was generated, and subcloned fragments were tested for their ability to complement the Unc phenotype (Fig. 2). When the rescuing ability of pDP#MMCb1 and a plasmid subclone was assessed in detail (Fig. 3), defects in egg laying, pharyngeal pumping and dauer forming ability were also rescued by the *C. briggsae* transgenes, similar to results obtained with *C. elegans unc-119* clones (Maduro and Pilgrim, 1995). This confirms that the cross-hybridizing region contains a functional homologue of *C. elegans unc-119*.

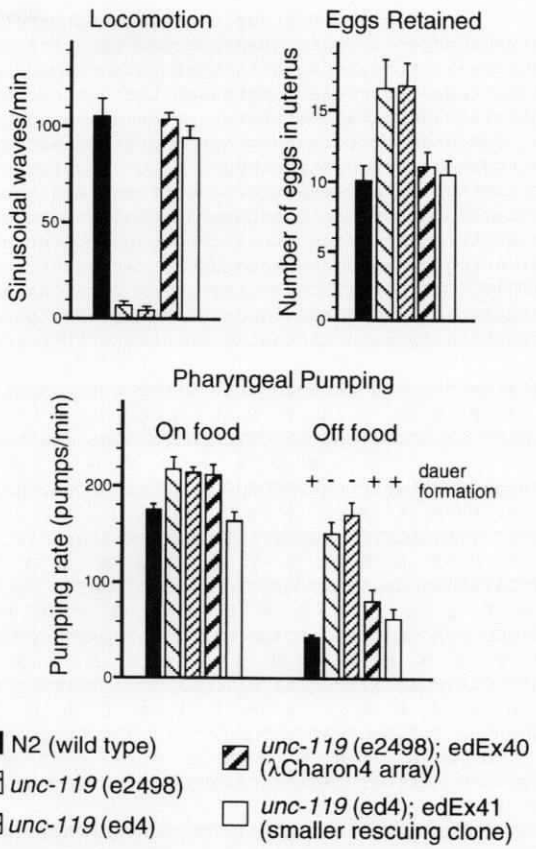


Fig. 3. Quantitation of defects in wild-type, *unc-119* mutants and mutants carrying the putative *C. briggsae unc-119* gene on transgenic arrays. The e2498 mutation results from a transposon insertion 3′ to exon IV, and ed4 results from a nonsense mutation in exon IV (Maduro and Pilgrim, 1995). Bars show the standard error of the mean (SEM). Data collection and the dauer-formation assay were performed as described (Maduro and Pilgrim, 1995). The array *edEx40* contains pDP#MMCb1 and pRF4; *edEx41* contains pDP#MM080 alone.

2.2. The coding regions of the homologues are conserved

The sequence of the *C. briggsae unc-119* genomic region was determined (Fig. 4). Using the previously determined *C. elegans* sequence and the *C. elegans* consensus splice donor and acceptor sites, we determined the presumed intron/exon junctions for the *C. briggsae* gene (shown schematically in Fig. 2). With the exception of the first exon (see next section), determination of the intron/exon boundaries was facilitated by the high degree of coding sequence conservation. An excellent alignment between the two predicted proteins can be produced if two single amino acid (aa) gaps are allowed in the *C. briggsae* ORF. The remaining codons can be aligned with the *C. elegans* ORF and show 90% identity over 217 aa (Fig. 5). Unfortunately, this high degree of similarity does not allow identification of specific amino acids important for function. In *C. elegans*, we have shown that the codons specified by exon II (the first 20 aa) are partially dispensable for function, since genomic clones containing only the last three exons (exons III–V) are sufficient for partial phenotypic rescue (Maduro and Pilgrim, 1995). However, the amino acids encoded by exon II from the two strains are still well conserved (90% identical), suggesting that this part of the polypeptide probably has a function.

Fig. 5 also shows an alignment of the conceptual translation product of *unc-119* from both *C. elegans* and *C. briggsae*, as well as C27H5.1, a predicted ORF identified by the *C. elegans* genome sequencing project, and noticed previously as a potential *C. elegans* homologue of *unc-119* (Maduro and Pilgrim, 1995). When compared to both the *C. briggsae* and *C. elegans* homologues, 25% identity and 40% similarity are seen. The relative positions of the introns in the coding sequence is not conserved at all with C27H5.1 (data not shown).

```
          5'-gtctcatagattacatgagtaattggtctctatttctgaaagttccgatctttcaaatgttcagttaacagcagtgttcacaca  -1001
tatttgaccagccttcggattcgaaataaacaaactgtcttttccaaattgtttccttccgttttcaagtttgaatgactcatgtaagggtgcatccaaa   -901
aagaagaatttttcgaactagcgatgctgaagtgtggctagagcaaactagacagtcctaacagagccatttcgccagtttcgtaaagttaccatgaaac   -801
ctgcaatagaagggtcccaactctgttaagtttgctagtttttcttatctcctgtgcaagatatcaaaaatgtgtcaactgataaaatttaatgatgtca   -701
tgatttatcagttgacaacttttttatatctttcacaggaacagagataatatggctcaaagtgcaaacgtcctatggatcctatgcttgcgctttgagcc   -601
aatttatccaagtccttgtaaaaagtattcgaaaattgttaacggataaaatgtttattataatatcaaaaacaatttgtcagttgaccacttttgata   -501
tatttgacagaaaccgggatgaattggctcaaagttagggcgcccttctattacagggtttctgataacaaaccggttattaactcccaacaagggatga   -401
tttcaattcatcatgctcaaattgacccaaattaagttacatgacaaattcatcgccctttttcaactcttcttggctcatcatctgttattctgttccta   -301
'ttctgtctgcaccctatacccctttgcatactctctcgtcattcctttcttggatagtgtcttcttctccccagctctgctacttctatgacttgcgccc   -201
gctgcttttccgctcgttcctctcctctctgacgtcttcttctgcttctttttctccttccttttttcccttctcaatcttcctctttttcccatttctgtc   -101
aatcattcgaagaagaagaagaagaagagccctcattcatttcatttttttttctgtcggtgtgtgctgctcggttaagagtgagctctctattccacgt     -1
CTTCTTCTTTTTCTTCTTCGATTCGAATCAATCACTCCACCAAAACGCATTCGTTTTTGGGATTCACCCCGCGGTTCGCAACAGgtttctttttcaaatt    100
attagcgttataaatagaaaaatgtggagtttcaataaaaatgataatttacaaagtgtattttgattacatgtactcaaaaaggttgaaatttttcata    200
ccagttttccggaaatccatctgatatcattatcgtatttttctttttaaaaatgttttcaaaaaaaaacaaaatatagctggttatttggcaccctcta    300
                                                                                M  K  A  E  Q  Q  Q
attaccattttcctgtcacaccacactcttttccttcttccacttctttccgttttcagCCGCTTCCAACCAAACCGATATGAAAGCCGAGCAACAACAA    400
S  I  P  P  G  S  A  T  F  P  S  Q
TCGATTCCACCCGGCTCGGCGACCTTCCCGTCGCAGgtgagactcagaaactagagaaccgctcaactactcttgatatcgatttcatctttctctttct    500
                     M  P  R  P  P  P  S  T  E  Q  G  I  T  T  E  S  E  L  A  K  K  A  Q  I  T  P  N
ttgtgacttcatattccagATGCCACGGCCACCACCAAGCACCGAACAAGGAATCACAACGGAATCGGAGCTTGCGAAGAAAGCTCAAATCACTCCGAAC    600
D  V  L  A  L  P  G  I  T  Q  G                                                F  L  C  S  P  S  A  N
GACGTTTTAGCACTTCCGGGAATCACTCAAGgtatcttgctccctttggatttcagtttaaaacataaatttagGATTCTTATGCTCCCCATCTGCGAAC    700
I  Y  N  I  E  F  T  K  F  Q  I  R  D  L  D  T  E  Q  V  L  F  E  I  A  K  P  E  N  D  Q  E  N  D  E
ATCTATAACATCGAGTTCACCAAGTTCCAAATCCGTGATCTGGACACTGAGCAAGTGCTTTTCGAGATCGCCAAACCGGAGAACGATCAGGAGAATGATG    800
  S  P  Q  E  S  A  R  Y  V  R  Y  R  F  A  P  N  F  L  K  L  K  T  V  G  A  T  V  E  F  K  V  G  D
AGTCGCCACAGGAGTCGGCAAGATACGTGCGTTATAGATTTGCTCCAAACTTTTTGAAACTCAAAACGGTCGGAGCAACTGTGGAATTCAAAGTAGGAGA    900
 I  P  I  H  H  F  R  M  I  E  R  H  F  F  F  D  R  L  L  K  C  F  D  F  E  F  G  F  C  I  P  N  S
CATCCCAATCCATCATTTCCGAATGATCGAACGTCACTTCTTCTTTGATCGCCTTCTGAAGTGTTTTGACTTTGAATTCGGATTCTGTATTCCGAATTCA   1000
R  N  N  C  E  H  I  Y  E  F  P  Q  L  S  Q  Q  L  M
'CGAAACAACTGTGAACATATCTATGAGTTCCCTCAACTCTCTCAACAACTCAgtgagtcattattctaaaaagtacaattcaaaagactaatctcttttt   1100
    D  D  M  I  N  N  P  N  E  T  R  S  D  S  F  Y  F  V  D  N  K  L  V  M  H  N  K  A  D  Y  S  Y
cagTGGACGACATGATCAACAATCCAAACGAGACTCGTTCTGACAGCTTCTATTTCGTCGATAACAAACTCGTCATGCACAACAAAGCCGACTACTCATA   1200
 D  A
TGATGCATAAATATTTAATACAAAATGTTCTGGATAATTATTCTGTCGAATAGAAAAAACTCCAATGTGATTAAATTCCAATATTCCTGTCTGTTTTGTC   1300
CTTCCCTTCCCTCTCTGTCATGCATTCTAAGCTTTTCAGTTCCCCCCTTGTTTCTATATTTTTTCGCTGTCCTGTCACACTCGCTAAAAACACTAATCAC   1400
ACGGAAATCTGTTTTCAATAAAAACTCCAACTTTAactcattttcaatttcaactgaaagatttttcattagagaatgtccagcgaggaggatacattg    1500
tacccatcaaacttgcgcaaagatttattgtaggttttcagtacttttttgagggagaaaaagtaatctcactaaaactttaaaaaaactgatactggcac   1600
tcaaggatattgttctttttgagggcttccgggaaaatttgaaatttaaaggaaggaaaaaatcattgaaaagatttcaac-3'   1681
```

Fig. 4. The genomic nt sequence of *C. briggsae unc-119*. Predicted exons are shown in upper case, with conceptual translation above the sequence in single letter code above the first base of each codon. The start and stop codons and the putative polyadenylation sequence are underlined. **Methods:** Sequencing was performed using Sequenase 2.0 (United States Biochemical) on small-scale single stranded templates prepared as described (Sambrook et al., 1989), using universal primers for the pBluescript vector. One additional oligonucleotide, MMA1 (5′-AGTCGGCCTTATTGTGCATTAC-3′), originally designed for use in *C. elegans*, was stable as a sequencing primer in *C. briggsae* despite three base mismatches (underlined). The first exon was determined by RACE using a 5′-RACE System kit (BRL) according to the manufacturer's instructions. Total RNA from *C. briggsae* was prepared using a standard guanidine isothiocyanate protocol (Sambrook et al., 1989). The primer used for first strand synthesis was MMA21 (5′-AGTCGGCTTTGTTGTGCATGAC-3′) and the nested primer was MMA5 (5′-GGAGCATAGGAATTCTTGAGTGATTCC-3′). The approx. 275 bp product was reamplified, cloned and sequenced as for the *C. elegans* 5′-RACE product (Maduro and Pilgrim, 1995). Sequences were compiled and conceptually translated using DNA Strider 1.2 (Marck, 1988). This sequence is available through GenBank (accession No. U45326).

Since there do not appear to be any regions more conserved than others between *C. briggsae* and *C. elegans unc-119*, either there has been insufficient time for evolutionary change (unlikely, given the divergence of the intron sequences), or there is strong evolutionary pressure on the majority of the protein sequence. The latter is consistent with the estimated 20–50 Myr (million years) of evolution between the two species (Heschl and Baillie, 1990; Lee et al., 1992; Kennedy et al., 1993), and results of similar studies with other homologues. The 90% identity between *C. elegans* and *C. briggsae unc-119* is similar to that seen between homologues of a ubiquitin-like protein (Jones and Candido, 1993), but much higher than the 60–70% identity seen between the functional homologues of HLH-1 (Krause et al., 1994) or TRA-2 (Kuwabara and Shah, 1994). It was not surprising, therefore, to find that the *C. briggsae unc-119* gene could complement mutants of *C. elegans*, especially given that such interspecies rescue has been seen with the more dissimilar HLH-1 homologues (Krause et al., 1994).

2.3. Intron and flanking sequences are divergent

The location and size of the noncoding first exon of *unc-119* in *C. briggsae* was determined by 5′-RACE (rapid amplification of cDNA ends; Frohman et al., 1988). This exon is smaller than that of the *C. elegans* gene (84 bp compared to 128 bp) and bears no sequence similarity except for the occurrence of T-rich sequences

```
C. elegans   MKAEQQQQSIAPGSATFPSQMPRPPPVTEQAITTEAELLAKNQITPNDVL
C. briggsae  ·····-··················S···G····S··AK·A·····
C27H5.1      MATT  (C27H5.1 is shorter than the other two proteins)

                                  ‡          ‡      ‡
C. elegans   ALPGITQGFLCSPSANVYNIEFTKFQIRDLDTGHVLFEIAKPENETEENL
C. briggsae  ···············I·············EQ··········Q-··D
C27H5.1      ·TRHQDSKLSEKAESILAGFKLNWMNL··AE··K··WQST----------

             ◊   ‡      ‡           ‡                          ‡
C. elegans   QAQAESARYVRYRFAPNFLKLKTVGATVEFKVGDVPITHFRMIERHFFKD
C. briggsae  ESPQ·····························I··H·····F·
C27H5.1      EDM·DPK·EHKAHVPK·L··CR··SREIN·-TSS·K·EK··LEQ·VYL·G

                ‡ ‡  ‡    ◊                                   ‡
C. elegans   RLLKCFDFEFGFCMPNSRNNCEHIYEFPQLSQQLMDDMINNPNETRSDSF
C. briggsae  ···········I·····································
C27H5.1      TIIEEWY·D···VI·D·T·TWQNMI·AAPE··MFPPSVLSGNVVVET-L·

                ◊       ‡
C. elegans   YFVENKLVMHNKADYSYDA*
C. briggsae  ···D··············
C27H5.1      ·DGDL-··STSRVRLY··*
```

Fig. 5. Alignment of the *C. elegans* and *C. briggsae* predicted UNC-119 protein sequences, and the predicted product of the open reading frame C27H5.1 from *C. elegans*. ●, amino acid identical to *C. elegans*; −, gap introduced to optimize alignment; ◇, amino acid identical between C27H5.1 and *C. briggsae*; ‡, conservation of charged or hydrophobic amino acid; *, stop codon. The GenBank accession number for C27H5.1 is U14635.

(Fig. 4). While there is an apparently unused start codon in the *C. elegans* first exon (Maduro and Pilgrim, 1995), there is no counterpart in *C. briggsae*, suggesting it is not required for expression.

There is a polyadenylation signal (5′-AATAAA-3′) 206 bp after the predicted translation stop codon. From comparisons made of many *C. elegans* cDNA clones (Krause, 1995), the end of the mRNA is probably 13 bases downstream. This is comparable to the 3′ end of the *C. elegans* message, in which the polyadenylation signal occurs 158 base pairs after the stop codon (Maduro and Pilgrim, 1995). There is no apparent conservation between the translation stop and the end of the message, suggesting that unlike some *C. elegans* genes (e.g. *tra-2*; Goodwin et al., 1993), the expression of *unc-119* probably does not involve regulation by mRNA stability through the interaction of *trans*-acting factors with the 3′-*UTR*.

The intron sizes are quite different between the two species (shown schematically in Fig. 2). The intron between exons I and II is 890 bp long in *C. elegans*, but only 275 bp in *C. briggsae*. The largest difference is in the next intron, 1070 bp in *C. elegans* but only 83 bp in *C. briggsae*. The other introns are closer in length, with *C. elegans*/*C. briggsae* sizes of 161/43 and 488/51 bp. This pattern of smaller introns in *C. briggsae* has been seen in many genes, including *ges-1* (Kennedy et al., 1993), where the average *C. elegans* intron is eight times as large.

2.4. Putative promoter elements are conserved in the 5′ flanking region

Fig. 6 aligns the two nt sequences in a matrix comparison. Apart from the coding region, there is virtually no conservation of nt sequence between the two genes over the regions compared. The only exception is in the 5′ flanking region (Fig. 6b), where limited similarity is seen between regions rich in pyrimidine residues (on the upper strand), as well as three regions A, B and C (Fig. 6b and c). The A regions share limited simlarity of 15/22 bp, while the B and C regions are more similar: the centers of these regions share 23/26 bp (B) and 26/31 bp (C). At the center of region C is the sequence 5′-TGTCAAT-3′, which is the VPE1 consensus sequence, identified in the promoter region of the vitellogenin genes from the two species (MacMorris et al., 1994). The *C. elegans* regions A and B contain a 6/7 bp match to this same consensus, which has been shown to be important for high-level expression of the vitellogenin genes (Spieth et al., 1985; MacMorris et al., 1994). The relative 5′ to 3′ order of these three regions is conserved, although the distribution of the pyrimidine-rich segments is different. For example, in *C. briggsae*, regions A and B overlap and are not preceded by a pyrimidine-rich segment, while the opposite is true in *C. elegans*. Collectively, the regions are closer to the start of transcription in *C. elegans* by about 100 bp; it is curious that sequence C in *C. elegans* overlaps with the first

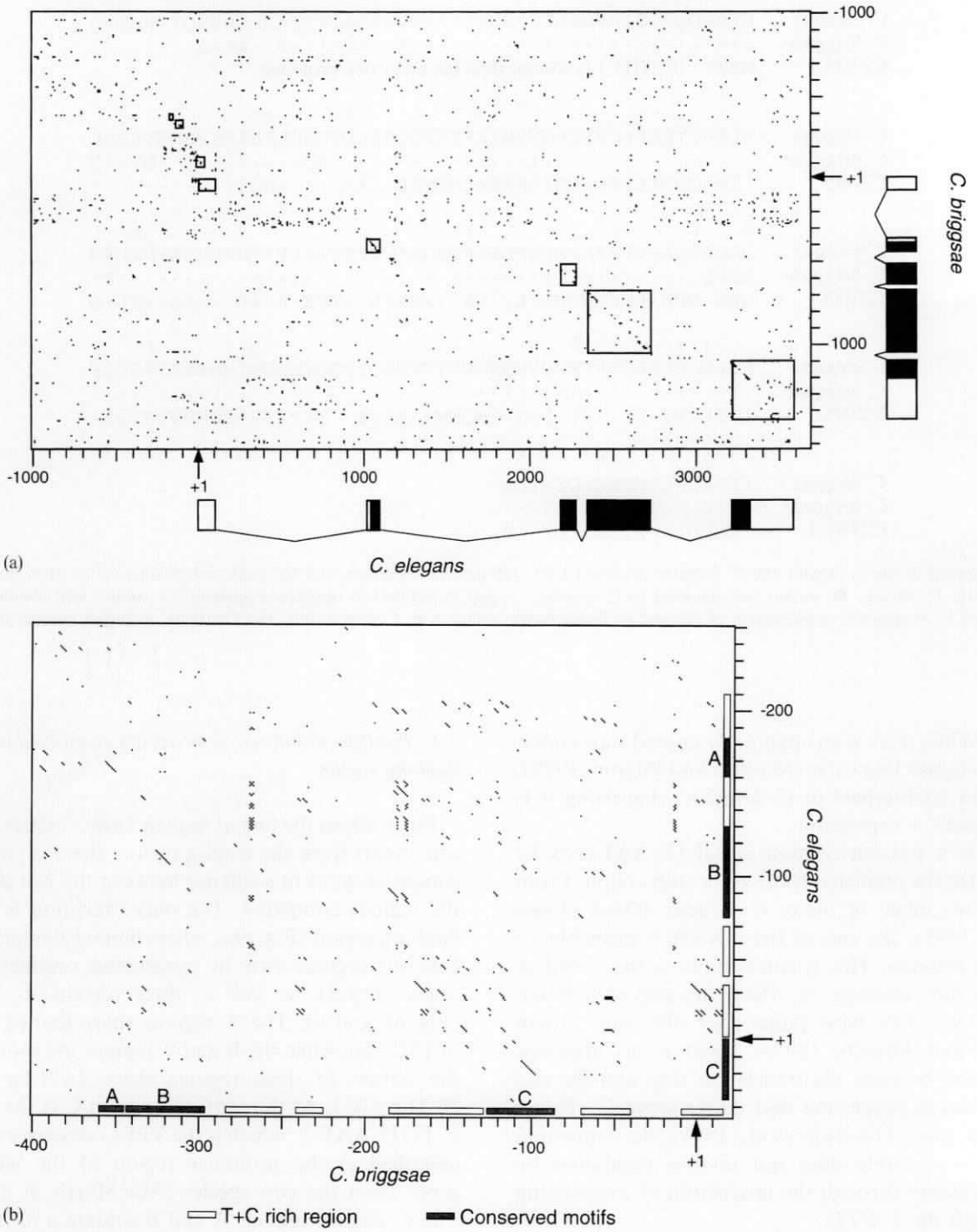
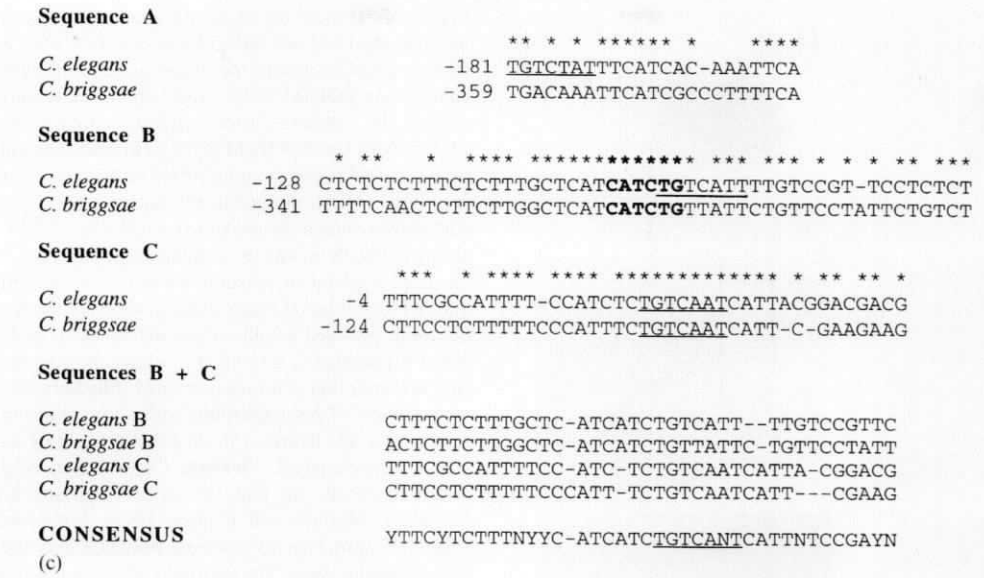


Fig. 6. Alignment of *unc-119* nt sequences from *C. elegans* and *C. briggsae*. Numbers refer to base pairs of nt sequences as shown in Fig. 4 (*C. briggsae*) and Maduro and Pilgrim (1995) (*C. elegans*). (**a**) Dot matrix alignment of entire coding and flanking regions. Boxed regions along the axis refer to exons of the genes; shaded boxes correspond to the protein coding regions of the transcript. Boxed regions within the matrix correspond to the overlap of exon sequences from the two species, as well as conserved 5′ flanking sequences. (**b**) Enlargement of 5′ flanking regions. Conserved regions A, B and C (see below) are highlighted along the axes by shaded boxes, pyrimidine-rich regions are highlighted by open boxes. (**c**) Similarity between the 5′ flanking regions of the *unc-119* gene. Numbers give distances 5′ to the start of transcription (+1). Asterisks indicate conserved base pairs, dashes where gaps were introduced to maximize alignment. Underlined sequences highlight 6 of 7 bp match to VPE1 *C. elegans* promoter element (MacMorris et al., 1994). Bold letters in region B match the 'E-box' sequence 5′-CATCTG-3′ (Krause et al., 1994). An alignment of sequences B and C, along with a consensus, is shown immediately below. **Methods:** Alignment was produced using DNA Strider 1.2 (Marck, 1988), using Stringency of 7, Window of 7 and Scale of 10 for panel **a**, and Stringency of 5, Window of 5 and Scale of 1 for panel **b**.



(c)

exon, which suggests that it may not be a promoter element at all. Without evidence for alternative transcriptional initiation, a compelling argument for the maintenance of region C cannot be made. None of these sequences is present in the immediate upstream region of the C27H5.1 ORF. The functional significance of these regions, as well as the significance of the VPE1 elements, has not been experimentally tested.

2.5. *Reporter gene expression is conserved between the two species*

Evidence for functional conservation of *cis*-acting sequences comes from comparison of reporter gene expression. The *C. elegans unc-119* null mutants ed3 and ed4, which contain nonsense mutations in exon IV (Maduro and Pilgrim, 1995), were made transgenic for both a reporter gene fusion and a *C. elegans* genomic *unc-119* clone, such that animals carrying the transgenes have a wild type phenotype. Animals containing a *C. elegans* or *C. briggsae unc-119* gene fusion to βgal were prepared as described (Fire et al., 1990). Photographs of differential interference contrast (DIC) microscopy of fixed and stained animals are shown in Fig. 7.

Temporal expression is very similar, as transgene activity is seen very early in the embryo and continues through adulthood. Spatially, staining is restricted mainly to the nervous system. Staining of neuronal structures, such as the ventral nerve cord, the pre-anal and lumbar ganglia, and nerve ring, is readily seen in both fusion strains. There is striking similarity at higher magnifications as well. In panels e and f, cells of the nerve ring and ventral nerve cord stain indistinguishably in placement and number of cells. Furthermore, some cells are seen to stain outside the nervous system, a phenomenon common to both constructs: cells are visible anterior to the anterior bulb of the pharynx, where no neuronal cell bodies are found (White et al., 1986). Both the *C. briggsae* and *C. elegans* photographs show staining similar to that previously reported for a *C. elegans unc-119* reporter gene (Maduro and Pilgrim, 1995).

To rule out differences due to amount of UNC-119 present in the fusions, both constructs were made to be very similar. The *C. briggsae* fusion contains approx. 620 bp of upstream DNA and 130 aa of UNC-119; the *C. elegans* fusion contains approx. 1160 bp of promoter and 101 aa of UNC-119. Both contain the conserved regions A, B and C. The similar expression pattern is consistent with the maintenance of the specificity of these *cis*-acting regions, and hence transcriptional control. Alternatively, transcription may occur at a basal level throughout the animal, and visualization of specific nervous system staining would arise from post-translational regulation, since the amino-terminal half of UNC-119 is present in both fusions. Such regulation might be maintained due to the high degree of amino acid similarity; however, when the same *C. elegans unc-119* promoter (without any coding sequences) is used to drive expression of another reporter, the green fluorescent protein (GFP) from *Aequoria victoria* (Chalfie et al., 1994), specific nervous system expression still occurs (data not shown), suggesting regulation of *unc-119::*reporter expression is primarily transcriptional.

It is possible that similar neuronal expression at the transcriptional level is due to smaller promoter elements scattered throughout the control region, but the conservation of position and sequence of the A, B and C elements, as well as the existence of a known transcriptional regulatory element at the core of regions B and C, make these good candidates for putative neuronal control elements.

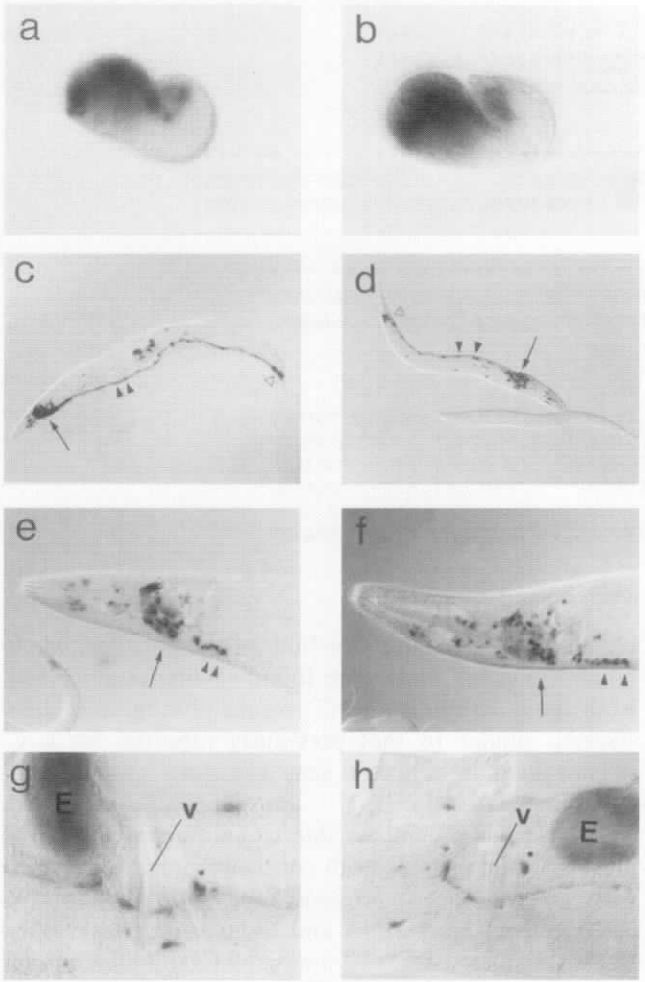


Fig. 7. X-gal staining of *C. elegans* hermaphrodites transgenic for *unc-119::*βgal fusions. Left column, *unc-119* (ed4); edEx42 (*C. briggsae* fusion; edEx42 contains the fusion plasmid pDP#MM082 and the rescuing clone pDP#MM016); right column [except (**h**)], *unc-119*(ed3); edEx43 (*C. elegans* fusion; edEx43 contains the fusion plasmid pDP#MM081 and pDP#MM016); (**h**) wild-type animal transgenic for pRF4 (rol-6^{D} marker) and pDP#MM082. Despite the nuclear localizing signal (NLS) present in the reporter gene constructs, staining is still visible outside of nuclei (as verified by DAPI staining; data not shown). Panels **a** and **b**, comma-stage embryos showing anteriorly localized staining in presumptive neuronal cell precursors; (**c**) adult and (**d**) fourth larval stage animals, showing staining in the nerve ring (arrows), pre-anal ganglia (open arrowheads) and ventral nerve cord (filled arrowheads); (**e**) and (**f**), adults showing staining in the nerve ring (arrows) and ventral nerve cord (filled arrowheads); (**g**) and (**h**), ventral view of hermaphrodite vulva (V) showing the ventral nerve cord and a VC neuron (*). A diffusely staining early embryo (E) is visible in each panel. **Methods:** Construction of pDP#MM081, predicted to encode the first 101 aa of *C. elegans* UNC-119, has been described (Maduro and Pilgrim, 1995). For construction of the *C. briggsae* fusion, two intermediate plasmids were used to generate suitable restriction sites. The plasmid UR#91 (a gift from W. Wadsworth) contains an *Eco*RI-*Hin*dIII fragment cloned into pBluescript SK+; this fragment was removed from the polylinker of pUC19 which contained a 120 bp insertion at the *Bgl*II site. UR#91 was digested with *Eco*RI and *Bgl*II (both enzymes BRL) to liberate the 120 bp insertion and part of the pUC polylinker, and allow the insertion of a *Bam*HI-*Eco*RI fragment from *C. briggsae*. This intermediate plasmid, pDP#MM070, was digested with *Hin*dIII (Pharmacia) and *Pst*I (BRL) to allow its subcloning into similarly digested pPD22.04 (Fire et al., 1990), which is predicted to join the first 130 aa of *C. briggsae* UNC-119 with βgal.

3. Conclusion

The *unc-119* genes from *C. elegans* and *C. briggsae* provide another example in which conservation of coding sequence, coupled with maintenance of only a small number of control elements, allow similarity of function and expression between two species that are separated by tens of millions of years of evolution. Here we have demonstrated that inter-species sequence comparisons provide a means by which homologous genes and candidate promoter elements may be identified.

Acknowledgement

We wish to thank Terry Snutch and David Baillie for providing the *C. briggsae* genomic library, and Andrew Fire for donating fusion vectors. This work was supported by a grant from the Natural Science and Engineering Research Council.

References

Berks, M. and the *C. elegans* genome mapping and sequencing consortium (1995) The *C. elegans* genome sequencing project. Genome Res. 5, 99–104.

Burns, N., Grimwade, B., Ross-Macdonald, P.B., Choi, E.-Y., Finberg, K., Roeder, G.S. and Snyder, M. (1994) Large-scale analysis of gene expression, protein localization, and gene disruption in *Saccharomyces cerevisiae*. Genes Dev. 8, 1087–1105.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.C. (1994) Green fluorescent protein as a marker for gene expression. Science 263, 802–805.

Fire, A., Harrison, S.W. and Dixon, D. (1990) A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. Gene 93, 189–198.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specifc oligonucleotide primer. Proc. Natl. Acad. Sci. USA 85, 8998–9002.

Goodwin, E.B., Okkema, P.G., Evans, T.C. and Kimble, J. (1993) Translational regulation of *tra-2* by its 3′ untranslated region controls sexual identity in *C. elegans*. Cell 75, 329–339.

Heschl, M.F.P. and Baillie, D.L. (1990) Functional elements and domains inferred from sequence comparisons of a heat shock gene in two nematodes. J. Mol. Evol. 31, 3–9.

Hodgkin J., Plasterk, R.H.A. and Waterston, R.H. (1995) The nematode *Caenorhabditis elegans* and its genome. Science 270, 410–414.

Jones, D. and Candido, E.P.M. (1993) Novel ubiquitin-like ribosomal protein fusion genes from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. J. Biol. Chem. 268, 19545–19551.

Kennedy, B.P., Aamodt, E.J., Allen, F.L., Chung, M.A., Heschl, M.F.P. and McGhee, J.D. (1993) The gut esterase gene (*ges-1*) from

the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. J. Mol. Biol. 229, 890–908.

Krause, M. (1995) Transcription and translation. Methods Cell Biol. 48, 483–512.

Krause, M., Harrison, S.W., Xu, S.-Q., Chen, L. and Fire, A. (1994) Elements regulating cell- and stage-specific expression of the *C. elegans* MyoD family homolog *hlh-1*. Dev. Biol. 166, 133–148.

Kuwabara, P.E. and Shah, S. (1994) Cloning by synteny: identifying *C. briggsae* homologues of *C. elegans* genes. Nucleic Acids Res. 22, 4414–4418.

Lee, Y.H., Huang, X.Y., Hirsh, D., Fox, G.E. and Hecht, R.M. (1992) Conservation of gene organization and trans-splicing in the glyceraldehyde-3-phosphate dehydrogenase-encoding genes of *Caenorhabditis briggsae*. Gene 121, 227–235.

MacMorris, M., Spieth, J., Madej, C., Lea, K. and Blumenthal, T. (1994) Analysis of the VPE sequences in the *Caenorhabditis elegans vit-2* promoter with extrachromosomal tandem array-containing strains. Mol. Cell. Biol. 14, 484–491.

Maduro, M. and Pilgrim, D. (1995) Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. Genetics 141, 977–988.

Marck, C. (1988) 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. 16, 1829–1836.

Mello, C.C., Kramer, J.M., Stinchcomb, D. and Ambros, V. (1991) Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J. 10, 3959–3970.

Pilgrim, D., McGregor, A., JŠckle, P., Johnson, T. and Hansen, D. (1995) The *C. elegans* sex determining gene *fem-2* encodes a putative protein phosphatase. Mol. Biol. Cell 6, 1159–1171.

Prasad, S.S. and Baillie, D.L. (1989) Evolutionarily conserved coding sequences in the *dpy-20-unc-22* region of *Caenorhabditis elegans*. Genomics 5, 185–198.

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

Snutch, T.P. (1984) Ph.D. Thesis, Simon Fraser University, Burnaby, BC.

Southern, E. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503–517.

Spieth, J., Denison, K., Kirtland, S., Cane, J. and Blumenthal, T. (1985) The *C. elegans* vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes. Nucleic Acids Res. 13, 5283–5295.

White, J.G., Southgate, E., Thomson, J.N. and Brenner, S. (1986) The structure of the nervous system of *Caenorhabditis elegans*. Phil. Trans. R. Soc. Lond. B Biol. Sci. 314, 1–340.

Zucker-Aprison, E. and Blumenthal, T. (1989) Potential regulatory element of nematode vitellogenin genes revealed by interspecies sequence comparison. J. Mol. Evol. 28, 487–496.