Mol Gen Genet (1993) 236:289–298

MGG

Molecular analysis of two genes between *let-653* and *let-56* in the *unc-22(IV)* region of *Caenorhabditis elegans*

Marco A. Marra, Shiv S. Prasad*, and David L. Baillie

Institute for Molecular Biology and Biochemistry, Department of Biological Sciences, Simon Fraser University, Burnaby, B.C., Canada, V5A 1S6

Received July 10, 1992 / Accepted July 31, 1992

Summary. A previous study of genomic organization described the identification of nine potential coding regions in 150 kb of genomic DNA from the *unc-22(IV)* region of *Caenorhabditis elegans*. In this study, we focus on the genomic organization of a small interval of 0.1 map unit bordered on the right by *unc-22* and on the left by the left-hand breakpoints of the deficiencies *sDf9*, *sDf19* and *sDf65*. This small interval at present contains a single mutagenically defined locus, the essential gene *let-56*. The cosmid C11F2 has previously been used to rescue *let-56*. Therefore, at least some of C11F2 must reside in the interval. In this paper, we report the characterization of two coding elements that reside on C11F2. Analysis of nucleotide sequence data obtained from cDNAs and cosmid subclones revealed that one of the coding elements closely resembles aromatic amino acid decarboxylases from several species. The other of these coding elements was found to closely resemble a human growth factor activatable Na^+/H^+ antiporter. Pairs of oligonucleotide primers, predicted from both coding elements, have been used in PCR experiments to position these coding elements between the left breakpoint of *sDf19* and the left breakpoint of *sDf65*, between the essential genes *let-653* and *let-56*.

Key words: *Caenorhabditis elegans* – Genomic organization – Na^+/H^+ antiporter – Dopa decarboxylase

Introduction

Understanding the mechanisms by which contiguous blocks of genes are regulated during eukaryotic development requires knowledge of the identity and organization of genes in the genome. With the ultimate aim of identifying the gene content of a specific region, the 2 map unit *unc-22(IV)* region, defined by the deficiency *sDf2* in *Caenorhabditis elegans*, has been subjected to extensive genetic analysis in our laboratory (Rogalski and Baillie 1985; Clark et al. 1988; Clark 1990; Clark and Baillie 1992). Screens for recessive lethal mutations have resulted in the identification of 36 essential genes between the breakpoints of *sDf2* (Clark 1990; Clark and Baillie 1992). Based on this figure, a truncated Poisson calculation predicts that this region is approximately 70% saturated for mutations in essential genes (Clark 1990; Clark and Baillie 1992). A number of overlapping deficiencies have been isolated, which subdivide the region into smaller zones (Clark 1990; Clark and Baillie 1992). These features, and the availability of overlapping cosmid clones (Coulson et al. 1986, 1988), provided the framework for a detailed molecular characterization of the region.

In this paper, we focus on the identification and characterization of coding elements located in a 0.1 map unit (m.u.) region between *let-653* and *let-56* on linkage group (LG)IV. This region contains the left-hand breakpoints of the deficiencies *sDf9*, *sDf19* and *sDf65*, and at present contains only one mutagenically defined locus, the essential gene *let-56*. This locus is deleted by all three of these deficiencies (Clark and Baillie 1992). A cDNA, recovered using a subclone of the cosmid C11F2 as a probe, had been mapped to the region deleted by the deficiency *sDf9* (Prasad and Baillie 1989). Subsequently, Clark and Baillie (1992) demonstrated that the cosmid C11F2 could rescue *let-56* mutant individuals; therefore, at least part of C11F2 had to lie between *let-653* and *unc-22*. By using interspecies hybridization to the related species *C. briggsae* and subsequent Northern blot analyses, Prasad and Baillie (1989) had mapped five transcripts to C11F2. Four of these transcripts mapped to adjacent *Pst*I fragments, denoted "Q" and "R". Here, we report nucleotide sequence data from cosmid subclones and cDNAs isolated using these adjacent *Pst*I fragments as probes, and assign tentative identities to two genes based on similarities to sequences in the SWISS and EMBL databanks. In addition, using our sequence data and

* *Present address:* Department of Ophthalmology, University of British Columbia, Room 330 2550 Willow St., Vancouver, B.C., Canada, V5Z 3N9

Correspondence to: M.A. Marra

polymerase chain reaction (PCR) methodology (Barstead and Waterston 1991), these two genes have been positioned between *let-653* and *let-56*.

Materials and methods

The genetic nomenclature follows the recommendations of Horvitz et al. (1979). The cosmid C11F2 was obtained from A. Coulson and J. Sulston of the Medical Research Council, Cambridge, UK.

Nematode strains and culture conditions. Nematodes were maintained on petri plates containing nematode growth medium (NGM) streaked with *Escherichia coli* OP50 (Brenner 1974). The wild-type strain N2 (var. Bristol) was obtained from the stock collection at the Medical Research Council, Cambridge, UK. The deficiency *sDf19* was isolated by Moerman (1980) and positioned by Rogalski and Baillie (1985). *sDf19* has a dominant twitcher phenotype and is homozygous lethal, usually arresting development at the embryonic stage; however, *sDf19* homozygous animals will occasionally hatch. *sDf19* breaks in the *unc-22* gene and extends to the left of *let-56* (Rogalski and Baillie 1985), but does not delete *let-653* (Clark and Baillie 1992; Fig. 1). The deficiency *sDf65* was isolated by D.V. Clark (Clark and Baillie 1992). *sDf65* deletes *let-56* but not *let-653* on the left, and extends as far right as *let-52* (Clark and Baillie 1992). Animals homozygous for this deficiency are arrested as L1 larvae (Clark and Baillie 1992).

Labelling of DNA probes. DNA fragments were separated on 0.7% Seaplaque (FMG) gels containing 0.5 μg/ml ethidium bromide, and excised over 360 nm ultraviolet light. Approximately 10 to 50 ng of isolated DNA was labelled by the oligolabelling technique described by Feinberg and Vogelstein (1983). Probes with a specific activity of 1.0×10^8 to 2.0×10^9 cpm/μg were synthesized by this procedure. Probes were denatured by immersion in a boiling water bath for 10 min prior to addition to the hybridization solution.

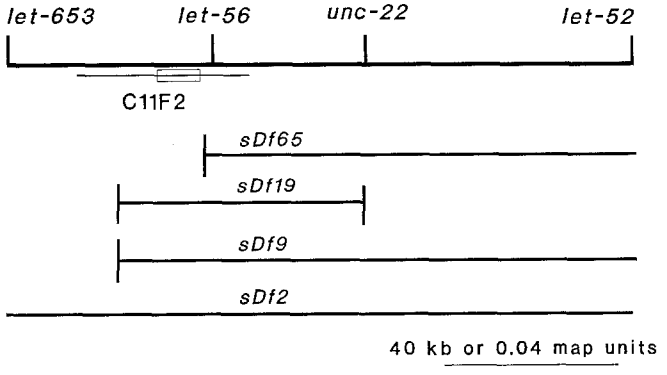

Fig. 1. Genetic map of the *let-653–let-52* interval. Genes are indicated *above* the heavy horizontal line. Deficiencies subdividing the region are shown *below* the genetic map. The cosmid C11F2 is shown beneath *let-56;* the *open box* shows the region on C11F2 that contains the antiporter-like gene and the decarboxylase-like gene

Hybridization of probes to DNA filters. DNA filters were prepared as previously described (Prasad and Baillie 1989). Filters were prehybridized in 5 × SSPE, 0.2% sodium dodecyl sulphate (SDS) and 5 × Denhardt's at the same temperature used during hybridization [1 × SSPE is 0.18 M NaCl, 10 mM sodium phosphate, 1 mM EDTA, pH 7.4; 1 × Denhardt's is 0.02% bovine serum albumin (BSA), 0.02% Ficoll and 0.02% polyvinylpyrrolidone]. Hybridization with [^{32}P]dATP labelled probe was carried out in a fresh aliquot of the above solution at 68° C for 16 h. Filters were washed in 0.1 × SSPE, 0.2% SDS at the same temperature. Filters were exposed to preflashed Kodak X-OMAT K film at −70° C for the required period (4 h to 3 days).

Polymerase chain reaction (PCR) primers designed and used in this study. Oligonucleotide primers 18–20 bp in length were designed using the program OLIGO (Rychlik and Rhoads 1989). "Control" primers designed from sequences on LGV were provided by K. McKim. Control primers 265F and 265R were used in reactions involving *sDf19* homozygotes. Control primers K087 and K083 were used in reactions involving *sDf65* homozygotes. Primers R1 and R11 were predicted from our dopa decarboxylase-like sequence (Fig. 3). Primers MAM2 and DB19 were predicted from our Na^+/H^+ antiporter-like sequence (Fig. 4). Primers QFOR1 and QREV1 were predicted from fragment Q sequence generated using M13 −40 or reverse primers, respectively. Primers RF1 and RR1 were predicted from fragment R sequence generated using M13 −40 or reverse primers, respectively. The sequences of the oligonucleotide primers are (5′ to 3′): R1,CGGCGGAGAAGAACAAGC; R11, TACTCGCATCTTTACCATCG; MAM2, AGGTTTCTCAGTTTTTGG; DB19, GCTATGATGGAGGAAAGTG; QFOR1, CTCAACGTCATCAGCAGATC; QREV1, TTTATCCCATTTTTCGCTAC; RF1, AATCAGAGGAGACCGAAG; RR1, TGACTTTTTGATGCCTGC.

Ordering fragment Q with respect to fragment R on C11F2. Fragment Q and fragment R (Fig. 2) were cloned into the *Pst*I site of pUC19. The ends of Q and R were sequenced using commercially available M13 reverse and −40 primers. A primer predicted from each end of fragment Q was used in conjunction with a primer predicted from each end of R in pairwise PCR experiments using approximately 1 ng of C11F2 DNA as the amplification template. Only the primers QFOR1 and RF1, designed from the sequences obtained using the M13 −40 primers, produced an amplification product, thus unambiguously ordering Q with respect to R on C11F2. PCR parameters were 94° C (5 min), 58° C (1 min), 72° C (3 min) for one cycle, followed by 25 cycles of 94° C (1 min), 58° C (1 min), 72° C (3 min), and one cycle of 94° C (1 min), 58° C (1 min) and 72° C (10 min). All PCR reagents were obtained from Perkin-Elmer Cetus.

Screening of a C. elegans *cDNA library.* Approximately 3.0×10^5 phage from a lambda ZAP (Stratagene) *C. elegans* cDNA library (Barstead and Waterston 1989) were screened with oligolabelled fragment R (Prasad and

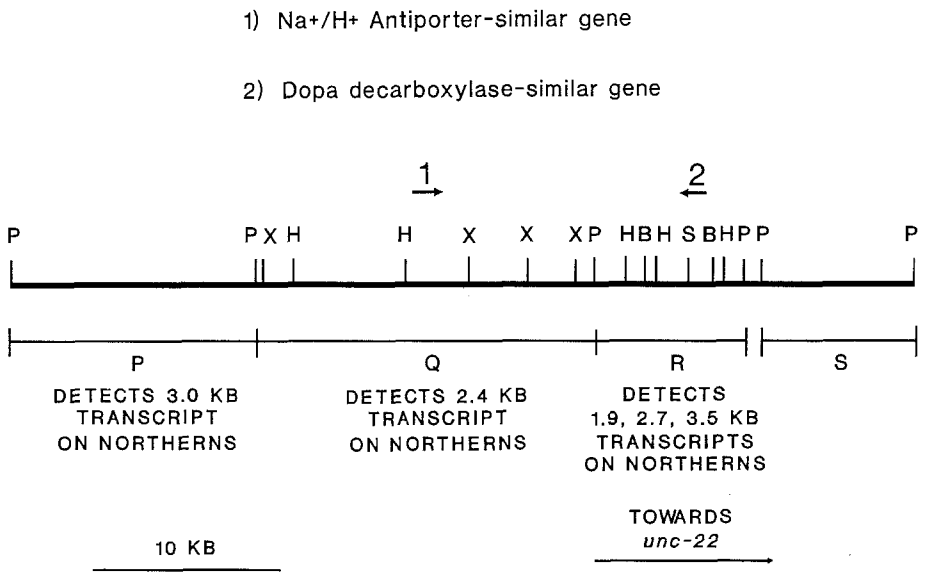

Fig. 2. Restriction enzyme map of fragments Q and R. Fragments P and S are shown for reference, but were not mapped in this study. The directions of transcription of the *Caenorhabditis elegans* Na^+/H^+ antiporter-like gene and the *C. elegans* 3,4-dihydroxyphenylalanine (dopa) decarboxylase-like gene are indicated by *arrows* above the restriction map. P, *Pst*I; X, *Xba*I; H, *Hin*dIII; B, *Bam*HI; S, *Sph*I

Baillie 1989) as a probe. A single positive plaque was purified after three screens using the method of Benton and Davis (1977). The cDNA was recovered in the *Eco*RI site of the pBluescript SK − phagemid following protocols obtained from Stratagene. Approximately 5.0×10^5 phage from the same library were screened with the C2 partial cDNA originally recovered by Prasad and Baillie (1989) using fragment Q as a probe. Six cDNAs were isolated; sequence analysis of the ends of these cDNAs revealed that the longest cDNA (2.2 kb) was a fusion cDNA. The second longest cDNA (designated 7 M) was 1.9 kb in length, and contained an additional 600 bp of Na^+/H^+ antiporter sequence as well as containing all of the C2 cDNA. The 600 bp fragment (designated 7 M–5′) was gel isolated after digestion with *Eco*RI, labelled, and used to screen additional 3.0×10^5 phage from the same cDNA library. Six cDNAs were isolated as a result of this screen, and all were determined to represent either the 7 M cDNA, or the fusion cDNA previously isolated. 7 M–5′ was used to identify an 11 kb *Xba*I subclone of Q designated X611. X611 is expected to contain the remainder of the coding sequence for the Na^+/H^+ antiporter.

Determining the direction of the C. elegans *Na^+/H^+ antiporter reading frame.* The ends of X611 were sequenced with T3 and T7 primers. The sequence obtained with T3 primer overlapped the sequence obtained from fragment Q using M13 reverse primer. The complement of the X611 sequence obtained with T7 primer was found to match the *C. elegans* Na^+/H^+ antiporter cDNA sequence. This information, along with the orientation of Q with respect to R, shows that the *C. elegans* Na^+/H^+ antiporter is transcribed from left to right on the genetic map, opposite to the direction of *unc-22* transcription.

Determining the direction of the C. elegans *dopa decarboxylase reading frame.* The dopa decarboxylase cDNA was found to have a single *Sph*I restriction site. The pBluescript SK − vector in which the cDNA was cloned contains no *Sph*I site. Double digestion of the cDNA with *Sph*I and *Eco*RI produces three fragments, one of which (3 kb) corresponds to the vector. A 1200 bp fragment represents the 3′ portion of the cDNA. A 700 bp fragment represents the 5′ portion of the decarboxylase cDNA. Digestion of fragment R with *Sph*I produces two bands on agarose gels, because fragment R has an internal *Sph*I site, and there is a *Sph*I site in the pUC19 polylinker. One of these bands is 3.5 kb and the other is 8.0 kb in length. Southern blotting of the *Sph*I-digested fragment R DNA onto nylon membranes followed by hybridization with the labelled 5′-specific decarboxylase *Sph*I-*Eco*RI restriction fragment and subsequent autoradiography demonstrated that the 3.5 kb *Sph*I restriction fragment contained the decarboxylase 5′ end (data not shown). This result indicates that the direction of transcription is from right to left on the genetic map, in the same direction as *unc-22* and opposite to that of the *C. elegans* Na^+/H^+ antiporter.

DNA sequence determination. The 1300 bp C2 cDNA isolated by Prasad and Baillie (1989) was used to construct deletion derivatives for DNA sequencing using the procedure described by Henikoff (1987). The additional 600 bp of Na^+/H^+ antiporter cDNA obtained in screens of the Barstead and Waterston (1989) lambda ZAP cDNA library using C2 as a probe were sequenced using custom designed oligonucleotide primers, which were synthesized on a model 391 ABI oligonucleotide synthesizer using standard phosphoramidite chemistry. The entire cDNA detected using fragment R as a probe was likewise sequenced using custom designed oligonucleotide primers. Fragments Q and R were subcloned into the *Pst*I site of pUC19, and their ends were sequenced using commercially available M13 forward and reverse primers. Approximately 2 to 5 μg of supercoiled plasmid DNA was obtained either by gel purification followed by phenol extraction, or by centrifugation through Miniprep columns (Pharmacia), and subjected to the dideoxy sequencing method of Sanger et al. (1977). Purified tem-

plates were sequenced as described by Hattori and Sakaki (1986) using Sequenase reagents (USB) or following the recommendations of T. Snutch (personal communication). Genomic DNA sequences corresponding to *C. elegans* Na^+/H^+ antiporter and decarboxylase sequences were determined to confirm the cDNA sequences. This was achieved by applying the same custom designed oligonucleotide primers, determined from the cDNA sequence, to fragments Q and R or by sequencing deletion subclones of clone X611 or fragment R constructed using the Erase-a-base kit (Promega). Deletion subclones were sequenced on an ABI automated sequencing machine using reagents and protocols supplied by ABI.

DNA sequence analysis. Nucleotide sequences were aligned, formatted, and translated using the Eyeball Sequence Editor (ESEE) program (Cabot and Beckenbach 1989). Amino acid sequences predicted using ESEE were used to search the latest available releases of the SWISS (release 17) and translated EMBL (release 25) databanks. Searches were done with the FASTA and TFASTA programs of Pearson and Lipman (1988). Hydrophobicity plots of the putative *C. elegans* Na^+/H^+ antiporter were generated using TGREASE (Pearson and Lipman 1988).

PCR using worms homozygous for deficiencies. PCR primers designed from *C. elegans* Na^+/H^+ antiporter and *C. elegans* dopa decarboxylase sequences were applied to developmentally arrested animals that were homozygous for the deficiences *sDf19* or *sDf65*. The primers MAM2 and DB19 were used in PCR reactions involving animals homozygous for the deficiency *sDf19*. These primers amplify a 600 bp band in reactions involving wild-type animals. This band is not seen in PCR involving *sDf19* homozygotes, indicating the absence of a priming site(s). The primers R1 and R11 were used in PCR reactions involving animals homozygous for the deficiency *sDf65*. These primers amplify a band of approximately 800 bp in reactions involving wild-type animals. A band of this size is seen in reactions involving *sDf65* animals, indicating the presence of both priming sites. Five arrested embryos (*sDf19*) or larvae (*sDf65*) were used as the amplification template for PCR experiments involving these deficiencies. PCR reactions were performed following a whole-worm protocol devised by Barstead and Waterston (1991), suitably modified for embryos (R.J. Barstead, personal communication) and using reagents supplied in the GeneAmp PCR kit (Perkin Elmer Cetus). Reaction parameters were as previously described (see above section on ordering of fragment Q with respect to fragment R on C11F2).

Results

Fragment R encodes a decarboxylase

The 8 kb fragment R had detected three mRNAs of 1.9, 2.7 and 3.5 kb in length in Northern blot analyses (Prasad and Baillie 1989; Fig. 2). We have used R as a probe to screen a *C. elegans* cDNA library (Barstead and Waterston 1989) and have identified a single positive plaque out of 3.0×10^5 screened. This cDNA was recovered and subjected to restriction enzyme and sequence analyses. The cDNA is 1924 bp in length, in agreement with the result from Northern analysis of 1.9 kb, and contains a single long open reading frame having the coding potential for 624 amino acids. The cDNA may be incomplete, as three potential exons have been detected in the genomic DNA immediately upstream from the 5′ end of the cDNA. The direction in which the coding element represented by the cDNA is transcribed has been inferred by analysing the cosmid and cDNA restriction maps (see the Materials and methods and Fig. 2). The gene is transcribed from right to left on the genetic map, in the same direction as *unc-22* (Benian et al. 1989). A TGA stop codon and a consensus poly(A) addition site are identifiable, commencing at positions 1876 and 1907, respectively, in the nucleotide sequence (Fig. 3).

The predicted amino acid sequence was used to search the SWISS (release 17) and translated EMBL (release 25) databanks using the FASTA or TFASTA programs, respectively (Pearson and Lipman 1988). Significant similarities to a variety of pyridoxal cofactor-dependent amino acid decarboxylases were detected. Human, rat and *Drosophila melanogaster* dopa decarboxylases were found to be 39.6%, 39.6% and 37% identical to the *C. elegans* sequence over 470, 465 and 477 amino acids, respectively. Rat and human histidine decarboxylases were discovered to be 38.9% and 39.1% identical to our sequence over 465 and 468 amino acids, respectively. A comparison of the human dopa decarboxylase protein and the *C. elegans* protein is presented in Fig. 3. Regions of amino acid similarity are restricted to a stretch of 470 amino acids near the amino-terminal end of our sequence. The first 70 amino acids and also the remaining carboxy-terminal 128 amino acids had no significant similarity to any of a number of decarboxylases in the SWISS databank.

A potential pyridoxal cofactor-binding site is located at position 368 in the predicted protein (Fig. 3). This putative binding site contains a perfectly conserved lysine residue, which functions as the site of attachment for the pyridoxal cofactor in pyridoxal cofactor-dependent decarboxylases. Immediately preceding this lysine residue are proline and serine residues, which are conserved in human and rat histidine decarboxylases. In human, *Drosophila* and rat dopa decarboxylases, the serine is replaced by a histidine residue.

Fragment Q encodes a Na^+/H^+ antiporter

Immediately to the left of fragment R on C11F2 is the 18 kb *Pst*I fragment Q (Prasad and Baillie 1989). Fragment Q had detected a 2.4 kb mRNA when used to probe Northern blots of *C. elegans* RNA (Prasad and Baillie 1989; Fig. 2). The use of Q to screen a lambda gt10 cDNA library (provided by B. Meyer) had resulted in the identification of a 1.3 kb cDNA, which was named C2 (Prasad and Baillie 1989). Our nucleotide sequence analy-

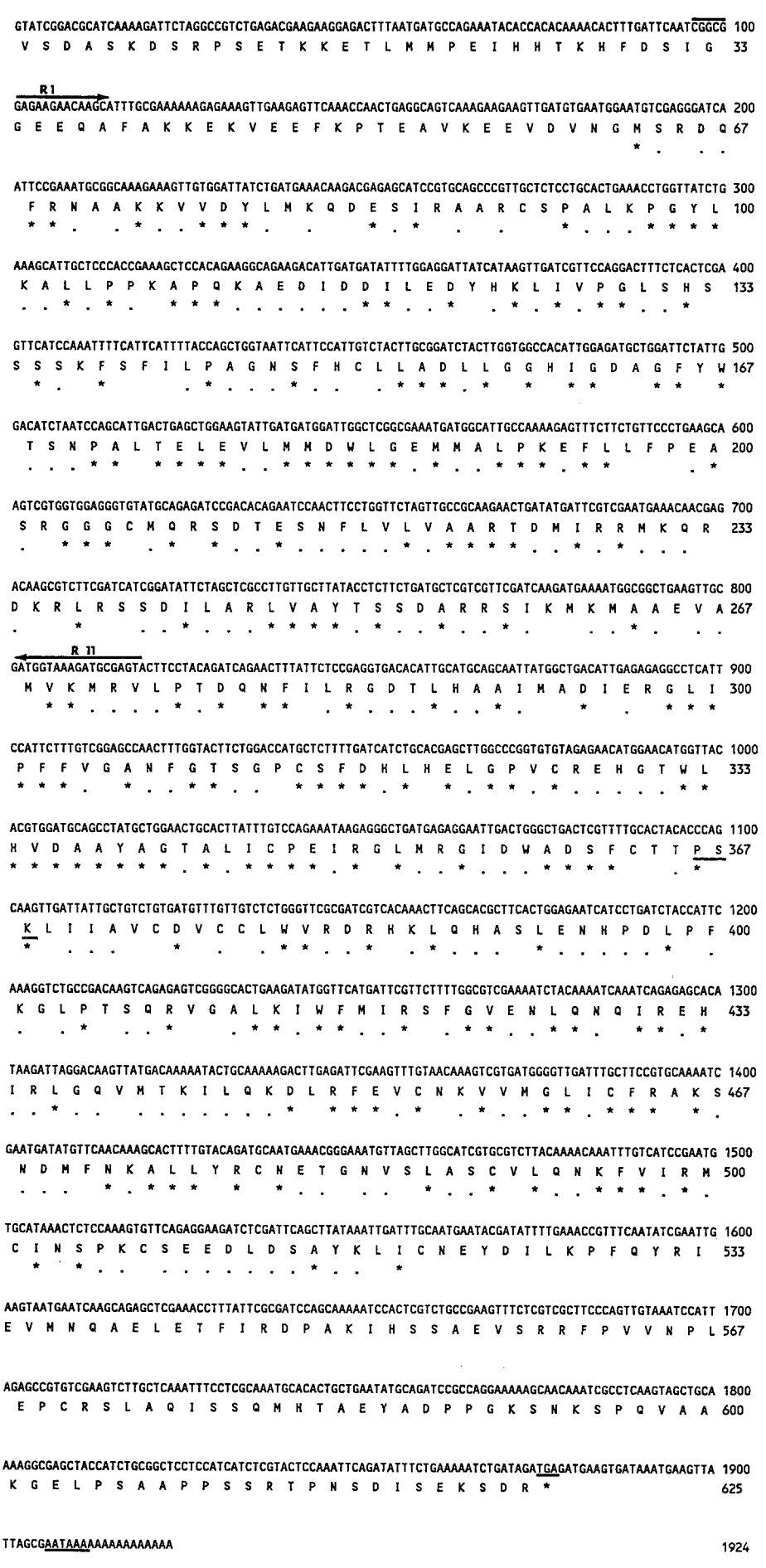

Fig. 3. The *C. elegans* dopa decarboxylase-like cDNA and inferred amino acid sequence. Amino acid residues identical to human dopa decarboxylase residues are indicated by an *asterisk*. A *dot* indicates a conservative amino acid replacement. The potential pyridoxal cofactor-binding site at amino acid positions 366–368 is *underlined*. A potential TGA stop codon and polyadenylation signal are likewise indicated in the nucleotide sequence. The positions of oligonucleotide primers R1 and R11 are indicated by *arrows* above the nucleotide sequence

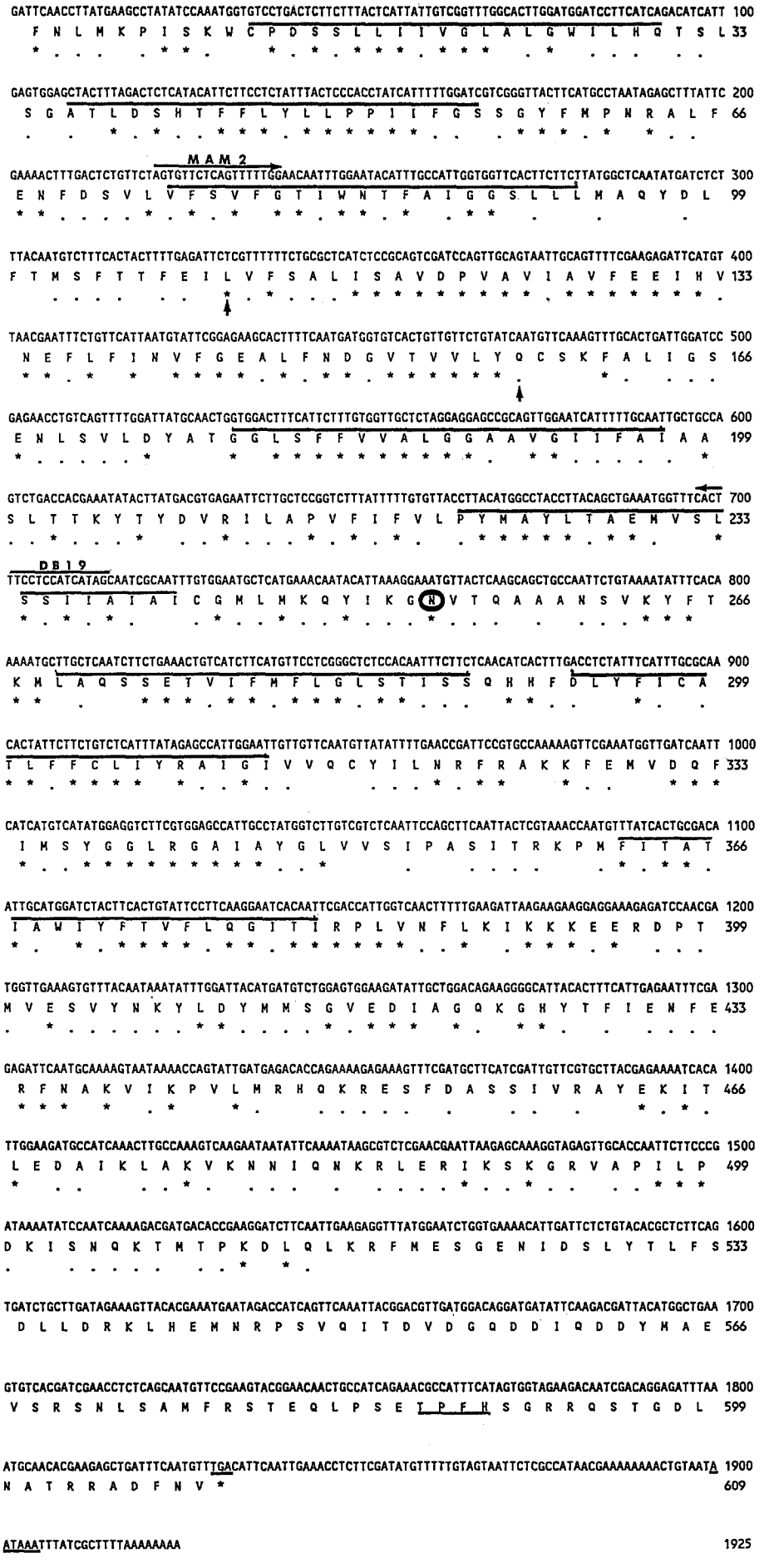

Fig. 4. *C. elegans* Na^+/H^+ antiporter-like cDNA and inferred amino acid sequence. Amino acid residues identical to human antiporter residues are indicated by an *asterisk*. A *dot* indicates a conservative amino acid replacement. The *circled residue* at amino acid position 253 is a conserved putative glycosylation site. *Underlined* amino acid residues at positions 586–589 indicate a potential phosphorylation site. A stop codon and polyadenylation signal are *underlined* in the nucleotide sequence. Regions of the *C. elegans* protein aligning with human antiporter transmembrane domains are indicated by *lines above* the amino acid sequence. The positions of oligonucleotide primers MAM2 and DB19 are indicated by *arrows* above the nucleotide sequence. The portion of the protein between the arrowheads (between the third and fourth transmembrane domains) exhibits a high degree of identity to a region of the human antiporter protein predicted to extend extracellularly

sis of this cDNA has revealed it to be incomplete, representing the 3′ portion of a molecule possessing significant similarity to a human growth factor activatable Na^+/H^+ antiporter (Sardet et al. 1989). The 1.3 kb cDNA was used as a probe to screen another cDNA library (Barstead and Waterston 1989), from which twelve cDNAs were recovered in screens of approximately 8.0×10^5 plaques. Restriction enzyme analysis of these cDNAs demonstrated the largest to be approximately 1.9 kb in length. The direction of transcription of the gene represented by these cDNAs has been inferred by analysing the restriction map of fragment Q and by DNA sequence analysis (see the Materials and methods). The gene is transcribed from left to right on the genetic map, opposite to the direction of transcription of the decarboxylase gene located on fragment R (Fig. 2). Nucleotide sequence analysis of the 1.9 kb cDNA revealed the cDNA to contain all of the previously isolated 1.3 kb C2 cDNA, and an additional 600 bp extending further towards the 5′ end of the gene. The cDNA is 1925 nucleotides in length, and has the potential to encode a protein of 609 amino acids. Based on the Northern result of 2.4 kb achieved by Prasad and Baillie (1989) and the lack of a consensus initiation site for translation, we believe the 1.9 kb cDNA to be incomplete. A TGA stop codon and a single consensus poly(A) addition site are located at nucleotide positions 1830 and 1900, respectively (Fig. 4). The predicted amino acid sequence was used to search releases of both the SWISS (rel.17) and translated EMBL (rel.25) databanks using the FASTA or TFASTA programs of Pearson and Lipman (1988). Our sequence is 39.5% identical over 521 amino acids to a human Na^+/H^+ antiporter. When conserved amino acid replacements are taken into consideration, our sequence is 79.5% similar to the human antiporter (Fig. 4).

Sardet et al. (1989) predicted the human Na^+/H^+ antiporter to contain ten transmembrane domains. The sequence we have obtained from *C. elegans* shows extensive identity to eight of these transmembrane domains (Fig. 4). In the *C. elegans* sequence the number of amino acids that reside in these predicted transmembrane domains totals 160 residues. Of these, 84 are identical to residues in the transmembrane portions of the human protein. Therefore, the percentage identity between the *C. elegans* sequence and the human sequence is 52.5% in the predicted transmembrane regions. When conserved amino acid replacements in these domains are considered, the two amino acid sequences are 84.4% similar (Fig. 4). The *C. elegans* sequence also shows strong similarity to a region of the human protein predicted by Sardet et al. (1989) to extend extracellularly. This region contains *C. elegans* amino acid residues 110 to 157 (Fig. 4). Seventy-three percent (35 of 48) of the amino acids in this region are identical in both sequences. When conserved amino acid replacements are considered, the two sequences are 96% similar in this domain. In addition, one of the putative glycosylation sites predicted in the human Na^+/H^+ antiporter sequence (Sardet et al. 1989) is conserved in the *C. elegans* protein (Fig. 4). The carboxy-terminal region of the human Na^+/H^+ antiporter

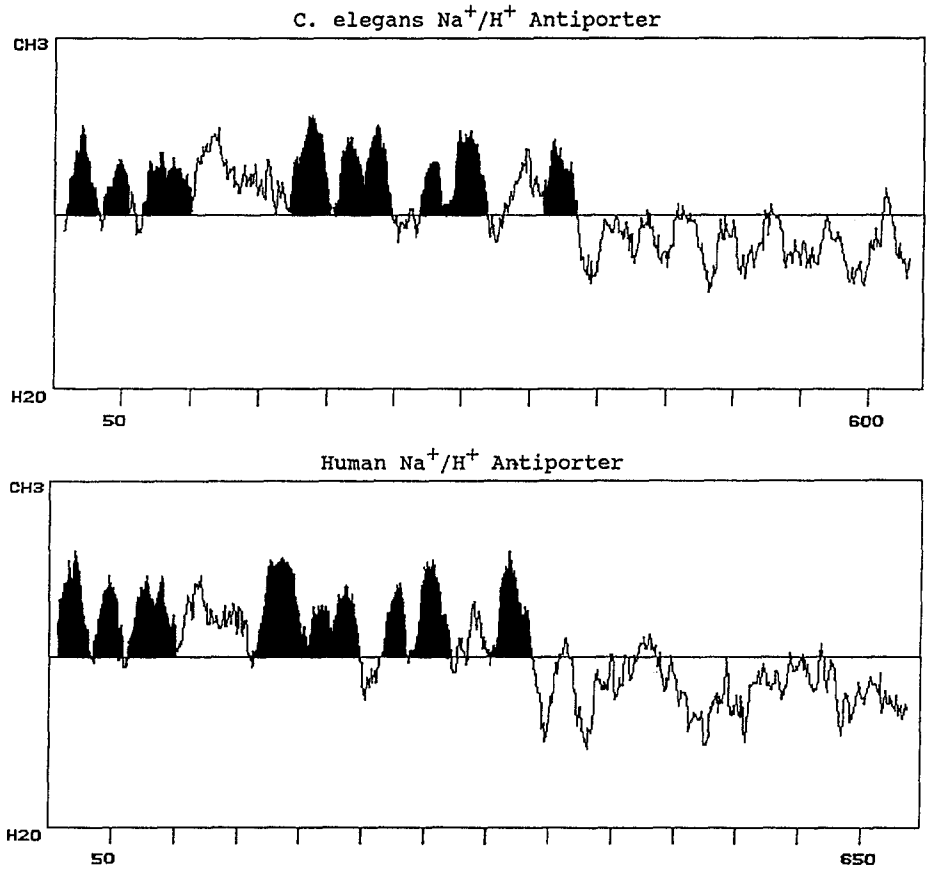

Fig. 5. TGREASE hydropathy profiles of the putative *C. elegans* Na^+/H^+ antiporter and the human Na^+/H^+ antiporter. The first 120 amino acids of the human amino acid sequence have been omitted from this comparison, as we do not have the corresponding sequence from *C. elegans*. The *blackened areas* of the profile indicate the regions of the human protein predicted by Sardet et al. (1989) to span the membrane, and the equivalent regions in the *C. elegans* protein. The *C. elegans* sequence shows a high degree of similarity to the human sequence throughout the length of the *C. elegans* sequence

is hydrophilic, and is predicted to form a large, positively charged, cytoplasmic domain containing several potential phosphorylation sites (Sardet et al. 1989). The carboxy-terminal portion of the *C. elegans* protein has a similar hydrophilic domain with one consensus phosphorylation site (Thr/Ser-Pro-X-basic; described by Shenoy et al. 1989; see Fig. 4). A comparison of the hydrophobicity profiles of the two proteins reveals striking similarity of secondary structure (Fig. 5).

The dopa decarboxylase-like gene and the Na^+/H^+ antiporter-like gene lie between let-653 *and* let-56

In order to position the genes represented by the dopa decarboxylase-like cDNA and the Na^+/H^+ antiporter-like cDNA on the genetic map, we investigated whether the deficiencies that break between *let-653* and *unc-22* delete the DNA represented by the antiporter-like and decarboxylase-like cDNAs. Prasad and Baillie (1989) had already shown that the cDNA C2, which we now know to encode a Na^+/H^+ antiporter-like gene, mapped to sequences deleted by the deficiency *sDf9*. Since C2 was identified using fragment Q as a probe, and since Q lies to the left of R, *sDf9* must also delete the dopa decarboxylase-like gene. We now wished to determine whether the deficiencies *sDf19* and *sDf65* deleted the DNA containing the antiporter-like and decarboxylase-like genes. Pairs of PCR primers were designed from the 5′ regions of both the Na^+/H^+ antiporter-like cDNA and the dopa decarboxylase-like cDNA, and were applied to arrested deficiency homozygotes. Control primers designed from sequences on LGV were incorporated into the same reaction. PCR products resulting from the amplification of sequences between both the decarboxylase- and Na^+/H^+ antiporter-specific primers were not seen in *sDf19* homozygotes. Figure 6 shows the result obtained with the primers MAM2 and DB19, which were predicted from the Na^+/H^+ antiporter-like cDNA. From these data, we conclude that *sDf19* deletes at least the DB19 primer annealing site. The decarboxylase-like gene must also be deleted by *sDf19* because it is to the right of the antiporter-like gene; this has been confirmed (data not shown). Similar experiments conducted using the deficiency *sDf65* indicate that *sDf65* does not delete the primer annealing sites in the decarboxylase-like gene, and therefore cannot delete the antiporter-like gene. Thus, we have positioned the genes represented by these cDNAs between the left breakpoint of *sDf19* and the left breakpoint of *sDf65*, between *let-56* and *let-653*.

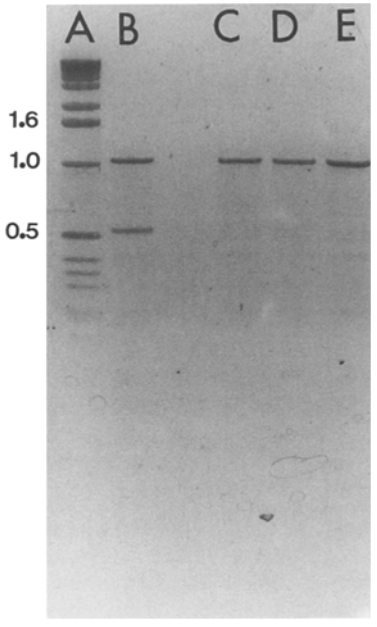

Fig. 6. Photograph of an ethidium bromide-stained agarose gel showing the results of polymerase chain reaction (PCR) on *sDf19* homozygotes. Lane A contains 1 kb DNA ladder (BRL); approximate sizes of marker DNA are indicated in kilobases. Lanes B–E show products of PCR reactions performed with DNA from various sources, using the primers MAM2 and DB19 and control primers specific for a region on linkage group V. Lane B is the wild-type (N2) control. Lanes C, D and E show products resulting from PCR on *sDf19* homozygotes. These lanes do not contain the 600 bp band indicative of *C. elegans* Na^+/H^+ antiporter amplified product (lane B). All lanes contain the upper 1.1 kb control band specific to sequences on linkage group V

Discussion

As part of our investigation into the organization of genes in the 2 map unit *unc-22(IV)* region defined by the deficiency *sDf2*, we have focused here on the 0.1 map unit interval between *let-653* and *unc-22*. We wished to identify the genes residing in this region, and their positions with respect to the deficiencies that subdivide the region, *sDf19* and *sDf65*. For the purpose of characterizing coding elements that map to this interval, cDNAs were obtained for two genes that reside on the cosmid C11F2. This cosmid has previously been shown to rescue the only mutagenically identified gene in the region, the essential gene *let-56* (Clark and Baillie 1992). Therefore, at least part of C11F2 had to be between *let-653* and *unc-22*. We have sequenced these cDNAs, and have assigned tentative identities to the coding elements based on their similarities to sequences in the EMBL nucleotide and SWISS amino acid databanks.

The cDNA isolated using fragment R as a probe was sequenced in its entirety, and the inferred amino acid sequence was used to search the SWISS and translated EMBL databanks. Amino acid identities of greater than 37% were detected with a variety of decarboxylases, including dopa decarboxylase from human, rat and *D. melanogaster*, and histidine decarboxylase from human and rat (Everleth et al. 1986; Ichinose et al. 1989; Tanaka et al. 1989; Joseph et al. 1990; Yamauchi et al. 1990). The best match to our sequence was obtained with L-aromatic amino acid (dopa) decarboxylase from human. This match was in excess of 39% amino acid identity. When conservative amino acid replacements are considered, the two proteins are 80% similar. These findings support the notion that our sequence is a member of the decar-

boxylase gene family and a possible candidate for a *C. elegans* L-aromatic amino acid decarboxylase.

The neurotransmitter dopamine is known to exist in *C. elegans,* and aromatic amino acid decarboxylase activity has been detected (Sulston et al. 1975). The biogenic amine serotonin is also present in *C. elegans* (Horvitz et al. 1982). In *Drosophila,* L-aromatic amino acid decarboxylase activity is required for the synthesis of both dopamine (Lunan and Mitchell 1969) and serotonin (Livingstone 1981). A number of mutations have been isolated in *Drosophila* that disrupt the dopa decarboxylase gene; the most common phenotype resulting from such lesions is lethality, probably due to defects in cuticle synthesis (Wright et al. 1982). In *C. elegans* a number of mutations have been isolated that result in the reduction or loss of serotonin and dopamine, including mutations in the genes *cat-1* and *cat-4* (Desai et al. 1988). Mutations in these genes result in animals that have a number of defects, including a reduced efficiency of male mating (Hodgkin 1983), and reduced levels of serotonin and dopamine in all serotonergic and dopaminergic cells (Sulston et al. 1975; Desai et al. 1988). However, no such mutation is known to map to the region in which the dopa decarboxylase-like gene that we have identified is located.

It has been noted that in the 30 kb of genomic DNA separating *let-56* from *unc-22* there are at least four extensive open reading frames, which encode integral membrane proteins having multiple membrane spanning domains. This information is based upon hydropathy profiles of peptides predicted from the genomic sequence generated by Benian et al. (1989). These predicted proteins have been used to search a variety of sequence databanks. Such searches have resulted in the identification of an open reading frame encoding a glucose transporter-like molecule (C. Fields, C. Soderlund, G. Benian, S.L'Hernault, K. Tobin and D.L. Baillie, personal communication), and another open reading frame encoding a protein similar to a rat gamma-aminobutyric acid (GABA) transporter (C. Fields, personal communication). The remaining two predicted proteins bear no significant similarity to any sequences currently in either the EMBL or SWISS databanks. The identification of the Na^+/H^+ antiporter gene from *C. elegans* raises to five the number of genes in this region predicted to encode integral membrane proteins, which possess multiple membrane spanning domains. These five coding elements are related in terms of the location of their predicted protein products, i.e. a cellular membrane. The Northern analysis conducted by Prasad and Baillie (1989) demonstrates that the coding elements located on fragments Q and R produce mRNA most abundantly in the L2 stage. Therefore, the genes on these *Pst*I restriction fragments may also be related in terms of temporal regulation of expression.

The human Na^+/H^+ antiporter has been cloned and sequenced, and mutants deficient in Na^+/H^+ antiporter activity have been isolated in hamster fibroblasts (Sardet et al. 1989, 1990; Pouyssegur et al. 1984). In vertebrates, the Na^+/H^+ antiporter regulates intracellular pH in all cells in which it is found, and plays an important role in signal transduction (Aronson and Boron 1986; Sardet et al. 1990). Analysis of mutant hamster fibroblast cells lacking Na^+/H^+ antiporter activity reveals that, in HCO_3^- free medium, growth of arrested mutant cells is retarded, presumably due to the inability of mutant cells to reach an internal pH permissive for DNA synthesis and release from G1/G0 arrest. In CO_2/HCO_3^- buffered media, this retardation of growth is abolished, presumably due to the presence of a Na^+ dependent Cl^-/HCO_3^- exchanger that is capable of regulating internal cellular pH. In all invertebrate animal cells examined, the primary regulator of internal pH is not the Na^+/H^+ antiporter but a different transporter, which mediates the uptake of Na^+ and HCO_3^- and the extrusion of Cl^- and H^+ (Boron 1983). Therefore, it seems likely that the Na^+/H^+ antiporter-like gene that we have identified in *C. elegans* serves some physiological function other than the regulation of cellular pH. The high degree of similarity (96%) between the predicted amino acid sequences of the human and *C. elegans* gene products in that region of the human protein predicted to extend extracellularly (Sardet et al. 1989) may yield some clue concerning what that function might be. The function of this extracellular domain, however, has yet to be elucidated.

The sequences reported in this paper have been positioned on the *C. elegans* genetic map. Using the PCR technique of Barstead and Waterston (1991), we have shown that the genes represented by both the antiporter-like and decarboxylase-like cDNAs fall between the left breakpoint of the deficiency *sDf19* and the left breakpoint of the deficiency *sDf65*. As these deficiencies fail to complement *let-56* yet complement *let-653,* the genes represented by both cDNAs lie between *let-653* and *let-56* (Fig. 1). This analysis has also served to localize *let-56* more precisely on the cosmid C11F2. The left-hand breakpoint of *sDf65* must fall within C11F2 because *sDf65* fails to complement *let-56,* and C11F2 rescues *let-56* (Clark and Baillie 1992). We have shown that *sDf65* does not delete the R1 priming site. Therefore, *let-56* must be between the R1 priming site and the end of C11F2, a distance of approximately 10 kb.

In summary, we have identified two genes from the cosmid C11F2, which has previously been shown to rescue *let-56*. One of these genes closely resembles the human growth factor activatable Na^+/H^+ antiporter. The identification of this gene raises to five the number of genes encoding integral membrane proteins having multiple transmembrane domains that map to the interval between *let-56* and *unc-22*. The other gene that we have identified most closely resembles L-aromatic amino acid (dopa) decarboxylases from human, rat and *Drosophila*. These two genes have been positioned on the genetic map between the left-hand breakpoint of the deficiency *sDf19,* and the left-hand breakpoint of the deficiency *sDf65,* between the genes *let-653* and *let-56*. This analysis has also localized *let-56* to the 10 kb region immediately to the right of the decarboxylase-like sequence on the cosmid C11F2.

Acknowledgements. We thank R.E. Rosenbluth, S.M. Gorski, B.F.F. Ouellette, A.M. Howell, A. Rose, K.S. McKim and C. Loer

for critical appraisals of the manuscript. We also wish to thank B.F.F. Ouellette for assistance with comparative DNA sequence analysis, and D. Suleman for technical assistance. This work was supported by an NSERC (Canada) predoctoral fellowship to M.A.M., an MDA of Canada predoctoral fellowship to S.S.P. and an NSERC operating grant to D.L.B.

References

Aronson PS, Boron WF (1986) Na^+/H^+ exchange, intracellular pH and cell function. Current topics in membranes and transport, volume 26. Academic Press, New York

Barstead RJ, Waterston RH (1989) The basal component of the nematode dense-body is vinculin. J Biol Chem 264:10177–10185

Barstead RJ, Waterston RH (1991) Vinculin is essential for muscle function in the nematode. J Cell Biol 114:715–724

Benian GM, Kiff JE, Neckelmann N, Moerman DG, Waterston RH (1989) Sequence of an unusually large protein implicated in regulation of myosin activity in *C. elegans*. Nature 342:45–50

Benton WD, Davis RW (1977) Screening gt10 recombinant clones by hybridization to single plaques *in-situ*. Science 196:180–182

Boron WF (1983) Transport of H^+ and of ionic weak acids and bases. J Membr Biol 72:1–16

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

Cabot EL, Beckenbach AT (1989) Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comp Appl Biol Sci 5:233–234

Clark DV (1990) The unc-22 IV region of *Caenorhabditis elegans:* genetic analysis and molecular mapping. PhD Thesis, Simon Fraser University, Burnaby, B.C., Canada

Clark DV, Baillie DL (1992) Genetic analysis and complementation by germ-line transformation of lethal mutations in the *unc-22 IV* region of *Caenorhabditis elegans*. Mol Gen Genet 232:97–105

Clark DV, Rogalski TM, Donati LM, Baillie DL (1988) The *unc-22(IV)* region of *Caenorhabditis elegans:* Genetic analysis of lethal mutations. Genetics 119:345–353

Coulson A, Sulston J, Brenner S, Karn J (1986) Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc Natl Acad Sci USA 83:7821–7825

Coulson A, Waterston R, Kiff J, Sulston J, Kohara Y (1988) Genome linking with yeast artificial chromosomes. Nature 335:184–186

Desai C, Garriga G, McIntire SL, Horvitz HR (1988) A genetic pathway for the development of the *C. elegans* HSN motor neurons. Nature 336:638–646

Everleth DD, Geitz RD, Spencer CA, Nargang FE, Hodgetts RB, Marsh JL (1986) Sequence and structure of the dopa decarboxylase gene of *Drosophila:* Evidence for novel RNA splicing variants. EMBO J 5:2663–2672

Feinberg A, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific acitivity. Anal Biochem 132:6–13

Hattori M, Sakaki Y (1986) Dideoxy sequencing method using denatured plasmid templates. Anal Biochem 152:232–238

Henikoff S (1987) Unidirectional digestion with Exonuclease III in DNA sequence analysis. Methods Enzymol 155:156–165

Hodgkin JA (1983) Male phenotypes and mating efficiency in *C. elegans*. Genetics 103:43–64

Horvitz HR, Brenner S, Hodgkin J, Herman RK (1979) A uniform genetic nomenclature for the nematode *Caenorhabditis elegans*. Mol Gen Genet 175:129–133

Horvitz HR, Chalfie M, Trent C, Sulston JE, Evans PD (1982) Serotonin and octopamine in the nematode. Science 216:1012–1014

Ichinose H, Kurosawa Y, Titanik K, Fujita K, Nagatsu T (1989) Isolation and characterization of a cDNA clone encoding human aromatic L-amino acid decarboxylase. Biochem Biophys Res Commun 164:1024–1030

Joseph DR, Sullivan P, Wang Y-M, Kozak C, Fenstermacher DA, Behrendsen ME, Zahnow CA (1990) Characterization and expression of the complementary DNA encoding rat histidine decarboxylase. Proc Natl Acad Sci USA 87:733–737

Livingstone MS (1981) Two mutations in *Drosophila* differentially affect the synthesis of octopamine, dopamine, and serotonin by altering the activities of two different amino acid decarboxylases. Soc Neurosci Abstr 7:117.1

Lunan KD, Mitchell HK (1969) The metabolism of tyrosine-*o*-phosphate in *Drosophila*. Arch Biochem Biophys 132:450–456

Moerman DG (1980) A genetic analysis of the *unc-22* region in *Caenorhabditis elegans*. PhD Thesis, Simon Fraser University, Burnaby, B.C., Canada

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence analysis. Proc Natl Acad Sci USA 85:2444–2448

Pouyssegur J, Sardet C, Franchi A, L'Allemain G, Paris S (1984) A specific mutation abolishing Na^+/H^+ antiport activity in hamster fibroblasts precludes growth at neutral and acidic pH. Proc Natl Acad Sci USA 81:4833–4837

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

Rogalski TM, Baillie DL (1985) Genetic organization of the *unc-22 IV* gene and the adjacent region in *Caenorhabditis elegans*. Mol Gen Genet 201:409–414

Rychlik W, Rhoads RE (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing, and in vitro amplification of DNA. Nucleic Acids Res 17:8543–8551

Sanger F, Coulson AR, Barrell BG, Smith AJH, Roe BA (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sardet C, Franchi A, Pouyssegur J (1989) Molecular cloning, primary structure, and expression of the human growth factor activatable Na^+/H^+ antiporter. Cell 56:271–280

Sardet C, Counillon L, Franchi A, Pouyssegur J (1990) Growth factors induce phosphorylation of the Na^+/H^+ antiporter, a glycoprotein of 110 Kd. Science 247:723–726

Shenoy S, Choi J, Bagrodia S, Copeland TD, Maller JL, Shalloway D (1989) Purified maturation promoting factor phosphorylates pp60c–src at the sites phosphorylated during fibroblast mitosis. Cell 57:763–774

Sulston J, Dew M, Brenner S (1975) Dopaminergic neurons in the nematode *Caenorhabditis elegans*. J Comp Neurol 163:215–226

Tanaka T, Horio Y, Teketoshi M, Imamura I, Ando-Yamamoto M, Kangawa K, Matsuo H, Kurodo M, Wada H (1989) Molecular cloning and sequencing of a cDNA of rat dopa decarboxylase: Partial amino acid homologies with other enzymes synthesizing catecholamines. Proc Natl Acad Sci USA 86:8142–8146

Wright TRF, Black BC, Bishop CP, Marsh JL, Pentz ES, Steward R, Wright EY (1982) The genetics of dopa decarboxylase in *Drosophila melanogaster* V. *Ddc* and *1(2)amd* alleles: Isolation, characterization and intragenic complementation. Mol Gen Genet 188:18–26

Yamauchi K, Ruriko S, Ohkawara Y, Tanno Y, Maeyama K, Watanabe T, Satoh K, Yoshizawa M, Shibuhara S, Takishima T (1990) Nucleotide sequence of the cDNA encoding L-histidine decarboxylase derived from human basophilic leukemia cell line, KU-812–F. Nucleic Acids Res 18:5891

Communicated by E. Bautz

Note added in proof. The *C. elegans* dopa decarboxylase sequence reported in this paper has been submitted to Genbank, and is available under Genbank accession number Z11576. The *C. elegans* Na^+/H^+ antiporter sequence reported in this paper has likewise been submitted to Genbank, and is available under Genbank accession number M23064.