The Journal of Neuroscience, April 1994, 14(4): 2290–2300

Cloning and Characterization of the Choline Acetyltransferase Structural Gene (*cha-1*) from *C. elegans*

Aixa Alfonso,[a] Kiely Grundahl, John R. McManus, and James B. Rand

Program in Molecular and Cell Biology, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104

We have cloned the *cha-1* gene from *Caenorhabditis elegans* using the method of transposon tagging. *cha-1* is the structural gene for ChAT, the enzyme that synthesizes ACh. Sequence analysis of cDNAs predicts a protein of 71.5 kDa; comparison of the deduced amino acid sequence with ChAT sequences from other species confirms that *cha-1* encodes ChAT. Comparison of cDNA and genomic sequences reveals that transcription is from right to left on the genetic map, and that some of the transcripts may result from *trans*-splicing of the 22-base spliced leader SL1. The *cha-1* gene is organized into 11 exons. The first exon contains only untranslated sequences, and is followed by an extremely long intron. The coding sequence of the *cha-1* transcript is disrupted by mutations in the *cha-1* gene. We have determined the sites of four transposon insertions and the end-points of two deletions that lead to the *cha-1* mutant phenotype; one of the deletions appears to eliminate gene function completely. Comparison of the *Drosophila*, rat, and *C. elegans* genes reveals conserved motifs and conserved intron sites.

[Key words: ChAT,* cha-1, *Caenorhabditis elegans, sequence analysis,* trans*-splicing, mutants]

ChAT (EC 2.3.1.6) is the enzyme that synthesizes ACh. ChAT has been purified and characterized from many vertebrate and invertebrate sources (Emson et al., 1974; Roskoski et al., 1975; Rossier, 1976; Eckenstein et al., 1981; Slemmon et al., 1982). Since ChAT is considered a differentiation-specific marker for cholinergic neurons, the localization and expression of the enzyme have been studied in a number of organisms (Eckenstein and Sofroniew, 1983; Levey et al., 1983; Gorczyca and Hall, 1987). DNA sequences have now been published for ChAT-encoding cDNAs from three mammals (pig, rat, and mouse) and from *Drosophila* (Itoh et al., 1986; Berrard et al., 1987; Brice et al., 1989; Ishii et al., 1990; Sugihara et al., 1990). In addition, genomic (exon–intron) structures have been published for the *Drosophila* and rat ChAT genes (Sugihara et al., 1991; Hahn et al., 1992). The sequence of a partial genomic clone of human ChAT has also been reported (Strauss et al., 1991).

In the nematode *Caenorhabditis elegans*, we can combine the tools of classical genetics and molecular genetics to study the function and regulation of genes and their products. We have been pursuing biochemical, genetic, and molecular studies of ChAT in *C. elegans*, in order to investigate cholinergic function and its control during neural development. Such studies are aided by the simplicity of the *C. elegans* nervous system: there are 302 total neurons whose position, morphology, lineage, and connectivity are essentially invariant (Sulston and Horvitz, 1977; Sulston et el., 1983; White et al., 1986). ACh is accepted to be the excitatory neurotransmitter at nematode neuromuscular junctions, and there is evidence for central cholinergic function as well (see Chalfie and White, 1988, for references).

Previous studies on *C. elegans* ChAT include the purification and biochemical characterization of the enzyme (Rand and Russell, 1985b), characterization of mutants deficient in ChAT (Rand and Russell, 1984; Hosono et al., 1985), and the genetic analysis of *cha-1*, the ChAT structural gene (Rand, 1989). Viable *cha-1* mutants have greatly reduced (<5% of the wild-type) ChAT activity; they are also uncoordinated, slow growing, and resistant to inhibitors of AChE (Rand and Russell, 1984), and they have significantly reduced ACh levels (Hosono et al., 1987; M. Nguyen and J. B. Rand, unpublished observations). Extremely severe mutations lead to complete developmental arrest (Rand, 1989).

We now report the cloning and sequence of *cha-1* from *C. elegans*. This includes complete genomic and cDNA sequences, and the intron–exon structure of the gene. We also present the structures of six *cha-1* ChAT-deficient mutations as well as a comparison of the exon-intron structures of the *Drosophila*, rat, and *C. elegans* genes.

Preliminary accounts of some of these results have appeared in abstract form (Alfonso-Pizarro et al., 1989, 1990).

Materials and Methods

Nematode culture conditions and strains. Nematodes were grown and maintained at 20°C on nematode growth medium (NGM) agar plates seeded with a lawn of *Escherichia coli* strain OP50-1 (Brenner, 1974; Johnson et al., 1988). Nematode strains used in this work include the wild-type Bristol N2 stock (Brenner, 1974) and PR1152 *cha-1(p1152)*, PR1166 *cha-1(p1156)*, PR1419 *cha-1(p1152) dpy-13(e184)*, RM49 *cha-*

Received Apr. 20, 1993; revised Sept. 30, 1993; accepted Oct. 7, 1993.

These studies were begun while A.A., K.G., and J.B.R. were affiliated with the Department of Zoology, University of Wisconsin–Madison. We are grateful to Karen Zaruba, Kate Guenther, Dan Drake, and Debbie Keith for technical assistance. We thank Stuart Kim, Bob Barstead, and Claudia Cummins for generously providing libraries, John Collins for transposon-specific probes, and John Sulston and Alan Coulson for assigning our clones to the *C. elegans* cosmid map and providing appropriate cosmids. We also thank Carl Johnson, Phil Anderson, Kathy Rand, and Lou Hersh for useful discussions and helpful suggestions, and Dick Russell for his help in the initial phases of this project. Paul Salvaterra, Lou Hersh, and Guy Benian kindly communicated results in advance of publication. Janet Duerr and Carl Johnson made helpful comments on the manuscript. This research was supported by Grants GM 38679 from the National Institutes of Health and BNS-8919780 from the National Science Foundation. A.A. was supported by postdoctoral fellowships from the Ford Foundation and NIH. Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR).

Correspondence should be addressed to James B. Rand, Program in Molecular and Cell Biology, Oklahoma Medical Research Foundation, 825 N.E. 13 Street, Oklahoma City, OK 73104.

[a]Present address: Department of Biological Sciences, University of Iowa, Iowa City, IA 52242.

1(p1154) dpy-13(e184), RM52 *unc-33(e204) dpy-13(e184)*, and PR1404 *unc-33(e204) vab-2(e96)*. Genetic nomenclature is according to Horvitz et al. (1979). We used TR638 as the transposition active strain; it contains a mutator allele (*mut-3*) that greatly increases the levels of Tc1 transposition in the germ line (Collins et al., 1987). To isolate nucleic acids, mixed populations of nematodes were grown and purified from liquid cultures containing *E. coli* strain RR1 as described by Sulston and Brenner (1974).

Transposon mutagenesis. For each round of selection, and to provide independent lines, individual wild-type hermaphrodites from TR638 were placed on 100 × 15 mm NGM plates heavily seeded with OP50-1. We set up 50–80 independent lines simultaneously, let them grow for approximately three generations, and the progeny (20,000–50,000 per line) were transferred to heavily seeded plates containing aldicarb (0.5 mM or 1.0 mM; obtained from Chem Service, West Chester, PA). Wild-type hermaphrodites are sensitive to these concentrations of aldicarb. They become paralyzed and hypercontracted, and they eventually die. After 7–21 d the aldicarb-containing plates were examined and aldicarb-resistant animals (i.e., those able to move and reproduce) were picked and transferred to plates without aldicarb.

Isolation of revertants. On normal growth medium (i.e., in the absence of aldicarb), *cha-1* animals are small, slow growing, and uncoordinated (Rand and Russell, 1984). We therefore looked for phenotypic revertants by screening between 3×10^5 and 1×10^6 mutant animals from each new mutant strain for faster-growing, better-moving, and/or larger-than-average animals. One such apparent revertant animal was found derived from strain RM1067, a strain containing the *cha-1(md1067)* mutation in the original, high-transposition background.

Genetic analysis. The methods and strategies for genetic analysis of *C. elegans* are those of Brenner (1974) and Herman and Horvitz (1980). Strategies used in the genetic analysis of *cha-1* and *unc-17* have been previously published (Rand and Russell, 1984; Rand, 1989). Three-factor genetic crosses were performed between the new *cha-1* mutations and linked markers on either side (*unc-33*, 0.4 map units to the left, or *vab-2*, 0.9 map units to the right). For each of these mutations, we isolated 30–35 independent Vab non-Unc-33 recombinants from + *cha-1* + / *unc-33* + *vab-2* heterozygotes. Two classes of recombinant chromosomes were expected: + *cha-1 vab-2* and + + *vab-2*. Approximately two-thirds of the Vab non-Unc-33 chromosomes contained the *cha-1* mutation, indicating that a recombination event had occurred between *cha-1* and *vab-2*. The other third contained the wild-type *cha-1* allele, indicating they had been derived from a recombination between *cha-1* and *unc-33*.

DNA analysis. Nematodes were disrupted by placing them in proteinase K [100 mM Tris Cl pH 8.5, 0.2 M NaCl, 50 mM EDTA, 0.5% SDS, 200 μg/ml proteinase K (BDH Biochemicals)] for 30 min at 65°C. Genomic DNA was extracted with phenol/chloroform/isoamyl alcohol (Ross, 1976), treated with ribonuclease (10 mM Tris Cl pH 8.0, 1 mM EDTA, 100 μg/ml RNase A), and isolated by ethanol precipitation (Maniatis et al., 1982). DNA fixation, Southern hybridization, and washes were performed according to standard methods (Maniatis et al., 1982). pGEM3Zf(+) (Promega) was used as the vector for all subcloning. Individual subclones are described in Figure 2. Plasmid pTc163 (obtained from John Collins and Phil Anderson, Univ. of Wisconsin) contains a Tc1 element cloned into the EcoRV site of vector pIBI76 and was used to identify Tc1-containing DNA fragments in Southern blots.

Screening of libraries, identification of cosmids, and isolation of genomic subclones. We screened 10^5 genomic phages from an N2 genomic library (prepared by Claudia Cummins, Univ. of Wisconsin) containing 15–20 kilobase (kb) MboI fragments in EMBL3, using the 0.85 kb SacI fragment as a probe. Cosmid DNA was purified based on the method of Birnboim and Doly (1979). We also screened 10^5 phages from a λgt10 cDNA library (Kim and Horvitz, 1990) and a λZap cDNA library (Barstead and Waterston, 1989). Genomic subclones (RM#1P, RM#15P, RM#16P, and RM#34P) were obtained using size-fractionated and gel-purified DNA fragments. Both phages and plasmids were purified to homogeneity using standard procedures (Maniatis et al., 1982).

Sequencing strategy. Sequencing of double- and single-stranded DNA was performed by the chain-termination method (Sanger et al., 1977) using the Sequenase version 2.0 kit (U.S. Biochemicals). Wild-type sequences were determined by "primer walking," using oligonucleotide primers synthesized at the Molecular Biology Resource Facility of the University of Oklahoma Health Science Center. The end-points of the deletions and sites of insertions were determined by sequencing asymmetric polymerase chain reaction (PCR) products (Gyllensten and Erlich, 1988) prepared from double-stranded templates. For the analysis of the Tc1 insertions we designed two transposon-specific primers that allowed us to determine the sequence at the insertion site, as well as the orientation of the transposon with respect to the flanking *cha-1* genomic sequence.

Sequence analysis. Sequence alignment and analysis were performed with the Genetics Computer Group software package (Devereux et al., 1984). Alignment of sequence pairs and calculation of percentage identity were done with the BESTFIT program, using system default values.

When the amino acid sequences for pig ChAT (Berrard et al., 1987), rat ChAT (Brice et al., 1989), and mouse ChAT (Ishii et al., 1990) are compared to the fly (Sugihara et al., 1990) and the nematode sequences, there are only five positions where the choice of mammal affects the presence or absence of a three-way identity between the mammal, the fly, and the nematode. Since there are 135 total three-way identities using the pig sequence, 134 using the rat sequence, and 133 using the mouse sequence, it is clear that the particular mammal chosen is not very important. We therefore used only one mammalian amino acid sequence for our comparisons (in Figs. 7, 8), and we chose the pig sequence because it was the first mammalian ChAT sequence to be published.

RNA analysis. Nematodes (2–3 gm) were ground to a fine powder in liquid nitrogen and resuspended in 3 vol of a solution containing 6 M guanidine thiocyanate, 10 mM KPO_4 pH 6.8, 1% *N*-lauroylsarcosine, and 5 mM dithiothreitol (DTT). The DNA was sheared with a Brinkman Polytron PT3000 and passages through a 22 gauge needle until the solution was no longer viscous. The cell debris was removed by centrifugation and the supernatant was layered over 5.7 M CsCl/10 mM EDTA and centrifuged overnight in an SW41 rotor. PolyA-selected RNA was isolated through binding to polyU-Sepharose 4B as described (Palatnik, et al., 1979). Total and polyA-selected RNAs were fractionated in a 1% agarose/formaldehyde gel and electroblotted to Hybond N+ membranes. Northern blots were hybridized and washed according to standard procedures (Maniatis et al., 1982).

For primer extension experiments, approximately 15 μg of total RNA from *C. elegans* was hybridized with 2 pmol of a 5′ end-labeled primer in 80% formamide, 0.4 M NaCl, 40 mM PIPES pH 6.4, 1 mM EDTA for 16 hr at 30°C. After ethanol precipitation, the hybrids were incubated with 50 U of AMV reverse transcriptase (Life Sciences) and 1 mM dNTPs in 50 mM Tris Cl pH 8.13, 40 mM KCl, 6 mM $MgCl_2$, 1 mM DTT, at 42°C for 2 hr. Samples were then treated with RNase, phenol extracted, and ethanol precipitated. Primer extension products were resuspended in 95% formamide, 20 mM EDTA (with 0.05% bromophenol blue/xylene cyanol) and resolved in a 6% acrylamide/7 M urea gel.

Accession numbers. The DNA sequences reported here have been submitted to GenBank, with accession numbers L08969 (*cha-1* composite cDNA) and L08970 (*cha-1* genomic sequence).

Results

Isolation of cha-1 *mutants in strains with active transposable elements*. We took advantage of the mobile element Tc1 to clone *cha-1* (Emmons et al., 1983; Rosenzweig et al., 1983; Collins et al., 1987; Herman and Shaw, 1987). Because mutations in *cha-1* confer resistance to AChE inhibitors such as aldicarb (Rand and Russell, 1984), we isolated aldicarb-resistant mutants from a transposition "active" strain, TR638 (see Materials and Methods). Since there are approximately 20 genes in addition to *cha-1* that can mutate to confer aldicarb resistance (Rand and Russell, 1985a; A. Alfonso, C. Johnson, M. Nguyen, and J. Rand, unpublished observations), it was necessary to screen a large number of resistant animals to obtain putative transposon-induced *cha-1* mutations. We screened approximately 7×10^5 haploid genomes, and we obtained 280 aldicarb-resistant mutants. Genetic complementation tests indicated that four of the 280 mutations were alleles of *cha-1*. All four mutations, when homozygous, led to the previously described set of *cha-1* phenotypes (Rand and Russell, 1984): very low ChAT activity, coiling uncoordinated behavior, slow growth, and small adult size, as well as aldicarb resistance.

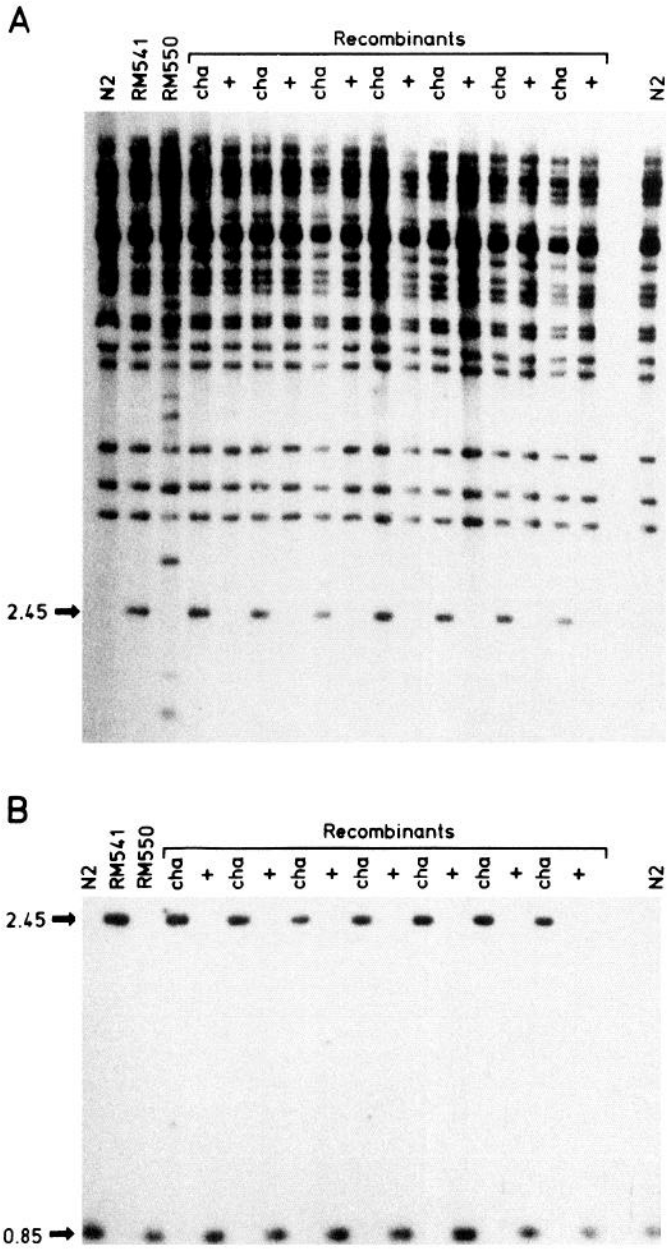

Figure 1. Southern blot analysis of *md1067*. Genomic DNA was isolated and digested with the restriction enzyme SacI. The fragments were resolved in a 1.0% agarose gel. *N2* is the wild-type reference strain; *RM541* is one of the outcrossed lines containing the *md1067* mutation; *RM550* is one of the outcrossed lines derived from the spontaneous phenotypic revertant obtained in a stock of *md1067*. The *lanes in the middle* represent independently isolated recombinant lines containing (*cha*), or lacking (+), the *cha-1(md1067)* mutation. Only 14 of 33 recombinants are shown. The same blot was hybridized sequentially to two different probes. *A* represents hybridization to a uniformly labeled Tc1 probe. The *arrow* points to the 2.45 kb Tc1-containing band present in all the "cha" strains and absent from all of the "+" strains. The additional Tc1 fragments detected in RM550 represent additional Tc1 elements still present in this strain as a result of differences in the outcrossing protocol. *B* represents hybridization to a uniformly labeled *cha-1*–specific probe. The probe was derived from the flanking DNA present in the 2.45 kb SacI fragment, obtained from RM541, after removal of the Tc1 sequences. The *lower arrow* points to the 0.85 kb wild-type *cha-1* fragment. The *upper arrow* points to the 2.45 kb fragment that results from the insertion of a Tc1 element into the 0.85 kb fragment.

cha-1 *mutants are associated with additional copies of Tc1.* All four mutants were outcrossed 8–10 times. This process retained those Tc1 elements near *cha-1*, reduced the overall Tc1 copy number to a level near that of wild-type Bristol N2, and stabilized the mutations. Hybridization of a Tc1-specific probe to genomic Southern blots of the four outcrossed mutants and Bristol N2 indicated that each of the four *cha-1* mutants had two to five additional copies of Tc1 (data not shown). To distinguish Tc1 elements within the *cha-1* gene from those in linked flanking regions, we isolated genetic recombinants resulting from crossovers between the *cha-1* mutations and closely linked markers to the left and to the right (see Materials and Methods). Two of the mutations, *md1067* and *md1143*, as well as their recombinant strains were analyzed by Southern blot analysis. Mutant strain RM541 *cha-1(md1067)* contained a novel 2.45 kb SacI Tc1-hybridizing fragment that was always associated with the presence of the mutant phenotype (Fig. 1*A*). Figure 1*A* shows a Southern blot analysis of 14 of the *md1067* recombinants. All 19 of the *md1067*-containing recombinant strains (only seven shown, lanes labeled "cha") and none of the 13 *md1067*-lacking recombinant strains (only seven shown, lanes labeled "+") contained the 2.45 kb SacI Tc1-hybridizing fragment. Similarly, the *md1143* mutation was correlated with a 3.7 kb EcoRI Tc1-hybridizing fragment (data not shown). The perfect correlation between the presence of the mutant phenotype and a novel Tc1-containing fragment strongly suggested that these two mutations resulted from the insertion of Tc1 into the identified restriction fragments.

Tc1, once inserted into a given gene, can excise (Emmons et al., 1986). Depending on the location of the insertion, and the precision of the excision, such an event can result in full or partial reversion of the mutant phenotype (Eide and Anderson, 1985, 1988). We screened (see Materials and Methods) between 3×10^5 and 1×10^6 animals from each mutant strain, and we isolated one phenotypic revertant of *cha-1(md1067)*. The phenotypic revertant (apparently wild-type) was crossed six times into a Bristol N2 background. A genomic Southern blot of the revertant strain lacked the *md1067*-specific Tc1-containing SacI fragment (Fig. 1*A*, lane RM550). Spontaneous loss of the uncoordinated phenotype was thus associated with loss of the Tc1-containing fragment as well as full restoration of ChAT activity (data not shown), which provided additional evidence that the *cha-1(md1067)* mutation was caused by the insertion of Tc1.

Cloning cha-1. We cloned the Tc1-containing 2.45 kb SacI fragment present in RM541. We removed almost all of the 1.6 kb Tc1 element and subcloned the flanking *cha-1*–specific sequences. These sequences were radiolabeled and hybridized to the genomic Southern blot containing the DNA from the recombinant strains. As illustrated in Figure 1*B*, a single hybridizing fragment was detected in each lane. This fragment was 0.85 kb in Bristol N2, in the *md1067*-lacking recombinant strains (Fig. 1*B*, lanes labeled "+") and in the revertant strain (Fig. 1*B*, lane RM550). As expected, the fragment was 1.6 kb larger, 2.45 kb, in the *md1067*-containing recombinant strains (Fig. 1*B*, lanes labeled "cha").

The 0.85 kb SacI fragment detected in Bristol N2 DNA a single hybridizing BamHI fragment of approximately 4.5 kb in size (data not shown). The 0.85 kb SacI fragment is contained entirely within the 4.5 kb BamHI fragment (see Fig. 2). Southern blot analysis with the 4.5 kb BamHI fragment in turn revealed an overlapping 4.3 kb EcoRI fragment located to the left on the physical map. We have subcloned all of the EcoRI and BamHI

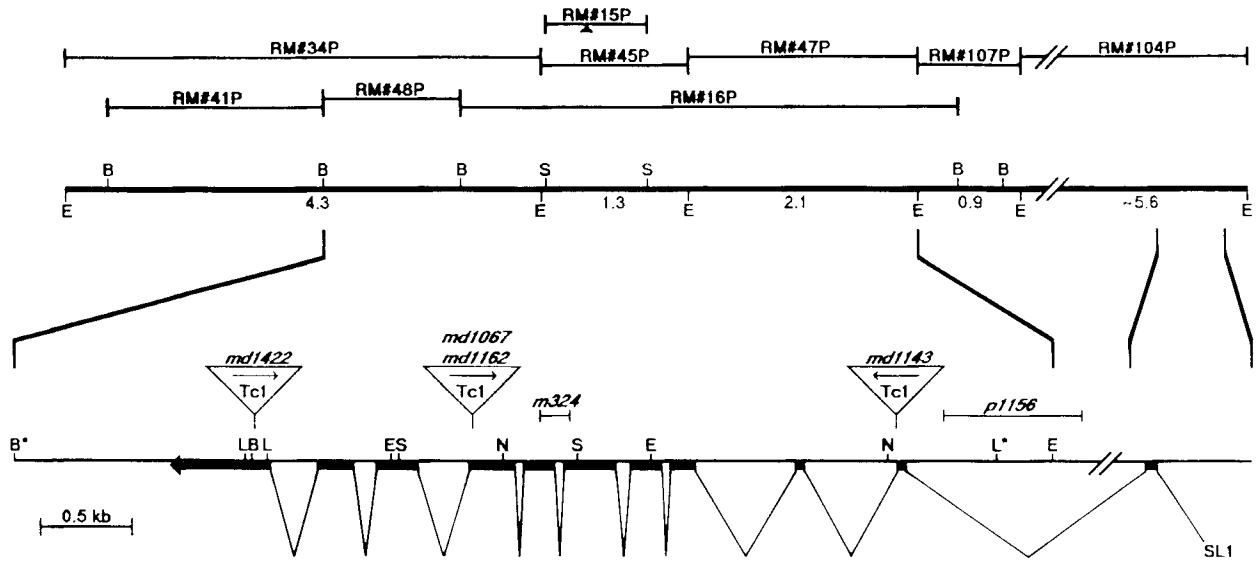

Figure 2. Physical map of the *cha-1* region of *C. elegans*. The *upper map* is a partial restriction map of the genomic region containing the *cha-1* locus. The orientation is the same as the genetic map (Rand, 1989). Sizes of the EcoRI fragments are given in kilobases. The extents of various subclones are indicated above the map; all were derived from wild-type DNA, except for RM#15P, which carries a small piece of the Tc1 element causing the *md1067* mutation (marked with an *arrowhead*). An enlargement of part of the map is shown at the *bottom*, indicating the structure of the *cha-1* cDNAs and the sites of six *cha-1* mutations. Transcription is *from right to left*. The *scale break* in both maps is to demonstrate the location of the first *cha-1* exon, near the right end of subclone RM#104P. The *trans*-splicing of the leader sequence SL1 to the first exon is indicated—this is the inferred 5′ structure of a known transcript class. Primer extension experiments indicate the existence of additional transcript classes with a different sequence/structure at the 5′ end. The *arrows* on the Tc1 elements indicate the orientation of the Tc1 open reading frame. Restriction sites: *B*, BamHI; *E*, EcoRI; *L*, SalI; *N*, NcoI; *S*, SacI. The *asterisks* on the SalI site on the right and the BamHI site on the left of the expanded map indicate the boundaries of the 5.2 kb region for which complete genomic sequence is available.

fragments present in this genomic region; they are diagrammed in Figure 2.

Screening of an N2 genomic library (C. Cummins and P. Anderson) identified seven positive phages. A. Coulson and J. Sulston (personal communication) kindly located two of these on a contig that had previously been shown to overlap YACs containing the gene *sup-29*. Since, on the *C. elegans* genetic map, *sup-29* lies approximately 0.5 map units to the right of *cha-1* (Edgley and Riddle, 1990), this provided additional evidence that our genomic clones were in the vicinity of the *cha-1* locus.

Southern blot analysis of mutants and alignment of the physical and genetic maps. The 0.85 kb SacI fragment and the 4.5 kb BamHI fragment were used to probe genomic Southern blots containing DNA from *cha-1* mutants. Figure 3 illustrates this analysis. We detected insertions (due to Tc1) associated with all four spontaneous mutations isolated from TR638 (Fig. 3, lanes 2–5) and deletions in two EMS-induced mutations, *m324* and *p1156* (Fig. 3, lanes 6, 7). Each of these mutations was positioned on the physical map (Fig. 2). The two deletion alleles, which had different positions on the physical map, had been previously studied by fine structure genetic analysis: *m324* is to the left of *p1156* (Rand, 1989). This information provided a left–right orientation for the physical map that was consistent with the "standard" genetic map.

To confirm this orientation, we also showed that *cha-1(md1067)* maps to the left of *p1156*. From a four-factor cross (see Materials and Methods), two non-Unc recombinants were obtained out of approximately 73,000 F_2 progeny scored. These data are consistent with expectations from previous experiments.

Isolation and sequence of cDNAs. The 4.5 kb BamHI fragment was used to probe a λgt10 cDNA library prepared from mixed stage *C. elegans* (Kim and Horvitz, 1990). Three phages were isolated and characterized. Two of these phages: RM#17L and RM#19L, appear to have identical 2.1 kb inserts. The insert in RM#17L hybridizes strongly to the three EcoRI fragments on the left side of the physical map and faintly to the 5.6 kb EcoRI fragment at the right end of the map in Figure 2. RM#18L is different: it contains a 0.67 kb insert whose hybridization is confined to the 0.9 kb EcoRI fragment on our map. We do not yet know what this molecule represents.

We determined the complete DNA sequence for the nearly full-length cDNA represented by RM#17L. This 2.1 kb cDNA contains an open reading frame, which encodes a deduced protein of 627 amino acids with a molecular weight of 71.5 kDa (Fig. 4). This value is in excellent agreement with the size of the ChAT protein (Rand and Russell, 1985b). Translation of the open reading frame starts with a methionine codon at position 76 on the cDNA (see Discussion). Three other cDNAs (RM#50P, RM#69P and RM#120P) that map to the *cha-1* region were isolated from a λZap library (Barstead and Waterston, 1989). Their inserts were characterized by restriction enzyme analysis using high-resolution gels and the fragment sizes appeared to be identical to those in the insert present in RM#17L, except for their termini. We sequenced the ends and the isolates differ slightly from RM#17L at their terminal 5′ and 3′ sequences (see below and Fig. 4). A comparison of the *C. elegans* ChAT sequence with other ChAT sequences is presented in the Discussion.

Structure of cha-1 *transcripts and detection of* cha-1 *mRNA.* We obtained complete sequence data for 5.2 kb of genomic DNA

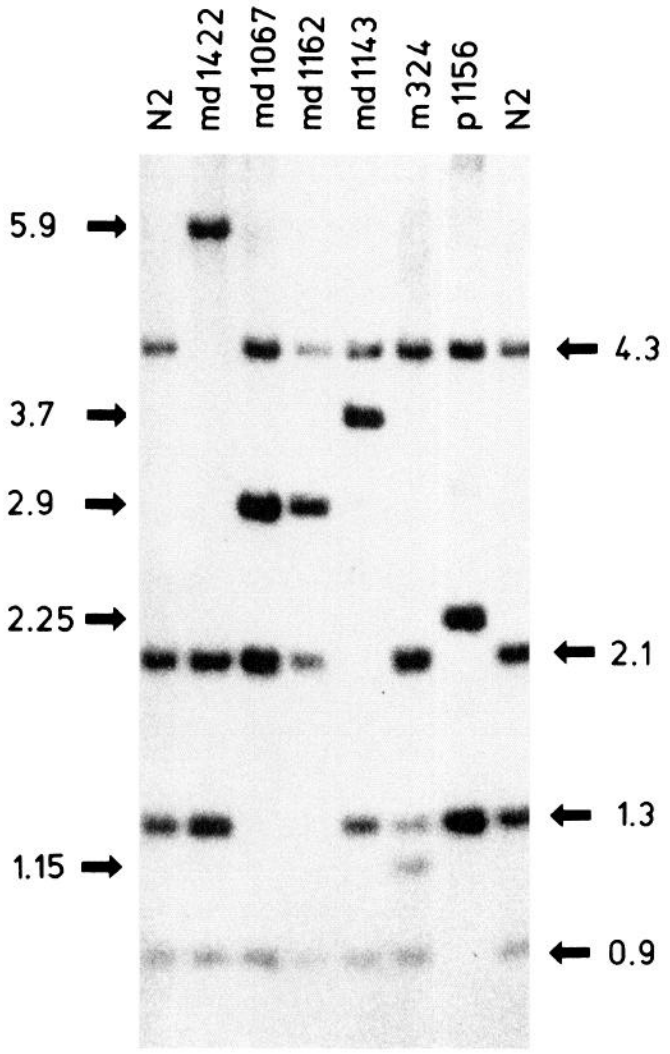

Figure 3. *cha-1* mutations associated with restriction fragment length polymorphisms. Genomic DNA was isolated from strains harboring *cha-1* mutations, and from the wild-type reference strain. Samples were digested with the restriction enzyme EcoRI, the fragments were resolved in a 1.0% gel, and the blot was probed with the genomic 4.5 kb BamHI fragment (RM#16P in Fig. 2). The *arrows to the right* identify the four hybridizing fragments in the wild-type N2: 4.3 kb, 2.1 kb, 1.3 kb, and 0.9 kb. The *arrows to the left* identify the altered fragment sizes. *md1422*, *md1067* (like *md1162*), and *md1143* are associated with a 1.6 kb insertion into the 4.3 kb, 1.3 kb, and 2.1 kb EcoRI fragments, respectively. *m324* showed two bands of comparable intensity, one the size of the wild-type fragment and the other 0.15 kb smaller. *m324* is an EMS-induced lethal allele that must be maintained as a heterozygote over the nT1 balancer chromosome. The smaller band represents the mutant sequences, carrying a 0.15 kb deletion. *p1156* is associated with a deletion of the 0.9 kb fragment and a shift in the 2.1 EcoRI fragment.

that includes the open reading frame present in the *cha-1* cDNAs (the region between the asterisks in Fig. 2). Comparison of genomic sequence to cDNA sequences allowed us to determine that the direction of transcription is right to left, and also provided us with the exon–intron boundaries (see Figs. 2, 4). In addition, we found that the first 75 nucleotides present in the cDNA represented by RM#50P were not contained in this genomic 5.2 kb region. Instead, most of this sequence was localized, by hybridization, to the right end of the 5.6 kb EcoRI fragment on the right end of our physical map (Fig. 2). Genomic sequence analysis confirmed the position of 66 of these nucleotides, and also indicated that they are preceded by a splice acceptor site. The nine nucleotides at the 5′ end of RM#50P do not precisely match any of the nearby genomic sequences; however, they are identical to the last nine bases of the *trans*-spliced leader sequence SL1. This 22-base sequence is known to be *trans*-spliced onto approximately 10% of the *C. elegans* mRNAs (Krause and Hirsh, 1987; Bektesh et al., 1988), and it therefore seems likely that at least some of the *cha-1* mRNAs are *trans*-spliced with SL1.

Following the putative SL1 sequence, the *cha-1* transcript consists of 11 exons that range in size from 50 to 477 nucleotides. The exons are interrupted by introns that range in size from 43 to 530 nucleotides, except for the first intron, which (for *C. elegans*) is unusually long, about 6.5 kb. The first exon is untranslated. The presumed protein initiation codon is at the very start of the second exon (see Figs. 2, 4).

There is evidence of heterogeneity in the 5′ ends of *cha-1* mRNAs. Primer extension experiments (Fig. 5) indicate three size classes of mRNAs. One size class extends 22 nucleotides past the *trans*-splice acceptor site, and thus appears to represent molecules with a full-length (22-base) SL1 *trans*-spliced as in RM#50P. The abundance of this class appears to be quite low, but the presence of a partial SL1 at the 5′ end of RM#50P confirms the existence of such a class of transcripts. The other two size classes appear to be more abundant, and correspond to molecules that, if unspliced, would extend 77 nucleotides and 91 nucleotides beyond the *trans*-splice acceptor site (Fig. 5). The structures of these longer 5′ ends have not yet been determined.

Northern blots containing total and polyA-selected RNAs from mixed stage nematodes were hybridized with labeled DNA probes specific for the *cha-1* transcript, and an RNA of about 2.3 kb was detected (Fig. 6). This size is consistent with the predicted size determined from the cDNA and primer extension results.

Sequence analysis of cha-1 *mutants.* The location of the insertion sites for the Tc1 elements and the boundaries of the deletions present in the RFLP-containing mutants (see Figs. 2, 4) were determined by sequencing PCR-amplified single-stranded DNA derived from each of the mutant strains (see Materials and Methods). *cha-1(md1067)* and *cha-1(md1162)* have identical Tc1 insertion sites in the eighth exon with the transposon inserted in the same orientation. *cha-1(md1422)* results from a Tc1 insertion in the eleventh exon. *cha-1(md1143)* is caused by insertion of a Tc1 element into an intron, 2 base pairs (bp) downstream from the end of the second exon. This insertion disrupts the sequence of the 5′ splice donor, which presumably interferes with the splicing of the intron at this site. We do not know yet how these insertions affect the structure of the transcripts, but we presume that there is still some residual ChAT activity in all of the Tc1 insertion mutants (see Discussion). The mutation *m324*, which leads to a larval lethal phenotype, results from a deletion of 157 bp: 112 bp of coding sequences from the sixth and seventh exons and the contiguous intron. This deletion results in a frameshift that generates a termination codon 10 amino acids downstream from the deletion site. The mutation *p1156* is associated with a 747 bp deletion, completely within the long first intron. The left end-point of this deletion is 194 bp upstream from 3′ acceptor site of the second *cha-1* exon. This deletion may remove some sequence important for the splicing of the long intron. A splice donor site is created by the sequences juxtaposed as a result of the deletion, but we do not know whether this site is used for splicing. Alternatively, it is

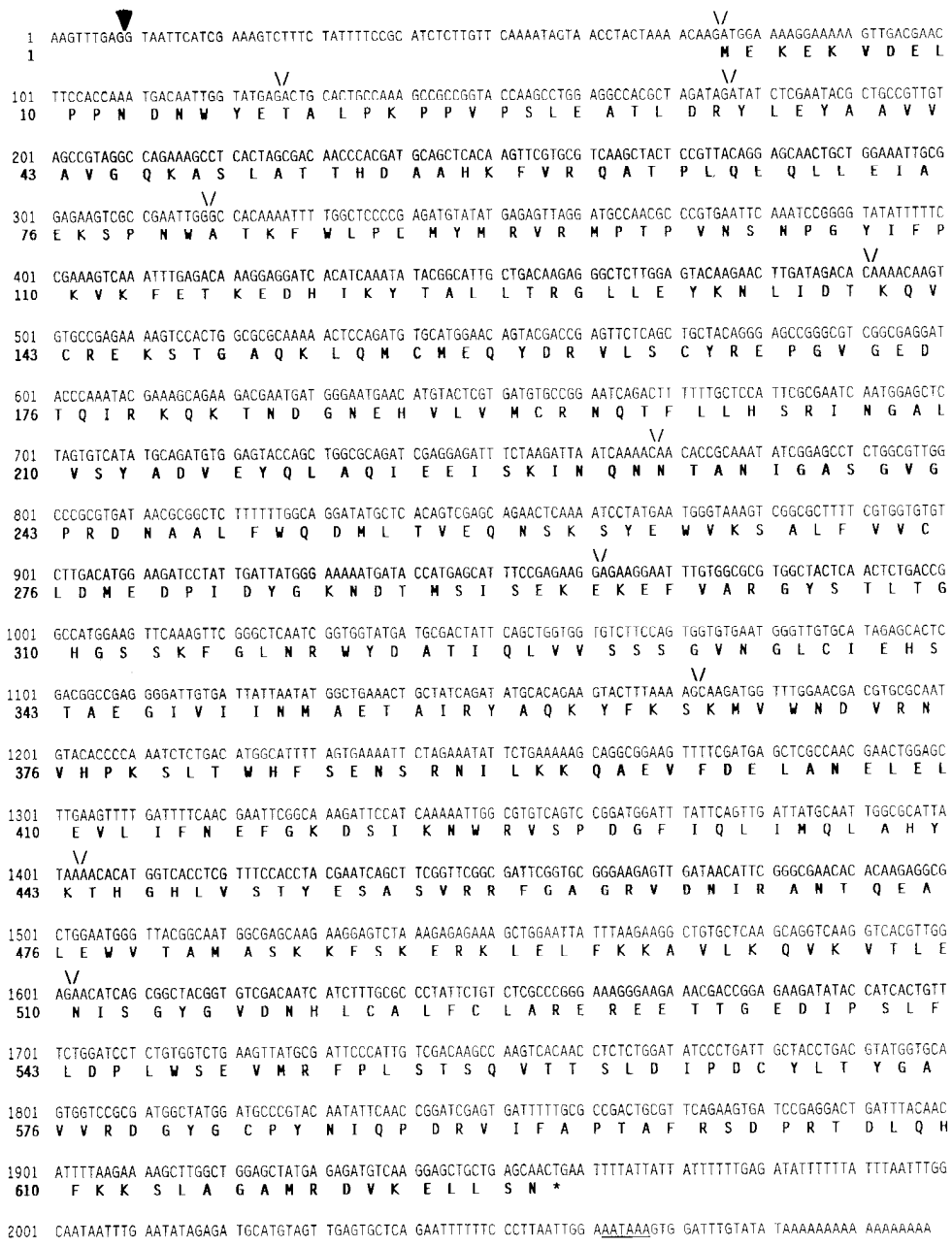

Figure 4. Composite nucleotide sequence of *cha-1* cDNAs, deduced ChAT amino acid sequence, and intron locations. Amino acids are identified by the standard one-letter code. The first nine nucleotides are apparently *trans*-spliced from the leader sequence SL1; the *solid triangle* between nucleotides 9 and 10 indicates the site at which (for at least some of the *cha-1* mRNAs) *trans*-splicing occurs. *Open arrowheads* indicate the location of introns. The putative polyadenylation signal (AATAAA) is *underlined*. RM#50P starts at nucleotide 1 and ends at 2078. RM#17L starts at nucleotide 17 and is polyadenylated following 2079. RM#69P starts at nucleotide 19 and is polyadenylated following 2081. Location of mutations (see Fig. 2): *md1067* and *md1162* are Tc1 insertions between nucleotides 1162 and 1163; *md1422* is a Tc1 insertion between nucleotides 1688 and 1689; *md1143* is a Tc1 insertion two nucleotides after the start of the intron following nucleotide 126; *m324* is a deletion removing 156 bp of genomic sequence corresponding to cDNA nucleotides 736–847.

possible that a point mutation elsewhere in the coding region is responsible for the observed phenotype.

Discussion

cha-1 *encodes ChAT.* The *cha-1* gene was originally defined genetically by a set of noncomplementing mutations that affected ChAT activity and led to resistance to AChE inhibitors (Rand and Russell, 1984). By correlating genetic with molecular data, we now provide three lines of evidence that the genomic region we have cloned is the *cha-1* locus. First, the position of these clones on the *C. elegans* physical map is consistent with the position of *cha-1* on the *C. elegans* genetic map. Second, the presence of a transposon insertion (*md1067*) into this region is perfectly correlated with the presence of the *cha-1* mutant phenotype even after genetic recombination very close to *cha-1*. Furthermore, spontaneous reversion of the *cha-1* mutant phenotype resulting from *md1067* was accompanied by excision of the transposon (Fig. 1*A*). Third, five other *cha-1* mutations are also associated with DNA alterations in this region (Fig. 3). The cDNAs described above appear to represent the *cha-1* transcript: they were the only cDNAs found that were derived from the region defined by *cha-1* mutations, and they contained most of the sites affected by these mutations.

We have previously provided genetic and biochemical evidence that *cha-1* is the structural gene for ChAT (Rand and Russell, 1984); the present study provides molecular evidence. There is striking similarity between the protein sequence deduced from our cDNAs and the sequences of mammalian and *Drosophila* ChATs (Fig. 7; see also below). Clearly, these genes are homologs.

Comparison of C. elegans *ChAT with other ChAT sequences.* Comparing the *C. elegans* ChAT sequence to the published ChAT sequences derived from mammals and *Drosophila* (Berrard et al., 1987; Brice et al., 1989; Ishii et al., 1990; Sugihara

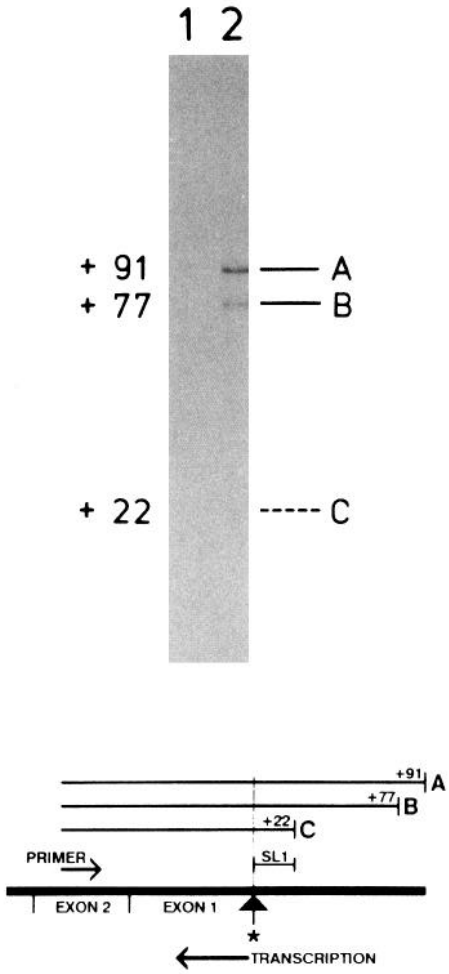

Figure 5. Primer extension of *cha-1*. *Top*, tRNA (15 μg; *lane 1*) or total RNA (*lane 2*) was used for primer extension (see Materials and Methods), and the resulting products were resolved on a polyacrylamide gel. In parallel lanes (not shown), a sequencing ladder was run, for precise identification of extension product sizes. *Numbers on the left* correspond to the distance from the *trans*-splice acceptor site; the *letters on the right* identify the three major extension products that were consistently observed. *Band C*, although quite faint, has been observed consistently in five experiments. *Bottom*, Diagram of the experiment and results. The left–right orientation is as in Figure 2, with the direction of transcription *from right to left*. The primer (*rightward arrow* above exon 2) was 81 nucleotides from the *trans*-splice acceptor site, indicated by the *asterisk* and *arrowhead*. The lengths of the three extension products (*A–C*) are indicated schematically, as is the 22-nucleotide *trans*-spliced leader sequence SL1.

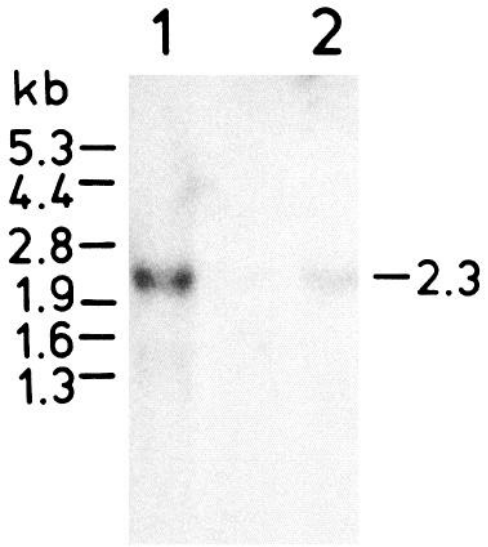

Figure 6. Northern blot of *cha-1*. *Lane 1*, 5 μg polyA-selected RNA; *lane 2*, 20 μg total RNA. See Materials and Methods for precise protocol. The positions of markers (in kilobases) are shown on the *left*.

et al., 1990) allows us to determine conserved structural elements that are presumably important for enzymatic activity or for other ChAT functions. When aligned as indicated in Figure 7, the *C. elegans* amino acid sequence is 34% and 36% identical to the *Drosophila* and porcine sequences, respectively, with the *Drosophila*–porcine identity at 40%. (This analysis, and what follows, is essentially unchanged if either the rat or mouse sequence is substituted for the pig sequence; see Materials and Methods). One significant difference between species is that the *Drosophila* protein is 60–70 residues longer at the N-terminal end (Fig. 7), although there is evidence that part of this sequence is removed by posttranslational cleavage (Selski et al., 1989).

An identity profile of the three sequences is given in Figure 8, and shows that there are 8–15 conserved domains (depending on how the precise boundaries are drawn) of varying size, and these domains are distributed along almost the entire length of the protein. Three of these domains are represented by high, sharp peaks, indicating that these are relatively small, isolated, highly conserved motifs. The first of these motifs is LPK–PVP–L (consensus residues 88–96, labeled "a" in Fig. 8), and it is the first part of the consensus to show any significant degree of identity. Interestingly, this motif does not occur until the third ChAT exon in *C. elegans, Drosophila* and rat (although in *C. elegans* and rat, this is the second coding exon). The other two small conserved motifs are HV–V–C–NQ (consensus residues 269–277, labeled "b" in Fig. 8), and F–LSTSQV (consensus residues 663–670, labeled "d" in Fig. 8). In addition to these relatively short conserved motifs, there are larger conserved domains indicated in Figure 8. The most pronounced of these is a 75 amino acid region (consensus 494–569, labeled "c" in Fig. 8). In the middle of this domain (515–556), each of the three pairwise species comparisons shows 30 of 42 consecutive residues identical, and 26 of the 42 residues are identical in all three of the sequences (Fig. 7). If conservative amino acid substitutions are allowed, then 32 of 42 positions are similar.

The sequence comparisons may be useful for determining essential functional sites in ChAT. Biochemical experiments have suggested that the active site of ChAT contains an arginine and a histidine (Currier and Mautner, 1974; Roskoski, 1974; Malthe-Sorenssen, 1976; Mautner et al., 1981). Comparison of the *Drosophila* and pig sequences revealed 14 conserved arginines. The three-way comparison now reduces the number of conserved arginine residues to seven (consensus 166, 325, 405, 545, 546, 551, and 556). It is noteworthy that the last four of these residues are tightly clustered within the large conserved domain discussed above. It should be possible to test the importance of these sites by *in vitro* mutagenesis.

Inspection of the three sequences in Figure 7 reveals the presence of only two conserved histidines (at consensus positions 269 and 427), each of which is in a domain of conserved sequence. In addition, there is a partially conserved histidine at positions 394–396; that is, the nematode histidine is two resi-

```
   fly   1  vasneastsa agsgpesaal fsklrsfsig sgpnspqrvv snlrgflthr lsnitpsdtg wkdsilsipk kwlstaesvd
   pig   1  .......... .......... .......... .......... .......... .......... .mpilektpp kmaakspsse
  worm   1  .......... .......... .......... .......... .......... .......... .......mek ekvdelppnd
CONSEN   1  ---------- ---------- ---------- ---------- ---------- ---------- --------pk k-----ps-d

   fly  81  efgfpdtLPK vPVPaLdeTm adYir..ale pittpaqler tkelirqFsa pqgigarLhq yLldkrEaed NWayyyWLne
   pig  20  e...epgLPK lPVPpLqqTl atYlr..cmq hlvpeeqfrr sqaivqqFga pgglgetLqq kLlerqEqta NWvseyWLnd
  worm  14  nw.yetaLPK pPVPsLeaTl drYleyaavv avgqkaslat thdaahkFvr q...atpLqe qLieiaEksp NWatkfWLpe
CONSEN  81  e---e--LPK -PVP-L--Tl a-Ylr--a-- -----aql-r t-----qF-a p-g-g--Lqq -Lle--E--- NWa--yWLne

   fly 159  MYmdiRiplP iNSnPgmvFp prrFktvhDv ahfaArLldG iLshremlDs gelpleraas rrrisrcawr sttaCwapvv
   pig  95  MYlnnRlalP vNSsPaviFa rqhFqdtnDq lrfaAnLisG vLsykallDs hsipidcakg qlsg...... .qplCmkqyy
  worm  90  MYmrvRmptP vNSnPgyiFp kvkFetkeDh ikytAlLtrG lLeyknliDt kqvcrekstg aqkl...... ..qmCmeqyd
CONSEN 161  MYm--R-plP vNSnPg-iFp ---F-t--D- --faA-L--G -Lsyk-llDs ---p-e-a-g ---------- ----Cm-qy-

   fly 239  glvSsrtrss crrgTmnd.. .....edrHV vViCrNQmyc vvlqasdrgk lSeseiasQi lyvlsdapcl pakpvpvGll
   pig 168  glfSsyrlpg htqdTlvaqk ssvmpepeHV iVaCcNQffv ldvvinfr.r lSegdlftQl rkivrmasne derlppiGll
  worm 162  rvlScyrepg vgedTqi.rk qktndgneHV lVmCrNQtfl lhsring.al vSyadveyQl aqieeiskin qnntaniGas
CONSEN 241  gl-Ssyr-pg ---dT----k -----e-eHV -V-CrNQ-f- l---in-r-- lSe-d---Ql --i---a--- -----piGll

   fly 312  taepRstwAr dremlqeder Nqrnleliet aqvvlCLDep lagnfnargf tgatptvhra gdrdetnmah emihGgGsey
   pig 247  tsdgRsewAe artvlvkdst Nrdsldmier ciclvCLDap ggmels.... .......... ....dtnral qllhGgGcsk
  worm 240  gvgpRdnaAl fwqdmltveq Nsksyewvks alfvvCLDme dpidygkndt msisek.... .ekefvargy stltGhGssk
CONSEN 321  t--pRs-wA- -r--l--de- N--sle-ie- a--vvCLD-p ---------- ---------- -----tnra- --lhGgGssk

   fly 392  nsgNRWfDkt mQliictdGt wGlcyEHScs eGIavvqllE kiyk.kieeh pdednglpqh hlppPerLeW hvgpqlqlrf
   pig 309  ngaNRWyDks lQfvvgrdGt cGvvcEHSpf dGIvlvqctE hllkhmvkss kkmvradsvs elpaPrrLrW kcspeiqgll
  worm 315  fglNRWyDat iQlvvsssGv nGlciEHSta eGIviinmaE tairyaqkyf kskmvwndvr nvh.PksLtW hfsensrnil
CONSEN 401  ng-NRWyDkt -Qlvv--dGt -Glc-EHS-- eGIv-vq--E ---k---k-- k-------v- -lp-P-rL-W h-sp--q--l

   fly 471  aqasksvdkc iddLdfyVyr yqsyGKtfIK scqvSsDvyI QlalQLAhYk lyGrLVaTYE SAStRRFlhG RVDcIRaast
   pig 389  assaeklqqi vknLdftVyk fddyGKtfIK qqkcSpDafI QvalQLAfYr lhGrLVpTYE SASiRRFheG RVDnIRsatp
  worm 394  kkqaevfdel aneLeleVli fnefGKdsIK nwrvSpDgfI QlimQLAhYk thGhLVsTYE SASvRRFgaG RVDnIRantq
CONSEN 481  a--ae--d-- ---Ldf-Vy- f--yGKtfIK ---vSpD-fI QlalQLAhYk lhGrLV-TYE SAS-RRF--G RVDnIRaat-

   fly 551  EALewakAmc qgeganvple sdredeeesr kvkfsiyskd hlreLfrcAv arQtevmvkn IlGngiDipL lgLreasiE.
   pig 469  EALhfvkAit dhasa..... .......... ......mpds eklLLlkdAi raQtqytvma ItGmaiDnhL lgLrelarE.
  worm 474  EALewvtAma skke...... .......... .......ske rkleLfkkAv lkQvkvtlen IsGygvDnhL caLfclarEr
CONSEN 561  EALewvkAm- ----a----- ---------- -------sk- -kleLfk-Av --Qt-vtv-n I-G-giDnhL lgLrelarE-

   fly 630  ...vtgemhe lFkDesynis qcFlLSTSQV acstds...f mgYGpVtprG YGcsYNphPe qivFcvsaFy scedTsasry
   pig 527  ...vckelpe mFtDetylms nrFvLSTSQV pttmem...f ccYGpVvpnG YGacYNpqPe silFcissFh gckeTsstkf
  worm 531  eettgedips lFlDplwsev mrFpLSTSQV ttsldipdcy ltYGaVvrdG YGcpYNiqPd rviFaptaFr sdprTdlqhf
CONSEN 641  ---v--e-pe lF-De-y--s -rF-LSTSQV -ts-d----f --YGpVvp-G YGc-YNpqPe -i-Fc-saF- sc--Ts---f

   fly 704  aKslqdsldi mrdLlqn*.. .......... .......... ..
   pig 601  aKaveesfie mkgLcslsqs gmgkplatke kvtrpsqvhq p*
  worm 611  kKslagamrd vkeLlsn*.. .......... .......... ..
CONSEN 721  aKsl--s--- mk-Llsn--- ---------- ---------- --
```

Figure 7. Comparison of *Drosophila*, pig, and *C. elegans* ChAT sequences. Numbering for each sequence (except the consensus) begins at the presumed translation start. Amino acids are identified by the standard one-letter code; those that are identical in all three sequences are shown in *uppercase*. Periods (.) indicate gaps inserted into the sequences to optimize alignment. The consensus sequence (*CONSEN*) has an *uppercase letter* for each three-way identity, a *lowercase letter* for each two-way identity, and a dash (–) for each position with no identity.

dues out of alignment with its mammalian and fly counterparts, but in a similar context. Carbini and Hersh (1993) have recently obtained evidence from *in vitro* mutagenesis experiments with *Drosophila* ChAT that the conserved histidine corresponding to 427 in the consensus sequence (*C. elegans*, 341) is the catalytically essential histidine.

Conserved domains and exon boundaries. The exon boundaries for *C. elegans* ChAT (10 introns), *Drosophila* ChAT (7 introns; Sugihara et al., 1991), and rat ChAT (introns 2–14; Hahn et al., 1992) are superimposed on the sequence identity profile in Figure 8. In general, no clear relationship appears to exist between the domains of conserved sequence and the exon structures of any of these ChAT genes, although all of the exons except the first two contain at least one conserved domain. However, there are striking similarities in the intron positions in the three ChAT genes. The *C. elegans* and rat genes have five introns in similar locations (four in identical positions, and one off by three nucleotides). The rat and *Drosophila* ChAT genes also have five introns in similar locations (three in identical positions, one off by one nucleotide, and one off by three nucleotides). The *C. elegans* and *Drosophila* genes share two intron sites, and these are the only two sites that are similar in all three species. Both of these sites occur near the N-terminus of the protein: the second introns of the three genes are within three nucleotides, and the *Drosophila* and rat third introns and the *C. elegans* fourth intron occur at identical sites.

Homology to other acyltransferases. Database searches revealed that *C. elegans* ChAT has significant similarity to a number of mammalian carnitine acyltransferases. This is reasonable because of the similar reactions catalyzed by these enzymes: the acyl donors are all CoA derivatives and the acyl acceptors (choline and carnitine) are similar structurally. Sequences have been published for carnitine octanoyltransferase (CarOT) from rat (Chatterjee et al., 1988) and for carnitine palmitoyltransferase (CarPT) both from rat (Woeltje et al., 1990) and from human (Finocchiaro et al., 1991). Each of these has approximately 26–27% identity with *C. elegans* ChAT. We evaluated the relatedness of different enzyme activities from a single species, and also the relatedness of a single activity in different species. In the rat, each of the three enzymes (ChAT, CarOT, CarPT) has approximately 31% sequence identity with the other two, suggesting that the enzyme activities, although related, are never-

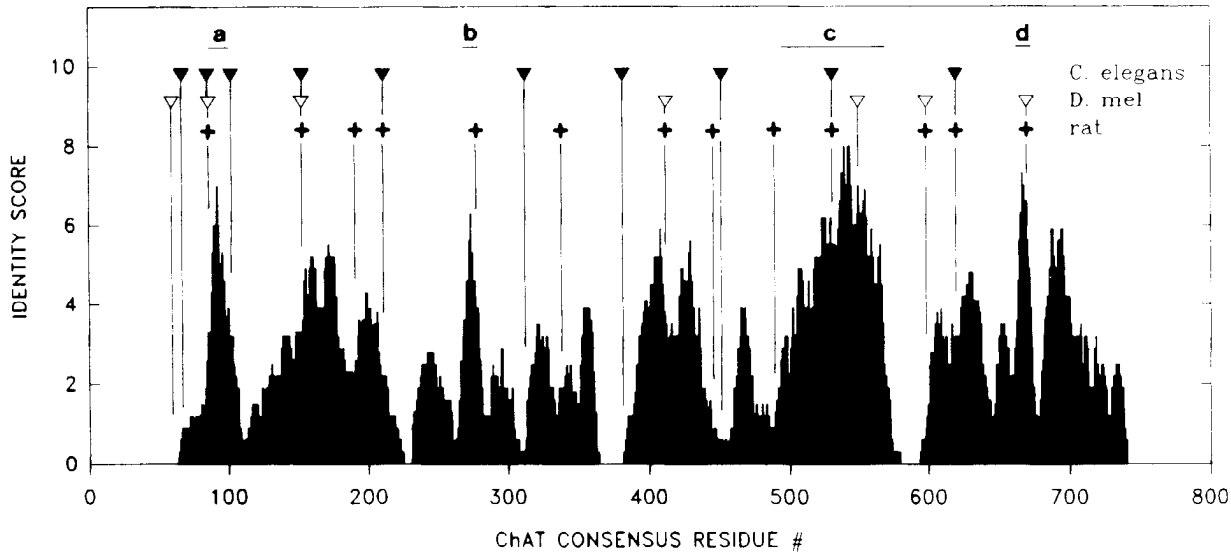

Figure 8. ChAT conserved domains and exon structures. The multiple sequence alignment in Figure 7 was scanned, one residue at a time, using a window of nine residues (from −4 to +4). Within this window, each three-way identity was given a score of 1.0, and each two-way identity was given a score of 0.3; the maximum possible identity score at any site is therefore 9.0. Random (i.e., unrelated) sequences would be expected to have an average identity score of approximately 0.4 (assuming that all amino acids are equally abundant). The four regions where the profile drops to the baseline (0–64, 225–230, 365–381, and 580–593) correspond to gaps inserted into one or two of the sequences to optimize the overall alignment. The *solid triangles* represent *C. elegans* intron sites; the *open triangles* represent *Drosophila* intron sites (Sugihara et al., 1991); the *plus signs* indicate rat intron sites 2–14 (coding sequence does not begin until the middle of the second intron; Hahn et al., 1992). The regions labeled *a–d* represent conserved domains discussed in Discussion.

theless quite distinct. Indeed, rat ChAT has slightly more similarity to *C. elegans* ChAT (36% identity) than it does to either of the rat carnitine acyltransferases. This argues that the three different acyltransferase enzymatic activities diverged from their common protein ancestor prior to the divergence of nematodes, insects, and vertebrates from their common animal ancestor.

The similarities between the ChAT, CarOT, and CarPT sequences are not confined to one or a few regions of the proteins, but rather are distributed along the molecules. In fact, most of the regions that display similarity among the ChAT sequences also show similarity among the more general class of acyltransferases, although with a poorer match. For example, the LPK–PVP–L motif present near the N-terminal of all ChATs becomes a motif of LP—P–P–L for the larger acyltransferase class (as has been previously observed by Finocchiaro et al., 1991). The 42 amino acid domain that has 26 three-way identities with ChATs (see above) shows 11 five-way identities (three ChATs, rat CarOT, rat CarPT).

Structure of the cha-1 *cDNAs.* The cDNA sequence and the derived intron–exon structure are based on data from five cDNAs, of which at least three are independent. The sequence and structure for the coding region of the gene were identical in all of the cDNAs; the only differences were slight variations in the 5′ and 3′ extent of the molecules (see Fig. 4). There is also some evidence from primer extension experiments for heterogeneity of the 5′ ends of *cha-1* mRNAs.

In the two cDNA libraries that were screened, *cha-1*–hybridizing plaques were rare, identified at a frequency of approximately 1 in 10^5. If this reflects the relative abundance of *cha-1* mRNA, then it is comparable to the relative abundance of ChAT protein, also estimated at approximately 1 part in 10^5 (Rand and Russell, 1985b).

Inspection of the *cha-1* cDNAs (and the corresponding genomic sequence) reveals some noteworthy features. The most striking of these is the likelihood that at least some of the *cha-1* mRNAs are *trans*-spliced with the spliced leader SL1, although this conclusion must remain tentative until additional and/or full length cDNAs are identified. *Trans*-splicing of mRNA is relatively common in *C. elegans*, and it is estimated that 10% of mRNAs are *trans*-spliced with one of two known leader sequences (Krause and Hirsh, 1987; Bektesh et al., 1988; Huang and Hirsh, 1989).

Another noteworthy feature of the *cha-1* cDNAs is the extremely long first intron. This ~6.5 kb intron is the third longest yet reported for *C. elegans*, exceeded only by an approximately 12 kb intron in the *lin-14* gene (Wightman et al., 1991), and a 7.4 kb intron in the *unc-22* gene (Benian et al., 1993). It is also of interest that the long *cha-1* intron separates the 5′ untranslated region of the transcript from the protein coding region. The ChAT structural gene from *Drosophila* also has a long first intron (>16.5 kb), although this intron lies within the protein coding sequence (Sugihara et al., 1991).

The remaining *cha-1* introns are unremarkable, by *C. elegans* standards, with lengths of 43–581 bp. Five of the introns lie within codons, and five lie between codons. The sequences around the 5′ and 3′ splice sites for all of the introns, including the long first intron, have excellent matches to the *C. elegans* splice site consensus sequences (Emmons, 1988).

Start of translation. We have assigned the ATG methionine codon at nucleotide position 76 as the start codon. This is the first ATG in the sequence; it is located at the beginning of the long open reading frame and has an in-frame termination codon nine bases upstream. However, this ATG has a poor, although minimally acceptable, fit with the Kozak (1987) consensus. Our sequence is ACAAGATGG, which has the important A at position −3 and the preferred G at +4, but only has one C at −4, rather than the preferred four Cs at positions −1, −2, −4, −5. It is known that *Drosophila* ChAT does not initiate protein

synthesis with a methionine, but instead uses a GTG valine codon (Sugihara et al., 1990). It has been suggested that a non-ATG initiation codon, or an ATG in a context with poor fit to Kozak's consensus might indicate the presence of translational regulation, and such regulation has been postulated (Sugihara et al., 1990) for *Drosophila* ChAT (non-ATG) and mammalian ChATs (poor context). Perhaps the *C. elegans* enzyme is also under translational regulation.

Transposon insertions into cha-1 *coding sequences can spare some gene function.* There is evidence to suggest that many Tc1 insertions do not lead to a null phenotype. For example, many of the Tc1-induced *unc-22* alleles have a weakly defective mutant phenotype (Moerman and Waterston, 1984). Since each of the four Tc1 *cha-1* mutations is homozygous viable, and since a complete loss of ChAT activity is lethal, each of these mutations must permit some residual gene activity. The Tc1 causing *md1143* is inserted into the splice donor site at the beginning of the second intron (Fig. 2). We assume that the effect of this mutation is to reduce the efficiency of mRNA splicing, but that there is enough residual splice donor activity to produce a small amount of normal *cha-1* mRNA. *md1422* lies in the final *cha-1* exon, and if it were not excised, the 89 amino acids at the C-terminal of the ChAT protein would be eliminated. Perhaps this truncated protein retains enough wild-type function to permit viability of the animals. *md1067* and *md1162* contain Tc1 insertions into the same site in the coding sequence near the end of the eighth exon. Perhaps the presence of the transposons activates a cryptic splice donor site upstream of the insertion site, thus "moving" the transposon into an intron. Such a mechanism has been demonstrated for a Tc1 insertion into coding sequences of the *mlc-2* gene of *C. elegans* (A. Rushforth and P. Anderson, personal communication), although inspection of the *cha-1* sequence near the *md1067* insertion site does not reveal any obvious alternative splice donor in the proper reading frame. Alternatively, it is possible that cryptic splice sites exist within the transposon, which could be weakly activated in certain contexts to remove most of the Tc1 sequences from the mutant *cha-1* transcripts. Such a mechanism has not been demonstrated in *C. elegans*, but has been shown to excise most of a *412* insertion and most of a *roo-412* compound insertion at a particular site in transcripts from the *Drosophila vermilion* gene (Pret and Searles, 1991). Sequencing of cDNAs from these Tc1 mutant strains should help to clarify these issues.

m324 *is a null mutation.* The *m324* mutation is a lethal *cha-1* allele (Rogalski and Riddle, 1988; Rand, 1989). Previous phenotypic and enzymatic analysis suggested that if there was any residual gene function, it was extremely low, but it was not possible to distinguish between an extreme hypomorph or a null allele (Rand, 1989). We have now shown that *m324* represents a 156 bp deletion, which effectively truncates the protein after residue 220 (see Results, and Figs. 2, 4). This represents only one-third of the wild-type protein, and does not include the catalytically essential histidine at residue 341 (see above); the *m324*-encoded gene product almost certainly lacks all ChAT activity. It will be interesting to determine whether a complete absence of neurotransmitter synthesis alters nervous system development.

p1156 *and* unc-17. *p1156* is genetically noteworthy because it fails to complement both *cha-1* and *unc-17* mutations (Rand and Russell, 1984), and because it lies between *cha-1* and *unc-17* on the genetic map (Rand, 1989). Although *unc-17* mutants have normal ChAT activity, they are resistant to AChE inhibitors and have elevated ACh levels (Brenner, 1974; Rand and Russell, 1984; Hosono et al., 1985), and we have recently obtained evidence that *unc-17* may encode a synaptic vesicle ACh transporter (Alfonso et al., 1993). We have now shown that *p1156* is associated with a 747 bp deletion near the 3′ end of the long *cha-1* intron (see Results). The failure of *p1156* to complement *unc-17* mutations suggested that *unc-17* coding or regulatory sequences might also be deleted by *p1156*. We have recently shown that this is indeed the case, and that all or part of the *unc-17* gene lies within the large *cha-1* intron (A. Alfonso, K. Grundahl, J. McManus, J. Asbury, and J. Rand, unpublished observations).

References

Alfonso A, Grundahl K, Duerr JS, Han H-P, Rand JB (1993) The *Caenorhabditis elegans unc-17* gene: a putative vesicular acetylcholine transporter. Science 261:617–619.

Alfonso-Pizarro A, Grundahl K, Rand JB (1989) Cloning the choline acetyltransferase structural gene from the nematode *C. elegans*. Soc Neurosci Abstr 15:62.

Alfonso-Pizarro A, Grundahl K, Rand JB (1990) Molecular analysis of the choline acetyltransferase structural gene from the nematode *C. elegans*. Soc Neurosci Abstr 16:199.

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

Bektesh S, Van Doren K, Hirsh D (1988) Presence of the *Caenorhabditis elegans* spliced leader on different mRNAs and in different genera of nematodes. Genes Dev 2:527–535.

Benian GM, L'Hernault SW, Morris ME (1993) Additional sequence complexity in the muscle gene, *unc-22*, and its encoded protein, twitchin, of *Caenorhabditis elegans*. Genetics 134:1097–1104.

Berrard S, Brice A, Lottspeich F, Braun A, Barde Y-A, Mallet J (1987) cDNA cloning and complete sequence of porcine choline acetyltransferase: *in vitro* translation of the corresponding RNA yields an active protein. Proc Natl Acad Sci USA 84:9280–9284.

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513.

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

Brice A, Berrard S, Raynaud B, Ansieau S, Coppola T, Weber MJ, Mallet J (1989) Complete sequence of a cDNA encoding an active rat choline acetyltransferase: a tool to investigate the plasticity of cholinergic phenotype expression. J Neurosci Res 23:266–273.

Carbini LA, Hersh LB (1993) Functional analysis of conserved histidines in choline acetyltransferase by site directed mutagenesis. J Neurochem 61:247–253.

Chalfie M, White J (1988) The nervous system. In: The nematode *Caenorhabditis elegans* (Wood WB, ed), pp 337–391. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Chatterjee B, Song CS, Kim JM, Roy AK (1988) Cloning, sequencing, and regulation of rat liver carnitine octanoyltransferase: transcriptional stimulation of the enzyme during peroxisome proliferation. Biochemistry 27:9000–9006.

Collins J, Saari B, Anderson P (1987) Activation of a transposable element in the germ line but not the soma of *Caenorhabditis elegans*. Nature 328:726–728.

Currier SF, Mautner HG (1974) On the mechanism of action of choline acetyltransferase. Proc Natl Acad Sci USA 71:3355–3358.

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395.

Eckenstein F, Sofroniew MV (1983) Identification of central cholinergic neurons containing both choline acetyltransferase and acetylcholinesterase and of central neurons containing only acetylcholinesterase. J Neurosci 3:2286–2291.

Eckenstein F, Barde Y, Thoenen H (1981) Production of specific antibodies to choline acetyltransferase purified from pig brain. Neuroscience 6:993–1000.

Edgley ML, Riddle DL (1990) The nematode *Caenorhabditis elegans*. Gen Maps 5:3.111–3.133.

Eide D, Anderson P (1985) Transposition of Tc1 in the nematode *C. elegans*. Proc Natl Acad Sci USA 82:1756–1760.

Eide D, Anderson P (1988) Insertion and excision of *C. elegans* transposable element Tc1. Mol Cell Biol 8:737–746.

Emmons SW (1988) The genome. In: The nematode *Caenorhabditis elegans* (Wood WB, ed), pp 47–79. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Emmons SW, Yesner L, Ruan K, Katzenberg D (1983) Evidence for a transposon in *Caenorhabditis elegans*. Cell 32:55–65.

Emmons SW, Roberts S, Ruan K-S (1986) Evidence in a nematode for regulation of transposon excision by tissue-specific factors. Mol Gen Genet 202:410–415.

Emson PC, Malthe-Sorenssen D, Fonnum F (1974) Purification and properties of choline acetyltransferase from the nervous system of different invertebrates. J Neurochem 22:1089–1098.

Finocchiaro G, Taroni F, Rocchi M, Martin AL, Colombo I, Tarelli GT, DiDonato S (1991) cDNA cloning, sequence analysis, and chromosomal location of the gene for human carnitine palmitoyltransferase. Proc Natl Acad Sci USA 88:661–665.

Gorczyca MG, Hall JC (1987) Immunohistochemical localization of choline acetyltransferase during development and in *cha*^{ts} mutants of *Drosophila melanogaster*. J Neurosci 7:1361–1369.

Gyllensten UB, Erlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. Proc Natl Acad Sci USA 85:7652–7656.

Hahn M, Hahn SL, Stone DM, Joh TH (1992) Cloning of the rat gene encoding choline acetyltransferase, a cholinergic neuron-specific marker. Proc Natl Acad Sci USA 89:4387–4391.

Herman RK, Horvitz HR (1980) Genetic analysis of *C. elegans*. In: Nematodes as biological models, Vol 1 (Zuckerman B, ed), pp 228–261. New York: Academic.

Herman RK, Shaw JE (1987) The transposable genetic element Tc1 in the nematode *C. elegans*. Trends Genet 3:222–225.

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

Hosono R, Kuno S, Midsukami M (1985) Temperature-sensitive mutations causing reversible paralysis in *Caenorhabditis elegans*. J Exp Zool 235:409–421.

Hosono R, Sassa T, Kuno S (1987) Mutations affecting acetylcholine levels in the nematode *Caenorhabditis elegans*. J Neurochem 49: 1820–1823.

Huang X-Y, Hirsh D (1989) A second trans-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. Proc Natl Acad Sci USA 86:8640–8644.

Ishii K, Oda Y, Ichikawa T, Deguchi T (1990) Complementary DNAs for choline acetyltransferase from spinal cords of rat and mouse: nucleotide sequences, expression in mammalian cells, and *in situ* hybridization. Mol Brain Res 7:151–159.

Itoh N, Slemmon JR, Hawke DH, Williamson R, Morita E, Itakura K, Roberts E, Shively J, Crawford GD, Salvaterra PM (1986) Cloning of *Drosophila* choline acetyltransferase cDNA. Proc Natl Acad Sci USA 83:4081–4085.

Johnson CD, Rand JB, Herman RK, Stern BD, Russell RL (1988) The acetylcholinesterase genes of *C. elegans*: identification of a third gene (*ace-3*) and mosaic mapping of a synthetic lethal phenotype. Neuron 1:165–173.

Kim SK, Horvitz HR (1990) The *Caenorhabditis elegans* gene *lin-10* is broadly expressed while required specifically for the determination of vulval cell fates. Genes Dev 4:357–371.

Kozak M (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148.

Krause M, Hirsh D (1987) A trans-spliced leader sequence on actin mRNA in *C. elegans*. Cell 49:753–761.

Levey AI, Armstrong DM, Atweh SF, Terry RD, Wainer BH (1983) Monoclonal antibodies to choline acetyltransferase: production, specificity, and immunohistochemistry. J Neurosci 3:1–9.

Malthe-Sorenssen D (1976) Choline acetyltransferase—evidence for acetyl transfer by a histidine residue. J Neurochem 27:873–881.

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Mautner HG, Pakula AA, Merrill RE (1981) Evidence for the presence of an arginine residue in the coenzyme A binding site of choline acetyltransferase. Proc Natl Acad Sci USA 78:7449–7452.

Moerman DG, Waterston RH (1984) Spontaneous unstable *unc-22 IV* mutations in *C. elegans* var Bergerac. Genetics 108:859–877.

Palatnik CM, Storti RV, Jacobson A (1979) Fractionation and functional analysis of newly synthesized and decaying messenger RNAs from vegetative cells of *Dictyostelium discoideum*. J Mol Biol 128: 371–395.

Pret A-M, Searles LL (1991) Splicing of retrotransposon insertions from transcripts of the *Drosophila melanogaster vermilion* gene in a revertant. Genetics 129:1137–1145.

Rand JB (1989) Fine-structure genetic analysis of the *cha-1–unc-17* gene complex in *Caenorhabditis*. Genetics 122:73–80.

Rand JB, Russell RL (1984) Choline acetyltransferase-deficient mutants of the nematode *Caenorhabditis elegans*. Genetics 106:227–248.

Rand JB, Russell RL (1985a) Molecular basis of drug-resistance mutations in the nematode *Caenorhabditis elegans*. Psychopharmacol Bull 21:623–630.

Rand JB, Russell RL (1985b) Properties and partial purification of choline acetyltransferase from the nematode *Caenorhabditis elegans*. J Neurochem 44:189–200.

Rogalski TM, Riddle DL (1988) A *Caenorhabditis elegans* RNA polymerase II gene, *ama-1 IV*, and nearby essential genes. Genetics 118: 61–74.

Rosenzweig B, Liao LW, Hirsh D (1983) Sequence of the *C. elegans* transposable element Tc1. Nucleic Acids Res 11:4201–4209.

Roskoski R Jr (1974) Choline acetyltransferase and acetylcholinesterase: evidence for essential histidine residues. Biochemistry 13:5141–5144.

Roskoski R Jr, Lim C-T, Roskoski LM (1975) Human brain and placental choline acetyltransferase: purification and properties. Biochemistry 14:5105–5110.

Ross J (1976) A precursor of globin messenger RNA. J Mol Biol 106: 403–420.

Rossier J (1976) Purification of rat brain choline acetyltransferase. J Neurochem 26:543–548.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467.

Selski DJ, Olschowka JA, Slemmon JR (1989) Evidence for precursor processing of choline acetyltransferase in *Drosophila melanogaster*. Soc Neurosci Abstr 15:63.

Slemmon JR, Salvaterra PM, Crawford GD, Roberts E (1982) Purification of choline acetyltransferase from *Drosophila melanogaster*. J Biol Chem 257:3847–3852.

Strauss WL, Kemper RR, Jayakar P, Kong CF, Hersh LB, Hilt DC, Rabin M (1991) Human choline acetyltransferase gene maps to region 10q11-q22.2 by *in situ* hybridization. Genomics 9:396–398.

Sugihara H, Andrisani V, Salvaterra PM (1990) *Drosophila* choline acetyltransferase uses a non-AUG initiation codon and full length RNA is inefficiently translated. J Biol Chem 265:21714–21719.

Sugihara H, Andrisani V, Salvaterra PM (1991) Genomic organization of *Drosophila* choline acetyltransferase. J Neurochem 57:1636–1642.

Sulston J, Brenner S (1974) The DNA of *C. elegans*. Genetics 77:95–104.

Sulston JE, Horvitz HR (1977) Post-embryonic cell lineages of the nematode *C. elegans*. Dev Biol 56:110–156.

Sulston JE, Schierenberg E, White JG, Thomson JN (1983) The embryonic cell lineage of the nematode *C. elegans*. Dev Biol 100:64–119.

White JG, Southgate E, Thomson JN, Brenner S (1986) The structure of the nervous system of the nematode *C. elegans*. Philos Trans R Soc Lond [Biol] 314:1–340.

Wightman B, Bürglin TR, Gatto J, Arasu P, Ruvkun G (1991) Negative regulatory sequences in the *lin-14* 3′-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. Genes Dev 5:1813–1824.

Woeltje KF, Esser V, Weis BC, Sen A, Cox WF, McPhaul MJ, Slaughter CA, Foster DW, McGarry JD (1990) Cloning, sequencing, and expression of a cDNA encoding rat liver mitochondrial carnitine palmitoyltransferase II. J Biol Chem 265:10720–10725.