The *Caenorhabditis elegans* cell death gene *ced-4* encodes a novel protein and is expressed during the period of extensive programmed cell death

JUNYING YUAN1,2,* and H. ROBERT HORVITZ2,†

1Program in Neuroscience, Harvard Medical School, Boston, MA 02115, USA
2Howard Hughes Medical Institute, Department of Biology, Room 56-629, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

*Present address: Cardiovascular Research Center, Massachusetts General Hospital-East, Charlestown, MA 02129, USA
†Author for correspondence

Summary

Mutations in the gene *ced-4* block almost all of the programmed cell deaths that normally occur during *Caenorhabditis elegans* development. We have cloned the *ced-4* gene using a *ced-4* mutation caused by the insertion of the transposon Tc4. When microinjected into a *ced-4* animal, a 4.4 kb DNA fragment derived from the wild-type strain and corresponding to the region of the Tc4 insertion in the mutant *ced-4(n1416)* rescues the Ced-4 mutant phenotype. The *ced-4* gene encodes a 2.2 kb RNA transcript. This mRNA is expressed primarily during embryogenesis, when most programmed cell deaths occur. The Ced-4 protein, as deduced from cDNA and genomic DNA clones, is 549 amino acids in length. Two regions of the putative Ced-4 protein product show some similarity to known calcium-binding domains.

Key words: *C. elegans*, programmed cell death, *ced-4*.

Introduction

Cell death is a fundamental aspect of animal development. Many cells die during the normal development of both vertebrates (Glucksmann, 1950) and invertebrates (Truman, 1984). These deaths appear to function in morphogenesis, metamorphosis and tissue homeostasis, as well as in the generation of neuronal specificity and sexual dimorphism (reviewed by Ellis et al., 1991a). An understanding of the mechanisms that cause cells to die and that specify which cells are to live and which cells are to die is essential for an understanding of animal development.

The nematode *Caenorhabditis elegans* is an appropriate organism for analyzing naturally occurring or programmed cell death (Horvitz et al., 1982). The generation of the 959 somatic cells of the adult *C. elegans* hermaphrodite is accompanied by the generation and subsequent deaths of an additional 131 cells (Sulston and Horvitz, 1977; Sulston et al., 1983). The morphology of cells undergoing programmed cell death in *C. elegans* has been described at both the light and electron microscopic levels (Sulston and Horvitz, 1977; Robertson and Thomson, 1982).

Many genes that affect *C. elegans* programmed cell deaths have been identified (Ellis and Horvitz, 1986; Ellis and Horvitz, 1991; Ellis et al., 1991b; reviewed by Ellis et al., 1991a). The activities of two of these genes, *ced-3* and *ced-4*, are required for the onset of almost all *C. elegans* programmed cell deaths (Ellis and Horvitz, 1986). When the activity of either *ced-3* or *ced-4* is eliminated, cells that would normally die survive instead and can differentiate into recognizable cell types and even function (Ellis and Horvitz, 1986; Avery and Horvitz, 1987; White et al., 1991). Genetic mosaic analyses have indicated that the *ced-3* and *ced-4* genes most likely act within dying cells, which suggests that the products of these genes are expressed within dying cells and either are cytotoxic proteins or control the activities of cytotoxic proteins (Yuan and Horvitz, 1990). To elucidate how *ced-3* and *ced-4* act within cells to cause their deaths, we have initiated molecular studies of these genes. In this paper, we describe the cloning and initial molecular characterization of the *ced-4* gene.

Materials and methods

General methods and strains

Techniques used for the culturing of *C. elegans* were essentially as described by Brenner (1974). All strains were grown at 20°C. DNA was prepared from worms grown on Petri dishes containing agarose seeded with *E. coli* strain HB101. RNA was prepared from mass cultures grown in liquid. Usually, the bacterial pellet from a 2 l overnight culture of *E. coli* HB101 grown in superbroth (12 g Bacto tryptone, 24 g yeast extract, 8 ml 50% glycerol, 900 ml H2O; after autoclaving, 100 ml 0.17 M KH2PO4 and 0.72 M K2HPO4 were added) was resuspended in 500 ml S basal medium (Brenner, 1974), and worms were added from one or two 10 cm Petri dishes in which the bacterial lawns had just been consumed. Worms were harvested about 4-5 days later by centrifug-
gation and washed in M9 buffer (Brenner, 1974). The yield was about 5-10 ml of packed worms.

Nomarski differential interference contrast microscopy was used to examine individual cells in living nematodes (Sulston and Horvitz, 1977). Methods for scoring the Ced phenotype of ced-1, ced-4 and ced-1;ced-4 double mutants have been described by Ellis and Horvitz (1986) and by Yuan and Horvitz (1990).

The wild-type parent of all mutant strains used in these experiments was C. elegans variety Bristol strain N2 (Brenner, 1974). The genetic markers used are listed below. These markers have been described by Brenner (1974), Hodgkin et al. (1988) and Finney et al. (1988). The strain TR679 carries the mutator mutant (r459) (Collins et al., 1987). The ced-4 alleles n1894, n1920, n1947, n1948, n2247 and n2273 have been characterized by us (unpublished results). Genetic nomenclature follows the standard system for C. elegans (Horvitz et al., 1979).

Genomic libraries

A 4.6 kb size-selected phage library was constructed from ced-4(nl1416) DNA as follows. Genomic DNA was digested with HindIII and run on a low-melting agarose gel. DNA migrating within the 4.6 kb size range was excised, and the low-melting agarose was removed by phenol extraction and precipitation (Maniatis et al., 1983). To total RNA by a poly(dT)-column (Maniatis et al., 1983). To this library had a total of 140,000 plaque-forming units (unpublished results). Genetic nomenclature follows the standard system for C. elegans (Horvitz et al., 1979).

LG I: ced-1(e1735), unc-54(r323)

LG III: unc-58(n1351), ced-4(n1162, n1416, n1894, n1920, n1947, n1948, n2247, n2273, n1416 n1712, n1416 n1713), unc-79(e1068), dpy-17(e164)

LG IV: unc-31(e928), ced-3(n717)

LG V: egl-1(e986), unc-76(e911)

Genomic libraries

Determination of DNA sequence

For determining DNA sequences, serial deletions were made according to Henikoff (1984). DNA sequences were determined using Sequenase and protocols obtained from US Biochemicals (Cleveland, OH). The ced-4 DNA sequence was confirmed by sequencing both strands of cDNA and genomic DNA clones.

Cloning of the cosmid fragment C10D8-5

The cosmid C10D8 was digested with EcoRI. Two EcoRI fragments of 2.2 kb (r5) and 2.4 kb (r7), both of which hybridized to a mixture of ced-4 cDNA subclones SK2-1 and SK2-2, were isolated. r7, which hybridized to SK2-1, the 3' half of ced-4 cDNA clone SK2, was cloned into the EcoRI site of plasmid pBSKII (Stratagene). The EcoRI site at the 3' end of r7 was deleted by digesting with SfiI, which cut once 0.2 kb to the 3' end of the insert, and SalI, which cut once in the polylinker, and then religating. The deleted r7 plasmid was linearized with EcoRI and ligated with EcoRI-digested r5, which hybridized to SK2-2, the 5' half of ced-4 cDNA clone SK2. Clones were analyzed for the correct orientation of the r5 insert based on the cDNA restriction map. One such correctly oriented clone was named C10D8-5.

DNA preparation, northern blot and primer extension

Total C. elegans RNA was extracted using guanidine isothiocyanate (Kim and Horvitz, 1990). Poly(A)+ RNA was selected from total RNA by a poly(dT)-column (Maniatis et al., 1983). To prepare stage-synchronized animals, we obtained eggs from gravid C. elegans adults grown at 20°C in liquid culture. A 5-10 ml sample of animals was treated with 50 ml of NaOCl/NaOH solution (10 ml NaOCl, 1 g NaOH, 40 ml H2O) for about 10 minutes with vortexing until the adults were dissolved. Eggs were centrifuged and washed three times with M9 buffer. Isolated eggs were allowed to hatch in S basal medium without food for 14 hours at 20°C with shaking. L1 larvae were collected by low-speed centrifugation after growth on E. coli HB101 for 2 hours, L2 larvae after 12 hours, L3 larvae after 24 hours, L4 larvae after 36 hours and adults after 48 hours. When we used DNA probes, our northern blot procedure was essentially that of Meyer and Casson (1986), except that we transferred RNA from the gel to the Gene Screen filter (DuPont, Wilmington, DE) by capillary action.

RNA preparation, northern blot and primer extension

Determination of DNA sequence

For determining DNA sequences, serial deletions were made according to Henikoff (1984). DNA sequences were determined using Sequenase and protocols obtained from US Biochemicals (Cleveland, OH). The ced-4 DNA sequence was confirmed by sequencing both strands of cDNA and genomic DNA clones.

Cloning of the cosmid fragment C10D8-5

The cosmid C10D8 was digested with EcoRI. Two EcoRI fragments of 2.2 kb (r5) and 2.4 kb (r7), both of which hybridized to a mixture of ced-4 cDNA subclones SK2-1 and SK2-2, were isolated. r7, which hybridized to SK2-1, the 3' half of ced-4 cDNA clone SK2, was cloned into the EcoRI site of plasmid pBSKII (Stratagene). The EcoRI site at the 3' end of r7 was deleted by digesting with SfiI, which cut once 0.2 kb to the 3' end of the insert, and SalI, which cut once in the polylinker, and then religating. The deleted r7 plasmid was linearized with EcoRI and ligated with EcoRI-digested r5, which hybridized to SK2-2, the 5' half of ced-4 cDNA clone SK2. Clones were analyzed for the correct orientation of the r5 insert based on the cDNA restriction map. One such correctly oriented clone was named C10D8-5.

Direction of transcription

The direction of transcription was determined by hybridizing northern blots with single-stranded RNA probes. The Bluescribe plasmid containing the insert pni1416 was linearized by digestion with either BamHI or HindIII, which cleaved at one or the other end of the insert. The linearized product was transcribed using T3 or T7 RNA polymerase, respectively, generating RNA from each strand. These RNA products were used to probe northern blots according to a protocol developed by Z. Liu and V. Ambros (personal communication). Filters were prehybridized in 50% formamide, 50 mM sodium phosphate (pH 6.5), 5 x SSC, 8 x Denhardt's, 0.5% SDS, 250 µg/ml salmon sperm DNA, then hybridized with probe at 55°C and washed in 4 x SSC, 0.1% SDS at 60°C x 3 times for 20 minutes each and then 2 x SSC, 0.1% SDS once at 60°C for 20 minutes. Northern blot experiments showed that the single-stranded RNA probe transcribed by T3 RNA polymerase hybridized to the 2.2 kb ced-4 mRNA, while the probe made by T7 RNA polymerase did not (data not shown). This result indicated that the direction of transcription is from the BamHI site toward the HindIII site of pni1416.

Microinjection and transformation

The procedure for microinjecting DNA into the gonad to obtain germline transformants was basically that of Fire (1986) with modifications introduced by J. Sulston (personal communication). Cosmid DNA to be injected was purified twice using CsCl-gradient centrifugation (Maniatis et al., 1983). Plasmid DNA to be injected was prepared by alkaline minipreps (Maniatis et al.,
C. elegans cell death gene ced-4

Results

Cloning of the ced-4 gene by transposon tagging

We isolated the ced-4 allele nl416 in the C. elegans strain TR679, which carries the mutator mut-2(r549) and shows an elevated frequency of transposition of several transposable elements (Collins et al., 1987; Yuan et al., 1991). The ced-4(nl1416) mutation is closely linked to a newly transposed copy of the C. elegans transposon Tc4 (Yuan et al., 1991). Using Tc4 as a probe, we cloned this novel Tc4 element and its flanking region as a 5 kb HindIII fragment from a 4-6 kb size-selected ced-4(n1416) genomic phage library. A 3 kb DNA fragment containing sequences adjacent to this Tc4 element was isolated by digesting the 5 kb HindIII fragment with BamHI. This 3 kb fragment, called pn1416, was cloned into the Bluescribe M13+ plasmid vector (Stratagene). When used as a probe on Southern blots, pn1416 recognized a 3.4 kb HindIII fragment in the wild type and in two spontaneous revertants of ced-4(nl1416), ced-4(nl1416 nl1712) and ced-4(nl1416 nl1713) (Yuan and Horvitz, 1991), and a 5 kb HindIII fragment in ced-4(nl416) animals (Fig. 1). These observations indicate that the Tc4 insertion in ced-4(nl1416) animals is responsible for their CED-4 mutant phenotype and suggest that pn1416 contains at least part of the ced-4 gene.

To isolate additional genomic DNA from the region of this Tc4 insertion, we used pn1416 to probe a C. elegans Bristol N2 genomic DNA phage library, provided by J. Sulston (personal communication). Five phage clones with inserts of 10 to 15 kb were isolated and shown to share a 3 kb BamHI-HindIII fragment that hybridized to pn1416 (data not shown). These phage clones were sent to J. Sulston and A. Coulson, who identified cosmids that hybridized to them and that were members of a 600 kb contig of overlapping cosmids (Coulson et al., 1986). Using the phage clones as probes to hybridize to Southern blots, we identified cosmids C10D8 as containing all regions of genomic DNA present in all five phage clones and in pn1416 (data not shown).

The CED-4 mutant phenotype can be rescued by a 4.4 kb DNA fragment

To identify ced-4(+) DNA capable of complementing the CED-4 mutant phenotype, we injected the cosmids C10D8 into the oocytes of ced-4(nl162) animals. To facilitate the identification of transgenic animals, a mutation in the unc-31 gene, which affects locomotion, was included as a marker for co-transformation (R. Hoskins and J. E. Sulston, personal communication; Kim and Horvitz, 1990). Cosmid C14G10, which contains the wild-type allele of unc-31 and does not have CED-4-rescuing activity (data not shown), was injected with cosmid C10D8 into ced-4(e1735);unc-31(e928);ced-4(nl162) DNA is 1.6 kb larger than that of the wild-type or the revertants, indicating that an insertion of this size is present in the ced-4(nl416) strain and is deleted in both revertants.

Ced-4 fusion protein and antibody preparation

To express a Ced-4 fusion protein in E. coli, we constructed (data not shown) a clone containing both the 5′ and 3′ halves of the ced-4 cDNA (SK2-2 and SK2-1) in the expression vector pET-5a (Rosenberg et al., 1987). The fusion protein expressed by this vector should include 11 amino acids of phage T7 gene 10 protein, 5 amino acids of linker and the 546 amino acids encoded by ced-4 cDNA SK2. The pJ76 plasmid, which encodes this fusion protein, was produced by this transformed strain as described by Rosenberg et al. (1987) and subjected to electrophoresis on a polyacrylamide gel. A band, with mobility equivalent to about 64 kDa, was excised and used to immunize three rabbits. Sera from all three rabbits tested positive on western blots (Towbin et al., 1979). These sera were purified using immunobLOTS (Harlow and Lane, 1988).

To identify additional genomic DNA from the region of this Tc4 insertion, we used pn1416 to probe a C. elegans Bristol N2 genomic DNA phage library, provided by J. Sulston (personal communication). Five phage clones with inserts of 10 to 15 kb were isolated and shown to share a 3 kb BamHI-HindIII fragment that hybridized to pn1416 (data not shown). These phage clones were sent to J. Sulston and A. Coulson, who identified cosmids that hybridized to them and that were members of a 600 kb contig of overlapping cosmids (Coulson et al., 1986). Using the phage clones as probes to hybridize to Southern blots, we identified cosmids C10D8 as containing all regions of genomic DNA present in all five phage clones and in pn1416 (data not shown).

The CED-4 mutant phenotype can be rescued by a 4.4 kb DNA fragment

To identify ced-4(+) DNA capable of complementing the CED-4 mutant phenotype, we injected the cosmids C10D8 into the oocytes of ced-4(nl162) animals. To facilitate the identification of transgenic animals, a mutation in the unc-31 gene, which affects locomotion, was included as a marker for co-transformation (R. Hoskins and J. E. Sulston, personal communication; Kim and Horvitz, 1990). Cosmid C14G10, which contains the wild-type allele of unc-31 and does not have CED-4-rescuing activity (data not shown), was injected with cosmid C10D8 into ced-4(e1735);unc-31(e928);ced-4(nl162) DNA is 1.6 kb larger than that of the wild-type or the revertants, indicating that an insertion of this size is present in the ced-4(nl416) strain and is deleted in both revertants.

Ced-4 fusion protein and antibody preparation

To express a Ced-4 fusion protein in E. coli, we constructed (data not shown) a clone containing both the 5′ and 3′ halves of the ced-4 cDNA (SK2-2 and SK2-1) in the expression vector pET-5a (Rosenberg et al., 1987). The fusion protein expressed by this vector should include 11 amino acids of phage T7 gene 10 protein, 5 amino acids of linker and the 546 amino acids encoded by ced-4 cDNA SK2. The pJ76 plasmid, which encodes this fusion protein, was produced by this transformed strain as described by Rosenberg et al. (1987) and subjected to electrophoresis on a polyacrylamide gel. A band, with mobility equivalent to about 64 kDa, was excised and used to immunize three rabbits. Sera from all three rabbits tested positive on western blots (Towbin et al., 1979). These sera were purified using immunobLOTS (Harlow and Lane, 1988).

To identify additional genomic DNA from the region of this Tc4 insertion, we used pn1416 to probe a C. elegans Bristol N2 genomic DNA phage library, provided by J. Sulston (personal communication). Five phage clones with inserts of 10 to 15 kb were isolated and shown to share a 3 kb BamHI-HindIII fragment that hybridized to pn1416 (data not shown). These phage clones were sent to J. Sulston and A. Coulson, who identified cosmids that hybridized to them and that were members of a 600 kb contig of overlapping cosmids (Coulson et al., 1986). Using the phage clones as probes to hybridize to Southern blots, we identified cosmids C10D8 as containing all regions of genomic DNA present in all five phage clones and in pn1416 (data not shown).

The Ced-4 mutant phenotype can be rescued by a 4.4 kb DNA fragment

To identify ced-4(+) DNA capable of complementing the CED-4 mutant phenotype, we injected the cosmids C10D8 into the oocytes of ced-4(nl162) animals. To facilitate the identification of transgenic animals, a mutation in the unc-31 gene, which affects locomotion, was included as a marker for co-transformation (R. Hoskins and J. E. Sulston, personal communication; Kim and Horvitz, 1990). Cosmid C14G10, which contains the wild-type allele of unc-31 and does not have CED-4-rescuing activity (data not shown), was injected with cosmid C10D8 into ced-4(e1735);unc-31(e928);ced-4(nl162) DNA is 1.6 kb larger than that of the wild-type or the revertants, indicating that an insertion of this size is present in the ced-4(nl416) strain and is deleted in both revertants.

Ced-4 fusion protein and antibody preparation

To express a Ced-4 fusion protein in E. coli, we constructed (data not shown) a clone containing both the 5′ and 3′ halves of the ced-4 cDNA (SK2-2 and SK2-1) in the expression vector pET-5a (Rosenberg et al., 1987). The fusion protein expressed by this vector should include 11 amino acids of phage T7 gene 10 protein, 5 amino acids of linker and the 546 amino acids encoded by ced-4 cDNA SK2. The pJ76 plasmid, which encodes this fusion protein, was produced by this transformed strain as described by Rosenberg et al. (1987) and subjected to electrophoresis on a polyacrylamide gel. A band, with mobility equivalent to about 64 kDa, was excised and used to immunize three rabbits. Sera from all three rabbits tested positive on western blots (Towbin et al., 1979). These sera were purified using immunobLOTS (Harlow and Lane, 1988).
Young L1 animals from these transformant strains had an average of 11.5 cell corpses in their heads, indicating that plasmid C10D8-5 restored *ced-4(+) activity as well as did cosmid C10D8 (Table 1).

**Identification of a ced-4 transcript**

We determined restriction maps of plasmid C10D8-5 and insert pn1416 and found that C10D8-5 overlapped with 2 kb of sequence in pn1416. (Fig. 2A). In northern blot experiments, we used both pn1416 and C10D8-5 to probe pol(A)+ RNA from populations of mixed developmental stages of wild-type (strain N2) and *ced-4(n1416)* animals. pn1416 hybridized to two transcripts in N2 animals: a 2.2 kb transcript and an 0.9 kb transcript (Fig. 2B). pn1416 hybridized to three transcripts from *ced-4(n1416)* animals: a 3.8 kb transcript, a transcript slightly larger than the wild-type 2.2 kb transcript, and a transcript slightly smaller than the wild-type 0.9 kb transcript. The 3.8 kb RNA contained Tc4 sequence (Fig. 2C), which suggests that this RNA resulted from the insertion of the 1.6 kb Tc4 sequence into the 2.2 kb *ced-4* transcript. The transcript slightly larger than the 2.2 kb wild-type transcript did not contain Tc4 sequence. This *ced-4(n1416)* RNA might have been an aberrant transcript containing sequences adjacent to the *ced-4* gene: when pn1416 was used as a probe, the wild-type 2.2 kb and this mutant 2.2 kb transcript were relatively similar in intensities (Fig. 2B), whereas when *ced-4* cDNA clone SK2-1 (see below) was used as a probe, this mutant 2.2 kb was not detected (Fig. 2E). These observations indicate that the *ced-4(n1416)* 2.2 kb transcript contained sequences from the *ced-4* region but did not contain sequences corresponding to at least the 3' half of the *ced-4* mRNA. The two revertants of *ced-4(n1416)*, *ced-4(n1416 n1712)* and *ced-4(n1416 n1713)*, contained both 2.2 kb and 0.9 kb transcripts of about the sizes of the wild-type transcripts (Fig. 2B). Thus, both the 2.2 kb and the 0.9 kb transcripts were altered in *ced-4(n1416)* animals, and both were restored in the two non-Ced revertants.

To determine whether one or both of these transcripts are encoded by *ced-4*, subclone C10D8-5, which rescued the *ced-4* phenotype, was used to probe a northern blot of RNA from N2 and *ced-4(n1416)* animals (Fig. 2D). C10D8-5 detected the wild-type 2.2 kb transcript, the *ced-4(n1416)* transcript slightly larger than the wild-type 2.2 kb transcript, and the *ced-4(n1416)* 3.8 kb transcript. C10D8-5 did not hybridize to the 0.9 kb transcript, indicating that this transcript is unlikely to be encoded by *ced-4*. C10D8-5 also detected a 1.4 kb transcript, which was not altered by the Tc4 insertion in *ced-4(n1416)* animals. Only a 470 bp *EcoRI-Stul* fragment at one end of C10D8-5 hybridized to this 1.4 kb RNA (data not shown). Since C10D8-5 did not contain the complete coding region for this RNA, and since this RNA was unaffected in *ced-4(n1416)* animals, this 1.4 kb RNA seems unlikely to be a *ced-4* transcript. The relationships among cosmid C10D8-5, insert pn1416 and the 0.9 kb, 1.4 kb and 2.2 kb transcripts are summarized in Fig. 2A.

On northern blots with *ced-4* cDNA clone SK2-1 (see below) as a probe (see below), the level of 2.2 kb transcript showed significant reduction in all three independently
derived EMS-induced *ced-4* mutants examined (Fig. 2E), strongly supporting the hypothesis that this 2.2 kb transcript is a *ced-4* transcript. That all of these distinct *ced-4* mutations (see below) caused reduced levels of a *ced-4* transcript could reflect either instability of all three mutant transcripts or a role for *ced-4* in regulating its own expression.

Fig. 2. *ced-4* encodes a 2.2 kb transcript. (A) The relationships among plasmid C10D8-5, plasmid insert pn1416 and three transcripts encoded in the *ced-4* region based upon DNA sequences, northern blots and restriction maps (see text). (B) pn1416 hybridizes to RNAs of 2.2 kb and 0.9 kb. A northern blot of poly(A)+ RNA from wild-type strain N2 and *ced-4*(n1416) animals, from two revertants of *ced-4*(n1416) [*ced-4*(n1416 n1712) and *ced-4*(n1416 n1713)] and from *ced-3*(n717) animals was probed with pn1416, a 3 kb fragment containing sequences adjacent to the site of Tc4 insertion in the mutant *ced-4*(n1416). pn1416 hybridized to RNAs of about 2.2 kb and 0.9 kb in N2 animals and to RNAs of 3.8 kb, slightly larger than 2.2 kb and slightly smaller than 0.9 kb in *ced-4*(n1416) animals. Both revertants had 2.2 kb and 0.9 kb transcripts with wild-type or near wild-type mobilities. *ced-3* animals appeared to have normal levels of both the 2.2 kb and the 0.9 kb transcripts. (C) The 3.8 kb transcript from *ced-4*(n1416) animals contains Tc4 sequence. A northern blot of poly(A)+ RNA from N2 and *ced-4*(n1416) animals was probed with Tc4-n1351, which contains the 1.6 kb Tc4 element present in the Tc4-induced mutant *unc-86*(n1351) as well as 4 kb of *unc-86* sequences (Yuan et al., 1991). Tc4-n1351 hybridized both to a 3.8 kb transcript of the Tc4-induced mutant *ced-4*(n1416) and to a 1.5 kb *unc-86* transcript in both *ced-4*(n1416) and N2 animals. (D) C10D8-5 hybridized to RNAs of 2.2 kb and 1.4 kb. A northern blot of poly(A)+ RNA from N2 and *ced-4*(n1416) animals was probed with C10D8-5, a 4.4 kb fragment that can rescue the *Ced-4* mutant phenotype. C10D8-5 hybridized to RNAs of 2.2 kb and 1.4 kb in N2 animals and of 3.8 kb, 2.2 kb and 1.4 kb in *ced-4*(n1416) animals. (E) The level of the 2.2 kb RNA transcript is reduced in *ced-4* mutants. A northern blot of total RNA from N2, *ced-4*(n1162), *ced-4*(n1416), *ced-4*(n1894) and *ced-4*(n1920) eggs was probed with 32P-labelled *ced-4* cDNA SK2-1. An actin 1 probe (Krause and Hirsh, 1984) was used as an internal control for the amount of RNA loaded in each lane. The ratios of the intensity of the 2.2 kb band to that of actin 1 band in N2, n1162, n1416 and n1894 were 0.5, 0.17, 0 and 0.12, respectively (data not shown).
Based upon these results, we conclude that the 2.2 kb RNA is a *ced-4* transcript. We do not know why the 0.9 kb RNA is also altered in *ced-4(n1416)* animals. Perhaps transcription of the 0.9 kb RNA is initiated incorrectly as a consequence of the nearby Tc4 element.

**ced-4 expression is primarily embryonic**

A northern blot containing RNAs from animals of different developmental stages showed that the 2.2 kb RNA was expressed primarily during embryogenesis (Fig. 3), consistent with the observation that 113 of the 131 programmed cell deaths in the *C. elegans* hermaphrodite are embryonic (Sulston and Horvitz, 1977; Sulston et al., 1983). The 2.2 kb RNA was relatively abundant during embryonic development: its level was about 20% that of actin 1 RNA (Edwards and Wood, 1983) during embryonic development, as determined by hybridizing duplicate northern blots with an actin 1 genomic DNA probe and an actin 1 cDNA probe of the same specific activity.

**The *ced-4* transcript is present in a *ced-3* mutant**

The activities of both *ced-3* and *ced-4* are required for programmed cell death (Ellis and Horvitz, 1986). One possibility is that one of these genes positively regulates the expression of the other. For this reason, we probed a Northern blot of wild-type strain N2 and *ced-3* poly(A)+ RNA with pn1416. This experiment showed that the 2.2 kb *ced-4* transcript was present at an apparently normal level in this *ced-3* mutant (Fig. 2B). Thus, the activity of the *ced-3* gene is unlikely to be necessary for the expression of the *ced-4* 2.2 kb transcript.

**Identification of *ced-4* cDNA clones**

To isolate cDNA clones of *ced-4*, we used pn1416 to probe a *C. elegans* cDNA phage library made from wild-type strain N2 mixed-stage RNA (Kim and Horvitz, 1990). Two cDNA clones were isolated. The two cDNA clones were isolated. The two cDNA clones (named SK1 and SK2) hybridized to the 2.2 kb *ced-4* transcript (data not shown). Both were about 1.8 kb in size, and both contained one 0.8 kb and one 1.0 kb EcoRI fragment. These EcoRI fragments were subcloned into plasmid vector Bluescribe M13+ (Stratagene). The two subclones derived from SK1 were named SK1-1 and SK1-2, and the two subclones derived from SK2 were named SK2-1 and SK2-2. The restriction maps of the SK1-1 and SK2-derived clones were the same (data not shown). Sequence analysis of the ends of the four cDNA subclones confirmed the equivalence of the SK1 and SK2 clones, except that SK1-2 contained a poly(A) sequence of more than 50 bp at its 5′ end. This poly(A) sequence was probably a cDNA cloning artifact, since SK1-2 contains the 5′ half of the cDNA (see below).

**The *ced-4* sequence**

We determined the DNA sequence of the SK2 1.8 kb cDNA clone. This sequence includes an open reading frame encoding 546 amino acids (Fig. 4A). This open reading frame is consistent with the results of northern blot analysis using single-stranded RNA probes (see Materials and methods). An ochre termination codon (TAA) is located in-frame near the 3′ end, indicating that the 3′ end of the 2.2 kb transcript is most likely included in this cDNA. The open reading frame extends to the 5′ end of the 1.8 kb cDNA, suggesting that this cDNA might lack the 5′ end of the *ced-4* coding region. We performed a primer extension experiment (see Materials and methods) and identified a major transcriptional initiation site 54 bp 5′ to the beginning of the *ced-4* cDNA SK2 and a minor initiation site 54 bp 3′ to the beginning of this cDNA (Fig. 4B). The first AUG codon after the presumptive major start site is located 9 bp 5′ to the beginning of the cDNA (Fig. 4A). If this site is used to ini-

---

**Table 2. Sites of mutations in the *ced-4* gene**

<table>
<thead>
<tr>
<th>Allele</th>
<th>Mutation</th>
<th>Nucleotide</th>
<th>Codon</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>n1162</td>
<td>C to T</td>
<td>1131</td>
<td>40</td>
<td>Q to ochre (TAA)</td>
</tr>
<tr>
<td>n2274</td>
<td>C to T</td>
<td>1428</td>
<td>139</td>
<td>R to opal (TGA)</td>
</tr>
<tr>
<td>n1920 and n2247</td>
<td>G to A</td>
<td>1744</td>
<td>first of 5′ splice donor of intron 3</td>
<td>Altered splicing</td>
</tr>
<tr>
<td>n2273</td>
<td>G to A</td>
<td>1929</td>
<td>first bp of 3′ splice acceptor of intron 3</td>
<td>Altered splicing</td>
</tr>
<tr>
<td>n1948</td>
<td>T to A</td>
<td>2117</td>
<td>258</td>
<td>I to N</td>
</tr>
<tr>
<td>n1947</td>
<td>C to T</td>
<td>2128</td>
<td>262</td>
<td>Q to amber (TAG)</td>
</tr>
<tr>
<td>n1894</td>
<td>G to A</td>
<td>3131</td>
<td>401</td>
<td>W to opal (TGA)</td>
</tr>
</tbody>
</table>

Nucleotide and codon positions correspond to the numbering in Fig. 4.

---

**Fig. 3.** The *ced-4* transcript is expressed mostly during embryogenesis. (a) A northern blot of poly(A)+ RNA from stage-synchronized animals was probed using pn1416, which hybridizes both to the 2.2 kb *ced-4* transcript and to a 0.9 kb transcript (see Fig. 2A). Only the 2.2 kb transcript is shown in this figure. The 0.9 kb transcript is expressed mostly in eggs and adults (data not shown). (b) The presence of RNA in all lanes was confirmed by loading 1/10 of each sample on another gel and probing a northern blot from this gel using the pn1416 cDNA probe and an actin 1 genomic DNA probe.
tiate protein synthesis, the Ced-4 protein would be 549 amino acids in length. The first AUG codon after the presumptive minor start site is located 130 bp 3' to the beginning of the cDNA. If this site is used, the Ced-4 protein would be 503 amino acids in length. Preliminary results using an anti-Ced-4 antibody raised against a Ced-4 fusion protein showed that endogenous Ced-4 protein is slightly smaller in molecular weight than a Ced-4 fusion protein of 562 amino acids expressed in E. coli. The longest hydrophobic region is a segment of 12 amino acids from residues 382 to 393.

The exon sequences of genomic clone C10D8-5 were identified by available computer search algorithms, either because of variability among the sequences or because the amino acids conserved are not clustered within the primary sequence. By direct inspection, we compared the sequence of the putative Ced-4 protein with the consensus sequence of the calcium-binding loop of the EF-hand domain (Tufty and Kretsinger, 1975; Kretsinger, 1987; Szébenyi and Moffat, 1986). We identified two regions of the Ced-4 protein that might bind calcium (Fig. 5).

The EF-hand is a 29 amino acid domain consisting of a helix-loop-helix region, with the loop portion coordinating calcium-binding via side-chain oxygens from serine, threonine, asparagine, aspartic acid, glutamine or glutamic acid. These residues occur at five of the vertices of an octahedron: X (position 10), Y (12), Z (14), −X (18), −Z (21). EF-hand amino acid sequences vary considerably in the residues present in the calcium-binding loop (Fig. 5), and some EF-hand domains have only one helical region (Kretsinger, 1987). Thus, searching protein data banks with a calcium-binding amino acid sequence could fail to identify a protein with the EF-hand motif.

In its putative calcium-binding loop, the first potential EF-hand-like sequence of the Ced-4 protein has four (positions Y, Z, −X, −Z) of the five conserved residues with oxygen-containing side chains (shown in bold), and the fifth position (X) has a tyrosine rather than an aspartic acid; tyrosine contains oxygen in its side chain. The second potential EF-hand-like sequence of the Ced-4 protein has three residues (positions Z, −X, −Z) that match the consensus sequence, and amino acids with oxygen-containing side chains at the other two positions. These observations suggest that these two regions of the Ced-4 protein might bind calcium. Like the Ced-4 protein, a number of known calcium-binding proteins, such as bovine intestinal calcium-binding protein (ICaBP) (Szébenyi and Moffat, 1986), rabbit troponin C (Collins et al., 1973), trypsinogen and villin (Doolittle, 1984; Dang et al., 1985) have only three or four conserved residues at these five positions (Fig. 4). The EF-hand domains in ICaBP and troponin C have been shown by X-ray crystallography to bind calcium. One major difference between the Ced-4 protein and the calcium-binding loop of the EF-hand consensus sequence is at position 15: the two Ced-4 sequences have a histidine and a glutamic acid, respectively, whereas most EF-hand-containing proteins have a glycine; this glycine has been suggested to be important for the turning of the loop (Kretsinger, 1987). However, a histidine is present at this position in a parvalbumin and an aspartic acid is present in another parvalbumin and also in a sarcoplasmic calcium-binding protein (Kretsinger, 1987) (Fig. 5). Thus, the presence of histidine or glutamic acid at position 15 does not rule out the possibility that these regions bind calcium.

The calcium-binding loop (positions 10-21) of the EF-hand is thought to be preceded (positions 1-9) and followed by alpha-helical domains (positions 22-29) (Kretsinger, 1987). Since position 3 of Ced-4 sequence 1 and positions 26 and 28 of Ced-4 sequence 2 are prolines, these regions might not form alpha-helices. However, the known calcium-binding protein galactose-binding protein (GBP) has a calcium-binding domain similar to that of the EF-hand (Fig. 5) but without the two helices; furthermore, position 29 of GBP is proline (Vyas et al., 1987). Thus, the Ced-4 protein need not contain such alpha-helical domains to calcium binding.

Based upon these considerations, it seems possible that the Ced-4 protein binds calcium or a similar divalent cation.
Fig. 4. Genomic organization and nucleotide sequence of ced-4 and deduced amino acid sequences. (A) The genomic sequence of the ced-4 region, as obtained from plasmid C10D8-5, which rescues the Ced-4 mutant phenotype. The deduced amino acid sequence of the Ced-4 protein is based upon the DNA sequence of the ced-4 cDNA SK2 and other experiments described in Results. The likely start sites of transcription are marked with downward arrows. The start of the cDNA is marked with an arrowhead. The positions of eight ced-4 mutations are indicated with upward arrows. Numbers on the sides indicate nucleotide positions, beginning with the start of C10D8-5. Numbers under the amino acid sequence indicate codon positions. Vertical lines between nucleotides indicate splice junctions. (B) A primer extension experiment to determine the ced-4 transcription initiation site(s). Lane 1, primer extension reaction using primer ATTGGCGATCCTCTCGA. Lane 2, DNA sequencing reaction using C10D8-5 as template and the same primer (see Materials and methods). The sequences around the transcription initiation sites are shown. The nucleotide numbers given in the figure (numbered as in A of this figure) indicate the possible start sites of transcription. (C) Western blot of wild-type strain N2 mixed-stage (lane 1), ced-4(n1416) mixed-stage (lane 2), wild-type egg (lane 3), ced-4(n1416) egg (lane 4) and bacteria-expressed protein (lane 5) probed using anti-Ced-4 antibody. Ced-4 fusion protein (pJ76) was made by cloning ced-4 cDNA SK2 into T7 expression vector pET-5a (Rosenberg et al., 1987; see Materials and methods), so that 546 amino acids of Ced-4 sequence were fused to 11 amino acids of T7 gene 10 protein and 5 amino acids of linker sequence. An arrow indicates that the Ced-4 fusion protein is similar in relative molecular mass to the endogenous Ced-4 protein, which is present in wild-type
(N2) but missing in ced-4(n1416) animals. The relative molecular mass standards were phosphorylase b, $97 \times 10^3$; bovine serum albumin, $66 \times 10^3$ (Hirayama et al., 1990); and ovalbumin, $43 \times 10^3$. (D) Genomic structure of the ced-4 gene and positions of ced-4 mutations. The sizes of exons and introns are indicated (in bp). The solid arrows indicate the positions of the Tc4 insertion in the ced-4(n1416) mutant and of eight EMS-induced mutations of ced-4. The solid arrow pointing right indicates the direction of transcription. The closed arrowhead indicates the site of the initiation of translation. The open arrowhead indicates the site of the ochre termination codon.
Fig. 5. Sequence similarities between the presumptive Ced-4 protein and some calcium-binding proteins. The calcium-binding loop of the 29 amino-acid EF-hand consists of the 12 amino acids from position 10 to position 21. The consensus sequence of this calcium-binding loop is shown at the top (Kretsinger, 1987). The positions indicated by X, Y, Z, −X, −Z correspond to vertices of an octahedron (see text). Amino acids are indicated by the single letter code. O, amino acid with an oxygen-containing side chain; *, non-conserved amino acid. Positions Y, Z and −X can have any of a number of oxygen side chain-containing amino acids; position X is usually aspartic acid, and position −Z is usually glutamic acid. Conserved amino acids are shown in bold face. Deviations from the EF-hand consensus sequence are underlined. The sequences of parvalbumins from carp muscle (Nockolds et al., 1972), the intestinal calcium-binding protein (ICaBP) (Szczephezy et al., 1981), troponin C (Collins et al., 1973) and calmodulin (Zimmer, W. E. et al., 1988; Babu et al., 1985) show canonical EF-hands. The hake and ray parvalbumins (Capony et al., 1973; Thatcher and Pechere, 1977), sarcoplasmic calcium-binding protein (SCBP) from the protochordate Amphioxus (Takagi and Konishi., 1986), trypsinogen (Bode and Schwager, 1975), fibrinogen (Doopiltte, 1984; Tang et al., 1985), villin (Hesterberg and Weber, 1983) and galactose-binding protein (GBP) (Vyas et al., 1987) show variations from the consensus sequence. GBP does not contain the helices of the EF-hand. The potential calcium-binding loop of sequence 1 of the hake parvalbumin (hpa) (see text). Conserved amino acids are shown in bold face. Deviations from the EF-hand consensus are underlined. The sequences of the 29 amino-acid EF-hand consists of the 12 amino acids from position 10 to position 21. The consensus sequence of this calcium-binding loop is shown at the top (Kretsinger, 1987). The positions indicated by X, Y, Z, −X, −Z correspond to vertices of an octahedron (see text). Amino acids are indicated by the single letter code. O, amino acid with an oxygen-containing side chain; *, non-conserved amino acid. Positions Y, Z and −X can have any of a number of oxygen side chain-containing amino acids; position X is usually aspartic acid, and position −Z is usually glutamic acid. Conserved amino acids are shown in bold face. Deviations from the EF-hand consensus sequence are underlined. The sequences of parvalbumins from carp muscle (Nockolds et al., 1972), the intestinal calcium-binding protein (ICaBP) (Szczephezy et al., 1981), troponin C (Collins et al., 1973) and calmodulin (Zimmer, W. E. et al., 1988; Babu et al., 1985) show canonical EF-hands. The hake and ray parvalbumins (Capony et al., 1973; Thatcher and Pechere, 1977), sarcoplasmic calcium-binding protein (SCBP) from the protochordate Amphioxus (Takagi and Konishi., 1986), trypsinogen (Bode and Schwager, 1975), fibrinogen (Doopiltte, 1984; Tang et al., 1985), villin (Hesterberg and Weber, 1983) and galactose-binding protein (GBP) (Vyas et al., 1987) show variations from the consensus sequence. GBP does not contain the helices of the EF-hand. The potential calcium-binding loop of sequence 1 of the putative Ced-4 protein is located at amino acids 77 to 88, and of sequence 2 is at amino acids 292 to 303.

Discussion

We cloned the ced-4 gene by isolating a Tc4 transposon-induced ced-4 allele and demonstrating that a 4.4 kb DNA fragment defined by the site of transposon insertion can rescue the Ced-4 mutant phenotype when microinjected into a ced-4 mutant strain. We identified a 2.2 kb RNA as a ced-4 transcript. This transcript is expressed mostly during embryonic development, when 113 of the 131 programmed cell deaths occur. Sequence analyses of cDNA and genomic clones indicated that the Ced-4 protein is 549 amino acids in length. The Ced-4 protein is highly hydrophilic, with no obvious transmembrane domains. Of the eight EMS-induced ced-4 alleles, one results in a single amino acid substitution and the other seven appear to prevent either ced-4 RNA splicing or completion of Ced-4 protein synthesis. These seven mutations establish the null phenotype of the ced-4 gene, confirming that ced-4 function is not essential for viability.

The putative Ced-4 protein has two regions that show some sequence similarity to known calcium-binding domains (Kretsinger, 1987). Although biochemical experiments are clearly essential to test if the Ced-4 protein can indeed bind calcium, it is interesting to speculate that Ced-4 activity and hence, programmed cell death might be modulated by calcium. Calcium has been implicated as an essential mediator of cell death in other organisms under a variety of conditions. For example, extracellular calcium is required for glucocorticoid-induced thymocyte death (Cohen and Duke, 1984), for the deaths of adult rat hepatocytes induced by certain toxins in vitro (Schanne et al., 1979), for agonist-induced muscle degeneration in mice (Leonard and Salpeter, 1979) and for neuronal cell death caused by oxygen deprivation or excitotoxicity (Coyle et al., 1981; Choi, 1987, 1988). Perhaps programmed cell death during C. elegans development is initiated by an increase in intracellular calcium, which activates the Ced-4 protein to become cytotoxic. Alternatively, certain cells seem to be protected against cell death by calcium (eg., Koike et al., 1989; Collins et al., 1991), so it is possible that increases in intracellular calcium levels inhibit the activity of the Ced-4 protein and thereby prevent programmed cell death.

The level of the ced-4 transcript in eggs is about 20% of the actin 1 transcript, which is very prevalent (Edwards and Wood, 1983). This level seems higher than might be expected if ced-4 were expressed only in dying cells, since in an embryo there are usually no more than two or three cells dying at the same time. These considerations suggest that ced-4 might be transcribed not only in dying cells but in other cells as well. Perhaps ced-4 activity, at least during embryonic development, is regulated at a post-transcriptional level. For example, the Ced-4 protein might have to interact with other proteins or other factors (such as calcium) to cause cell death. Since the ced-3 gene is also essential for programmed cell death in C. elegans, one possibility is that the activity of the Ced-4 protein is dependent upon ced-3 function.

Genetic mosaic analysis has indicated that ced-4 gene activity is needed within those cells that undergo programmed cell death (Yuan and Horvitz, 1990). The deaths of thymocytes in vitro induced by glucocorticoids and of vertebrate neurons in vitro in response to growth factor deprivation seem to be caused by the initiation of a cellular suicide program, since these deaths can be prevented by RNA and protein synthesis inhibitors (Cohen and Duke,
1984; Martin et al., 1988). Similarly, cell death in vivo during vertebrate neuronal development and prostate regression and during the metamorphosis of the moth Man - duca sexta can also be inhibited by blocking RNA or protein synthesis (Stanisic et al., 1978; Oppenheim et al., 1990; Fahrbach and Truman, 1987). Perhaps genes similar to ced-3 and ced-4 function in other organisms to cause cell death. The cloning of the C. elegans ced-4 gene provides a reagent that can be used to seek similar genes from other organisms.

We thank John Collins and Phil Anderson for the TR679 strain, Stuart Kim for cDNA libraries, John Sulston and Alan Coulson for a C. elegans genomic library and for mapping our ced-4 phage clones, Cori Bargmann for pointing out similarities between the presumptive Cед-4 protein and calcium-binding proteins, F. W. Studier and Li Lin for pET vectors, Ron Ellis, Michael Hengartner and Josh Kaplan for giving us the new ced-4 mutants n1894, n1920, n1947, n1948, n2247, n2273 and n2274, Hong Wu and Rudolf Jaenisch for letting us use their β counter, Greg Petsko for discussions about calcium-binding domains, Hong Zhu for help in making figures, and Barbara Osborne, Shai Shaham and Michael Hengartner for critical reading of the manuscript. This work was supported by US Public Health Service Research Grants B63061 and GM24663 and GM24943 and by the Howard Hughes Medical Institute. J.Y. was supported by a scholarship from Harvard University, by the Howard Hughes Medical Institute and by the Cardiovascular Research Center of the Massachusetts General Hospital. H.R.H. is an Investigator of the Howard Hughes Medical Institute.

References


Bode, W. and Schwager, P. (1975). The refined crystal structure of bovine beta trypsin at 1.8A resolution. II. Crystallographic refinement, calcium binding site, benzamidine binding site and active site at pH 7.0. J. Mol. Biol. 98, 693-717.


(Accepted 24 June 1992)