Homeobox containing genes in the nematode *Caenorhabditis elegans*

Nancy C.Hawkins* and James D.McGhee
Department of Medical Biochemistry, University of Calgary, Calgary, Alberta T2N 4N1, Canada

Received May 25, 1990; Revised and Accepted August 31, 1990 EMBL accession nos X52810–X52813 (incl.)

ABSTRACT

We designed a unique 36-mer oligonucleotide probe, based on the most highly conserved amino acid sequences of *Antennapedia*-like homeodomains and the codon bias of *Caenorhabditis elegans*. This probe was then used to isolate four classes of genes from a *C. elegans* genomic library. Sequencing reveals that we have isolated three new homeobox genes, designated *ceh-1*, *ceh-9* and *ceh-10*. The fourth homeobox gene, *ceh-11*, has recently been described by Schaller *et al* (Nucleic Acids Res. *18*, 2033–2036). The amino acid sequence of *ceh-1* is 87% similar to the honeybee H40 homeodomain, 85% similar to the Drosophila NK-1 homeodomain and 82% similar to the chicken *CHox3* homeodomain. The sequence *ceh-10* appears to be a member of the *paired* class of homeodomains. The other two sequences, *ceh-9* and *ceh-11*, remain unclassified. Three of the four sequences have at least one intron within the homeobox region. Transcripts of *ceh-10* and *ceh-11* are present in embryonic RNA but are greatly diminished in later developmental stages. Three of the four new genes have been placed on the *C. elegans* genomic map.

INTRODUCTION

Many proteins that control transcription contain a 'homeodomain' motif within their primary sequence (see for example, the compilation in 1). Indeed, if a newly-identified gene is found to contain a 'homeobox' sequence, this is taken as preliminary evidence that the gene will turn out to play some role in genetic regulation. By now, homeobox sequences have been identified in a wide variety of organisms and reasonably stringent criteria have been established for membership in a particular homeobox class (see review in 2).

Homeobox containing genes have been identified by two approaches. The first is to clone an interesting gene and then to discover that the gene sequence contains a homeobox motif. This approach has been successful with four genes, *mec-3* (3), *mab-5* (4), *unc-86* (5) and *lin-11* (6), that control particular cell fates during *C. elegans* development. The second approach is simply to identify sequences on the basis of cross-hybridization with known homeobox probes. While this approach has been highly effective with other organisms, until quite recently it has not worked with *C. elegans*. This failure was possibly because of sequence divergence or because of unusual codon bias or, as will be shown below, because *C. elegans* homeobox sequences tend to be split by introns, a feature not common in other organisms.

During the last year, three sets of homeobox genes have been isolated from *C. elegans* on the basis of low stringency hybridization. Burglin et al (7) have used a degenerate oligonucleotide probe to identify homeobox sequences in a *C. elegans* library and have estimated that the *C. elegans* genome might contain as many as 60 such sequences. Three homeobox regions were completely sequenced and another five regions sequenced partially. In addition, Kamb et al (8) have used PCR with degenerate primers to isolate several homeobox candidates in *C. elegans*; partial sequencing has identified homologs of the *Antennapedia*, *engrailed* and *paired* genes of *Drosophila*. Most recently, Schaller et al (9) have used a homeobox probe from the parasitic nematode *Ascaris lumbricoides* to isolate two further homeobox sequences from *C. elegans*, as well as a portion of a third homeobox candidate.

In the present paper, we use a unique oligonucleotide probe to isolate three new homeobox sequences from *C. elegans* as well as one of the sequences isolated by Schaller *et al.* (9). We report the sequences of the homeobox regions, compare these sequences to those of previously studied homeodomain containing genes and show that transcripts for two of the genes are enriched in early *C. elegans* embryos. Finally, three of the four sequences are assigned to positions on the *C. elegans* genomic map (10, 11).

MATERIALS AND METHODS

Worm growth and DNA isolation were by standard methods (12, 13). Essentially all recombinant methods were also standard (14); details can be found in (15). The oligonucleotide probe (see Figure 1 below) was obtained from the Regional DNA Synthesis Facility, University of Calgary. The *C. elegans* genomic library was prepared in λ EMBL4 and was kindly provided by Drs. Chris Link and William Wood, University of Colorado, Boulder.

For the Southern analysis shown below in Figure 2, 5 μg

* To whom correspondence should be addressed at Biology Department, Princeton University, Princeton, NJ 08544, USA

aliquots of *C. elegans* genomic DNA were digested to completion with various restriction endonucleases according to the manufacturers instructions, electrophoresed on a 1% agarose gel and blotted onto Zetaprobe membrane (BioRad) by capillary transfer in 0.4N NaOH (16). Blots were prehybridized for 2 hours at 42°C in 5×SSPE (where 1×SSPE = 0.15M NaCl, 0.1 mM EDTA, 0.1 mM $Na_4P_2O_7$, 25 mM sodium phosphate, pH 8.0) containing 20% formamide, 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin and 1% sodium dodecylsulphate. The addition of salmon sperm DNA was avoided since it contributed to high backgrounds. Oligonucleotide probes were labelled to a specific activity of about 5×10^8 cpm/μg using T4 polynucleotide kinase and [γ-^{32}P]-ATP (7000 Ci/mmol). Hybridization was in a fresh aliquot of the prehybridization solution, to which $1-2\times10^6$ cpm/ml of end-labelled oligonucleotide had been added. Filters were hybridized overnight at 42°C, washed for 15 minutes in 1×SSPE, 0.5% SDS at room temperature with shaking and then washed twice in the same solution for 20 minutes each at 47°C.

Di-deoxy sequencing was used throughout, with both single-stranded and double-stranded templates, and usually proceeded by preparing a set of nested unidirectional deletions as described in (17). Sequences shown in Figure 3 were determined from both strands of appropriate subclones; a more detailed account of the subcloning and sequencing strategy is given in (15). Some of the subclones containing the homeobox sequence *ceh-10* appeared to be unstable in the host strain JM109; such plasmids were found to be stable in the host strain JC8111 (18).

RNA was prepared from different developmental stages of *C. elegans* by disruption of worms in guanidinium isothiocyanate and centrifugation of the RNA through a cushion of cesium trifluoroacetate (19). Poly-A+ RNA was isolated by conventional oligo-dT cellulose chromatography. For the Northern analysis shown below in Figure 6, five μg of Poly-A+ RNA were electrophoresed on a 1.25% agarose gel containing formaldehyde (14) and transferred to Zetaprobe in 50 mM NaOH for 3 hours. The restriction fragments used as probes were labelled by random-priming (20).

RESULTS AND DISCUSSION

To design an oligonucleotide probe, we compiled sequences of *Antennapedia*-like homeobox motifs from a variety of organisms, selected the highly conserved region in the middle of the 'recognition helix' (2) and converted the amino acid sequence into a single nucleic acid sequence using the (rather extreme) *C. elegans* codon bias as compiled in (13). The resulting unique 36-mer probe is shown in Figure 1.

C. elegans genomic DNA was digested with a number of restriction endonucleases and a Southern blot was probed with the ^{32}P-labelled oligonucleotide. The resulting autoradiograph (Figure 2) shows that each digest produces a small number, usually 6−7, of distinct bands of roughly equal intensity. Lowering the hybridization or washing stringency does not greatly increase the number of detected bands (not shown) and we have probably identified the entire set of sequences detectable by the 36-mer probe. Two digests (BamHI and PvuII) show only 4−5 bands. Thus there is the possibility that several of the detected sequences could be clustered. However, short-range clustering can not be extensive since frequently cutting endonucleases, such as HaeIII, still reveal a maximum of 7 bands (data not shown).

Using the hybridization and washing conditions established in Figure 2, seven genomic equivalents of a *C. elegans* genomic library were screened and 42 positive clones were isolated, close to the number expected if the genome contains 7 separate fragments detectable by our probe. Sixteen of these clones were selected at random, plaque-purified and assigned to four distinct classes by restriction mapping. Sequences that hybridized to the 36-mer probe were isolated either by conventional subcloning or by the random subcloning of 200−400 base pair fragments produced from the bacteriophage DNA by sonication.

DNA sequencing immediately revealed that we had indeed isolated several new candidates for *C. elegans* homeobox sequences, as shown in Fig. 3: each proposed homeobox sequence is underlined and intron sequences are in lower case. By agreement with other workers in the field, the sequences have

```
      45                  50                  55
Gln-Ile-Lys-Ile-Trp-Phe-Gln-Asn-Arg-Arg-Met-Lys
 95  68 100  95 100 100 100 100 100 100  91 100
GTT TAG TTC TAG ACC AAG GTT TTG GCA GCA TAC TTC  5'
```

Figure 1. Design of the oligonucleotide probe. The upper line is a consensus amino acid sequence compiled for *Antennapedia*-like homeodomain containing proteins; the protein region shown corresponds to the 'recognition helix'. The middle line represents the % amino acid conservation at the particular protein position (as estimated at the time we designed the probe). The lower line is the resulting unique oligonucleotide probe, derived from the upper line by using the customary worm codon bias (13). The usual amino acid numbering convention for homeodomains is used throughout; this convention differs by 1 from that used in (2).

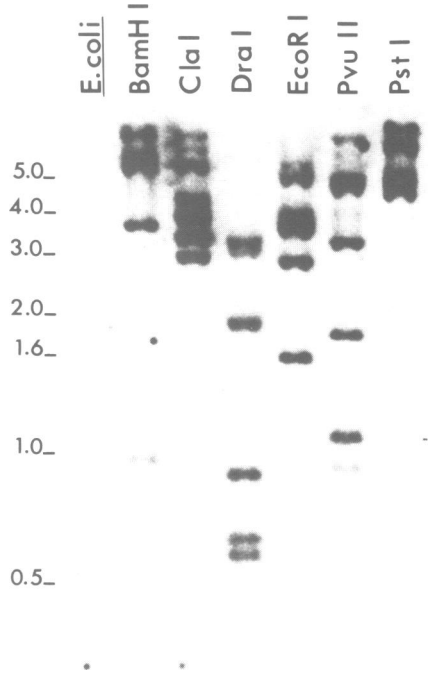

Figure 2. Autoradiogram of a Southern blot of *C. elegans* genomic DNA, hybridized to the ^{32}P-endlabelled 36-mer oligonucleotide shown in Figure 1. Lanes contained 5 μg of *C. elegans* DNA, digested with restriction enzymes as indicated. The lane designated *E. coli* contained 1 μg of *E. coli* DNA digested with the enzyme EcoRI. Hybridization and washing conditions are described in the Methods section. Autoradiographic exposure was 16 hours.

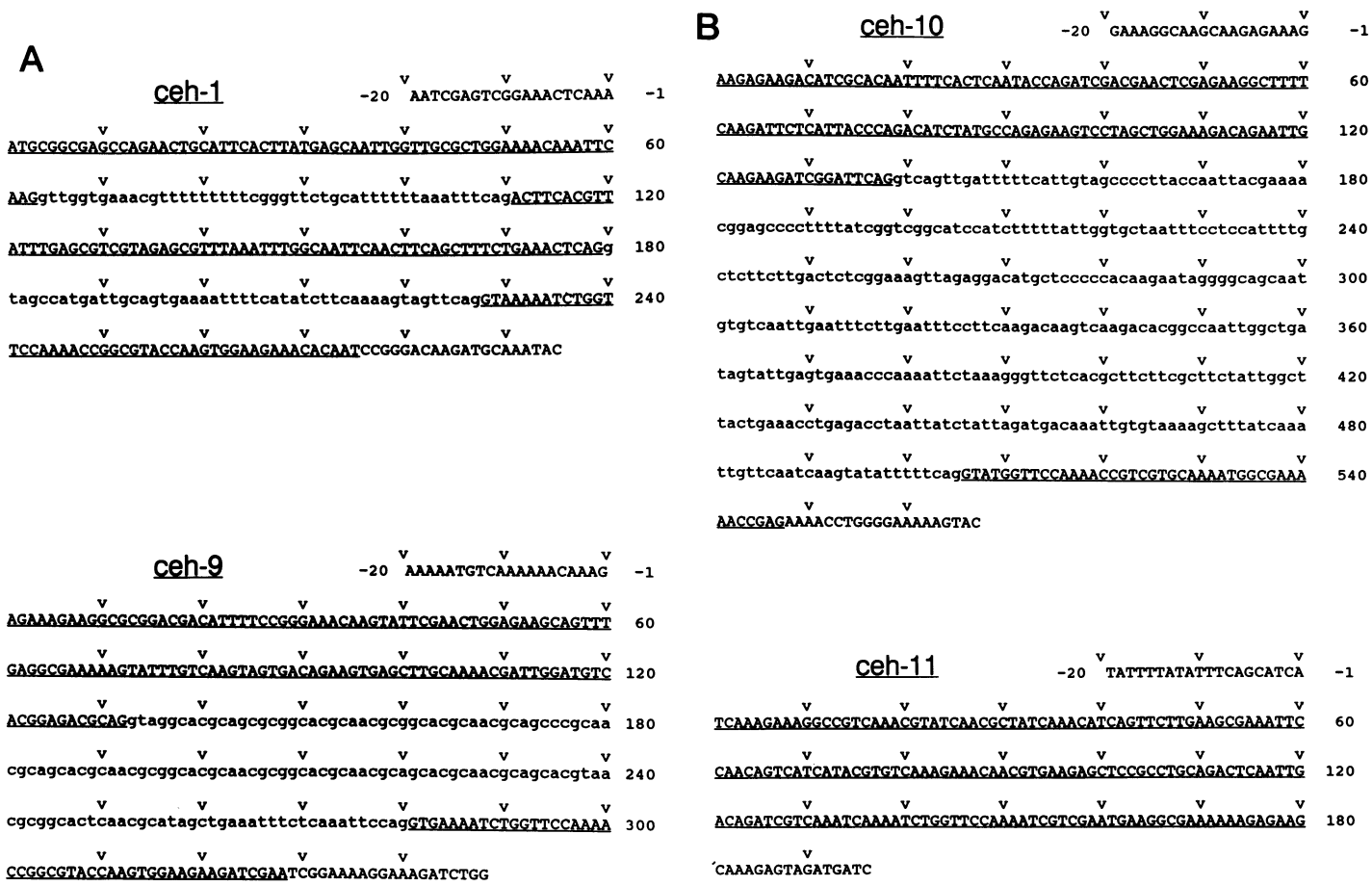

Figure 3. DNA sequence of the four candidate homeo-boxes. The homeobox coding sequences are underlined; the proposed intron sequences are indicated in lower case. Sequence is shown for 20 base pairs on either side of the homeobox sequence, (except for *ceh-11*, in which the cloning site is only 16 bp downstream).

been designated as *ceh-1*, *ceh-9*, *ceh-10* and *ceh-11*; *ceh* stands for *C. elegans homeobox* and the numbers reflect chronology of isolation. (See ref (7 and 9) for other members of the *ceh* class). Three of the four candidate clones contain one or more introns within the homeobox motif and our ability to identify a homeodomain amino acid sequence depends on our ability to recognize splice donor/acceptor sites. Fortunately, worm splicing signals are quite diagnostic and, in all cases, our proposed sites agree closely with the expected consensus sequence (13).

With the intron assignments shown in Figure 3, the four candidate sequences can be translated into amino acid sequences that show obvious similarities to established homeodomains in other proteins. Figure 4 shows the four sequences aligned with the consensus homeodomain sequence compiled in Table II of ref (2). Residues that are 100% conserved in all homeodomains of higher eukaryotes (arrows in Figure 4) are also 100% conserved in the four worm sequences. Residues that are highly (but not absolutely) conserved in all homeodomains are also highly conserved in the worm sequences. We therefore feel justified in claiming that the four sequences do indeed code for authentic homeodomain-containing proteins.

As shown in Figure 5A, the amino acid sequence of *ceh-1* is closely related (87%, 85% and 82% identity, respectively) to the honeybee H40 homeodomain sequence (21), to the *Drosophila* NK-1 homeodomain sequence (22) and to the chicken *CHox3* homeodomain sequence (23). The four sequences show no similarity in the seven amino acids preceding the homeodomain but show weak similarity (4/7, 3/7 and 3/7 matches) in the seven amino acids following the homeodomain. Figure 5A also shows that H40 and *CHox3* are highly similar to the *Drosophila* NK-1 sequence (58/60 and 56/60 matches, respectively); none of these sequences has yet been assigned a genetic function.

The similarity between *ceh-10* and the Drosophila *paired* gene can be increased from 60% to about 70% by allowing conservative amino acid replacements. Figure 5B compares the proposed amino acid sequence of the *ceh-10* gene with the homeodomain region of *paired* (24) and with the three other members of the *paired*-class genes: BSH4 and BSH9 from Drosophila (25) and the *Mix.1* gene isolated from Xenopus (26). It is clear that amino acid residues that tend to be conserved in *paired*-class genes (2) are also conserved in *ceh-10*. This similarity can be extended at least seven amino acids upstream of the homeodomain; the sequences show no similarity immediately downstream of the homeodomain. *Paired*-class genes also contain a second conserved sequence, the paired box, but we have not yet located such a sequence in the *ceh-10* gene. However, between the homeobox and the paired box, the Drosophila *paired*, BSH4 and especially the BSH9 gene contain a short stretch of amino acids that are rich in Ser and Gly residues (27). The twenty amino acids immediately upstream of the *ceh-10* homeobox are 75% Gly or Ser (data not shown), further suggesting that *ceh-10* is a *paired*-class gene.

Beyond the two similarities noted in Figure 5, the best matches that we could obtain with other homeodomains were as follows:

ceh-1 is 55% similar to *zen-2* of Drosophila; *ceh-9* is 53% similar, also to *zen-2*; *ceh-10* is 43% similar to *Hox2.1* of mouse, and; *ceh-11* is 57% homologous to *Antennapedia* class genes from sea urchin (*HBl*), Xenopus (*Xhox-36*), mouse (*Hox2.3*) and Drosophila (*Antp*). None of the four sequences shows particularly strong similarity either to each other or to any of the previously isolated *C. elegans* homeobox sequences; (for example, the best match shows that *ceh-11* is 55% similar to *mab-5*). Overall, the four sequences shown in Figure 4 are only 43−57% similar (at the amino acid level) to the *Antennapedia* sequence (i.e., the sequence that was used to derive the probe)

										10
CONSENSUS	---	---	---	---	ARG	---	---	TYR	---	---
ceh-1	Met	Arg	Arg	Ala	Arg	Thr	Ala	Phe	Thr	Tyr
ceh-9	Arg	Lys	Lys	Ala	Arg	Thr	Thr	Phe	Ser	Gly
ceh-10	Lys	Arg	Arg	His	Arg	Thr	Ile	Phe	Thr	Gln
ceh-11	Ser	Lys	Lys	Gly	Arg	Gln	Thr	Tyr	Gln	Arg
										20
CONSENSUS	---	GLN	---	---	---	LEU	---	---	---	PHE
ceh-1	Glu	Gln	Leu	Val	Ala	Leu	Glu	Asn	Lys	Phe
ceh-9	Lys	Gln	Val	Phe	Glu	Leu	Glu	Lys	Gln	Phe
ceh-10	Tyr	Gln	Ile	Asp	Glu	Leu	Glu	Lys	Ala	Phe
ceh-11	Tyr	Gln	Thr	Ser	Val	Leu	Glu	Ala	Lys	Phe
										30
CONSENSUS	---	---	---	---	TYR	---	---	---	---	---
ceh-1	Lys	Thr	Ser	Arg	Tyr	Leu	Ser	Val	Val	Glu
ceh-9	Glu	Ala	Lys	Lys	Tyr	Leu	Ser	Ser	Ser	Asp
ceh-10	Gln	Asp	Ser	His	Tyr	Pro	Asp	Ile	Tyr	Ala
ceh-11	Gln	Gln	Ser	Ser	Tyr	Val	Ser	Lys	Lys	Gln
										40
CONSENSUS	ARG	---	---	---	ALA	---	---	LEU	---	LEU
ceh-1	Arg	Leu	Asn	Leu	Ala	Ile	Gln	Leu	Gln	Leu
ceh-9	Arg	Ser	Glu	Leu	Ala	Lys	Arg	Leu	Asp	Val
ceh-10	Arg	Glu	Val	Leu	Ala	Gly	Lys	Thr	Glu	Leu
ceh-11	Arg	Glu	Glu	Leu	Arg	Leu	Gln	Thr	Gln	Leu
								▼	▼	50
CONSENSUS	---	---	---	GLN	---	LYS	ILE	TRP	PHE	GLN
ceh-1	Ser	Glu	Thr	Gln	Val	Lys	Ile	Trp	Phe	Gln
ceh-9	Thr	Glu	Thr	Gln	Val	Lys	Ile	Trp	Phe	Gln
ceh-10	Gln	Glu	Asp	Arg	Ile	Gln	Val	Trp	Phe	Gln
ceh-11	Thr	Asp	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln
	▼		▼							60
CONSENSUS	ASN	ARG	ARG	---	LYS	---	LYS	---	---	---
ceh-1	Asn	Arg	Arg	Thr	Lys	Trp	Lys	Lys	His	Asn
ceh-9	Asn	Arg	Arg	Thr	Lys	Trp	Lys	Lys	Ile	Glu
ceh-10	Asn	Arg	Arg	Ala	Lys	Trp	Arg	Lys	Thr	Glu
ceh-11	Asn	Arg	Arg	Met	Lys	Ala	Lys	Lys	Glu	Lys

Figure 4. Alignment of the amino acid sequences for the four candidate *C. elegans* homeodomains with the overall consensus sequence for homeodomains from higher eukaryotes (reference 2; Table IIA). The four arrows indicate residues that are absolutely conserved; the other consensus residues shown are highly but not absolutely conserved.

Both *ceh-1* and *ceh-9* have an intron positioned after amino acid 44 of the homeodomain and *ceh-10* has an intron after amino acid 46. An intron in this region appears to be a common feature of *C. elegans* homeobox sequences. Of the 15 complete or partial *C. elegans* homeobox sequences in which an intron position can be assigned, ten have an intron after either amino acid 43, 44 or 46. This position is within the highly conserved recognition helix (2, 28); thus, any exon-swapping that occurs during evolution would be likely to cause changes in the DNA binding specificity of the homeodomain.

We have used sequences outside the homeobox regions identified above to search for possible homologies with other sequences in the database. One significant (but curious) match was found: the intron of *ceh-9* contains ten 12-base pair repeats that are 84% similar to the repetitive domains of the circumsporozoite genes of the human parasite *Plasmodium malariae* and the monkey parasite *Plasmodium brasilianum* (29). Although there are two open reading frames through this region, (neither of which is a continuation of the proposed homeo-domain reading frame), we have no indication that the sequence appears as protein.

Figure 6 shows that two of the four homeo-box sequences can be detected as RNA transcripts in the *C. elegans* embryo. We first identified restriction fragments that would detect unique (or essentially unique) bands on a genomic Southern (data not shown). Using such a unique probe for *ceh-10*, a 1.25 kb transcript and a minor 1.85 kb transcript can be detected in poly-A+ RNA from wild type *C. elegans* embryos; both species decrease substantially in RNA isolated from larval stages (Figure 6). With a unique probe to a region of *ceh-11*, a weak 1.6 kb RNA species can be detected in embryos (after 11 days exposure); as with the *ceh-10* transcripts, the *ceh-11* transcript diminishes substantially in larvae (Figure 6). With unique probes to *ceh-1* and *ceh-9* (at similar specific activities), we have not yet been

A

	-1	10	20	30	40	50	60
ceh-1	KSSRKLK	MRRARTAFTY	EQLVALENKF	KTSRYLSVVE	RLNLAIQLQL	SETQVKIWFQ	NRRTKWKKHN PGQDANT
NK-1	GGGGGS*	P*********	****S*****	**T*****C*	*****LS*S*	T*********	********Q* **M*V*S
H40	RRWDRRE	A*********	*********	**T*****C*	*****LS*S*	T*********	********Q* **L*VIS
CHox3	AEASCA*	P*********	*********	RAT*****C*	*****LS*S*	T*********	********QH **A*GAA

B

	-1	10	20	30	40	50	60
Prd	GIALKRK	QRRCRTTFSA	SQLDELERAF	ERTQYPDIYT	REELAQRTNL	TEARIQVWFS	NRRARLRKQH TSVSGGA
BSH4	**P****	***S****T*	E**EA*****	S******V**	******T*A*	**********	********HS -------
BSH9	SVQ****	***S*****N	D*I*A***I*	A******V**	******S*G*	****V*****	*********L -------
Mix.1	ASLVPAS	***K**F*TQ	A***I**QF*	QTNM****HH	*****RHIYI	P*S******Q	****KV*R*G AKATKPI
ceh-10	*K*S***	K**H**I*TQ	Y*I****K**	QDSH*****A	**V**GK*E*	Q*D******Q	****KW**TE KTWGKST

Figure 5A. Amino acid sequence of the *ceh-1* homeodomain, aligned with the sequence of the *Drosophila* NK-1 (22), the honeybee H40 (21) and the chicken *CHox3* (23) homeodomains. Asterisks indicate the same residue as *ceh-1*. An additional seven amino acids are shown both 5′ and 3′ to the homeodomain. Figure 5B. Amino acid sequence of the *ceh-10* homeodomain, aligned with the sequence of the four members of the '*paired*' class of homeodomain-containing proteins: *paired* (24), BSH4 and BSH9 (25), and *Mix.1* (26). Asterisks indicate the same residue as *paired*. An additional seven amino acids are shown both 5′ and 3′ to the homeodomain.

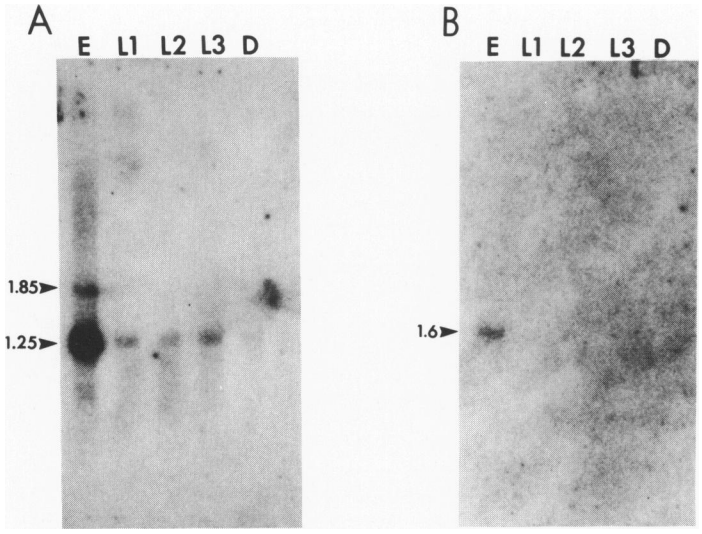

Figure 6. Northern blots of poly-A+ RNA from various developmental stages of *C. elegans*, hybridized with unique probes to either (A) *ceh-10* or (B) *ceh-11*. Each lane contained 5 μg of poly A+ RNA from either embryos (E), larval stages (L1, L2 and L3) or dauer larvae (D), fractionated on a 1.25% agarose gel containing formaldehyde. Embryos were obtained by hypochlorite digestion of gravid adults and thus are younger than 2–3 hours. Estimated RNA sizes in kilobases are indicated. Autoradiographs were exposed for (A) 7 days, or (B) 11 days.

TABLE 1. Genomic Locations of the Four *C. elegans* Homeobox Genes.

Homeobox Number	Original Bacteriophage	Description of Genomic Position
ceh-1	JM#L1001	Left end of the X chromosome on the *act-4* contig.
ceh-9	JM#L1002	Not yet linked to the map, but assigned to YAC.
ceh-10	JM#L1003	Left of Chromosome III, between *ubg-1* and *mlc-3*
ceh-11	JM#L1004	Centre of Chromosome III, tightly linked to *mab-5*.

able to detect RNA transcripts in either embryos or larvae.

The original bacteriophage clones that contained the four homeobox sequences were sent to Drs. Alan Coulson and John Sulston, MRC Labs, Cambridge, to be placed on the current physical map of the *C. elegans* genome (10, 11). Three of the four genes could be assigned map positions, as listed in Table 1. *ceh-11* was found to be tightly linked on the physical map to the gene *mab-5*, previously shown to contain a homeodomain (4). However, the sequence shown in Figure 3 clearly shows that *ceh-11* is distinct from *mab-5*. There is preliminary evidence (A. Chisholm and J. Hodgkin, personal communication) that a cosmid containing *ceh-11* rescues mutations in a gene called *egl-5*, which is known to be closely linked genetically to *mab-5*. The gene *egl-5* has been placed at an important position in the regulatory hierarchy that specifies the identity of a pair of neurons that innervate the hermaphrodite vulva (30). Thus, if this identification of *ceh-11* with *egl-5* is substantiated, once again a homeodomain-containing protein will have turned out to play a major role in determining cell fate.

In summary, we have isolated four new homeobox containing sequences from the nematode *Caenorhabditis elegans*. Based on experience with other organisms, it is probable that these genes, by virtue of this conserved motif, will code for proteins somehow involved in gene regulation. Future steps are obvious: cDNA clones must be isolated, temporal and spatial expression of both transcripts and proteins must be defined and the corresponding genetic loci identified. Only then will we be able to demonstrate that these genes are indeed involved in controlling *C. elegans* development.

ACKNOWLEDGEMENTS

This work was supported by the Medical Research Council of Canada, by the Alberta Heritage Foundation for Medical Research and by a Ralph Steinhauer Studentship to N.H. We should like to thank: B.Goszczynski and F.Allen for completing some of the DNA sequences; Drs. A.Coulson and J.Sulston for placing the clones on the *C. elegans* genomic map; an anonymous referee for pointing out the similarity of *ceh-1* with NK-1; Drs. T.Burglin, A.Chisholm, J.Hodgkin, M.Nirenberg and G.Ruvkun for communicating their results prior to publication, and; Mrs. Claire Berglind for expert preparation of the manuscript.

REFERENCES

1. Johnson, P.F. and McKnight, S.L. (1989) *Ann. Rev. Biochem.* **58**, 799–839.
2. Scott, M.P., Tamkun, J.W. and Hartzell, G.W. (1989) *Biochim. Biophys. Acta* **989**, 25–48.
3. Way, J.C. and Chalfie, M. (1988) *Cell* **54**, 5–16.
4. Costa, M., Weir, M., Coulson, A., Sulston, J. and Kenyon, C. (1988) *Cell* **55**, 747–756.
5. Finney, M., Ruvkun, G. and Horvitz, H.R. (1988) *Cell* **55**, 757–769.
6. Freyd, G., Kim, S.K. and Horvitz, H.R. (1990) *Nature* **344**, 876–879.
7. Burglin, T.R., Finney, M., Coulson, A. and Ruvkun, G. (1989) *Nature* **341**, 239–243.
8. Kamb, A., Weir, M., Rudy, B., Varmus, H. and Kenyon, C. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4372–4376.
9. Schaller, D., Wittman, C., Spicher, A., Muller, F. and Tobler, H. (1990) *Nucleic Acids Res.* **18**, 2033–2036.
10. Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7821–7825.
11. Coulson, A., Waterston, R., Kiff, J., Sulston, J. and Kohara, Y. (1988) *Nature* **335**, 184–186.
12. Brenner, S. (1974) *Genetics* **77**, 71–94.
13. Wood, W.B. (1988) The Nematode Caenorhabditis elegans. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
14. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
15. Hawkins, N.C. (1989) M.Sc. Thesis. University of Calgary, Calgary, Alberta, Canada.
16. Reed, K.C. and Mann, D.A. (1985) *Nucleic Acids Res.* **13**, 7207–7221.
17. Henikoff, S. (1987) *Methods Enzymol* **155**, 156–165.
18. Boissy, R. and Astell, C.R. (1985) *Gene* **35**, 179–185.
19. Okayama, H., Kawaichik, M., Brownstein, M., Lew, F., Yokota, T. and Atia, K. (1987) *Methods Enzymol.* **154**, 3–28.
20. Hodgson, C.P. and Fisk, R.Z. (1987) *Nucleic Acids Res.* **15**, 6295.
21. Walldorf, U., Fleig, R. and Gehring, W.J. (1989) *Proc. Natl. Acad. Sci. (USA)* **86**, 9971–9975.
22. Kim, Y. and Nirenberg, M. (1989) *Proc. Natl. Acad. Sci (USA)* **86**, 7716–7720.
23. Rangini, Z., Frumkin, A., Shani, G., Guttman, M., Eyal-Giladi, H., Gruenbaum, Y., and Fainsod, A. (1989) *Gene 76*, 61–74.
24. Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. and Noll, M. (1986) *Cell* **47**, 735–746.
25. Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. and Noll, M. (1986) *Cell* **47**, 1033–1040.
26. Rosa, F.M. (1989) *Cell* **57**, 965–974.
27. Baumgartner, S., Bopp, D., Burri, M. and Noll, M. (1987) *Genes and Development* **1**, 1247–1267.

28. Qian, Y.Q., Billeter, M., Otting, G., Muller, M., Gehring, W.J. and Wuthrich K. (1989) *Cell* **59**, 573–580.
29. Lal, A.A. de la Cruz, V.F., Collins, W.E., Campbell, G.H., Procell, P.M. and McCutchan, T.F. (1988) *J. Biol. Chem.* **263**, 5495–5498.
30. Desai, C., Garriga, G., McIntire, S.L. and Horvitz, H.R. (1988) *Nature* **336**, 638–646.