Nucleic Acids Research, Vol. 20, No. 12 2967–2969

Identification of a homeobox-containing gene located between *lin-45* and *unc-24* on chromosome IV in the nematode *Caenorhabditis elegans*

Morihiro Naito, Yuji Kohara[1] and Yoshikazu Kurosawa*
Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-11 and
[1]National Institute of Genetics, Yata, Mishima 411, Japan

Received April 10, 1992; Revised and Accepted May 15, 1992 EMBL accession nos Z11794, Z11795

ABSTRACT

Using two primers corresponding to helix 1 and helix 3 regions in the homeodomain, we subjected genomic DNA from *Caenorhabditis elegans* to amplification by the polymerase chain reaction. Sequence analysis of the amplified products revealed a new homeobox-containing gene, designated *ceh-19*. This gene was located between *lin-45* and *unc-24* on chromosome IV where no homeogene has previously been mapped.

INTRODUCTION

The homeobox was first identified as a sequence commonly observed in several homeotic and segmental genes in *Drosophila melanogaster* (1). To date, more than twenty genes involved in the developmental control of *Drosophila* have been shown to contain homeobox sequences (2). Burglin *et al.* (3) reported that the number of homeobox-containing genes (homeogenes) in the nematode *Caenorhabditis elegans* is at least 60. Using five kinds of 500- to 2,000-fold degenerate oligonucleotides which corresponded to helix 3 regions of homeodomains, they identified 49 putative homeogenes. The sequences of 22 of the genes have been determined. Five of the genes, *mec-3* (4), *mab-5* (5), *lin-11* (6), *unc-86* (7), and *unc*-4 (8) were identified from an analysis of mutants and fourteen of them, *ceh-1* through *14*, were obtained from cross-hybridization experiments (3, 9, 10). Kamb *et al.* (11) utilized a strategy involving the polymerase chain reaction (PCR) for analysis of homeogenes of *C.elegans*, and they identified three homeogenes, *ceh-15*, *en-3/5* and *prd-1/2*. Many of the homeogenes have already been mapped to the various chromosomes of *C.elegans* (3). In this study, we applied the PCR method to an identification of homeogenes in *C.elegans*, and we identified a new homeogene. This gene, designated *ceh-19*, was mapped between *lin-45* and *unc-24* on chromosome IV.

MATERIALS AND METHODS

Two kinds of primer were synthesized chemically . Primer I was 5′-GGGGATCCGARYTRGARAARGARTT, where Y denotes T or C and R denotes G or A. This base sequence corresponds to a *Bam*HI site, which is underlined, and the amino acid sequence ELEKEF. Primer II was 5′-GGGTCGACYCKYCKRTTYTG-RAACCA where K denotes G or T. This sequence corresponds to a *Sal*I site, which is underlined, and the amino acid sequence WFQNRR. PCR was performed as described elsewhere (12). In brief, the 100-μl reaction mixture contained 1 μg of genomic DNA, 2 μM of primer, 200 μM of each dNTP, and 2.5 units of DNA polymerase from *Thermus aquaticus* (Takara Co.) in 10 mM Tris-HCl (pH 8.4) / 50 mM KCl / 2.5 mM $MgCl_2$ plus 200 μg/ml gelatin. The mixture was subjected to 30 amplification cycles in a Perkin-Elmer/Cetus Thermocycler: 1 min at 94°C, 2 min at 55°C and 1 min at 74°C. After fractionation by electrophoresis in a 5% acrylamide gel, the bands were eluted from the gel, digested with both *Bam*HI and *Sal*I, and cloned into Bluescript vectors (Stratagene). Nucleotide sequences were determined by the chain-termination method (13). A cDNA library was made from a mixed-stage population of *C.elegans* using the ZAP cDNA synthesis kit (Stratagene). A YAC 'polytene filter' and YAC clones of *C.elegans* (14) were generous gifts from Dr. A. Coulson. The library was screened by the Benton-Davis method (15).

RESULTS AND DISCUSSION

Identification of a new homeogene, *ceh-19*, in *C.elegans*

Homeoboxes are regions of 60-amino acids with a helix-turn-helix structure (16, 17). The most highly conserved region is located in the C-terminal third of the homeodomain (helix 3 region) (2). In addition to this region, several positions in the helix 1 region are also conserved (2). As primer sequences for PCR, we chosed two blocks, ELEKEF from amino acids 15 to 20 and WFQNRR from amino acids 48 to 53. Similar sequences have already been used as primers for the identification of homeogenes in various animals by several groups (11, 18, 19). The frequencies of the amino acid residues in these two blocks within the known homeodomains of *C.elegans* are as follows: E(8/15), L(15/15), E(13/15), K(6/15), E(4/16) and F(15/16); W(20/20), F(20/20), Q(18/20), N(20/20), R(19/20) and R(20/20). Although the first block appears rather divergent, four

* To whom correspondence should be addressed

```
 P  L  Y  S  H  L  Y  N  M  Y  S  R  K  H  D  Y  K  R  S  R  K  S  V  C  E  R  K  P  R  Q  A
CCATTATATTCACACCTCTATAACATGTATTCTCGAAAGCATGATTACAAGAGAAGCAGAAAATCTGTCTGCGAAAGAAAACCACGACAAGCG    93
 Y  S  A  R  Q  L  D  R  L  E  T  E  F  Q  T  D  K  Y  L  S  V  N  K  R  I  Q  L  S  Q  T  L
TATTCAGCAAGACAGTTGGATAGATTGGAAACGGAATTTCAAACTGACAAATATCTGAGTGTGAACAAGAGAATCCAATTGTCTCAGACATTG   186
 N  L  T  E  T  Q  I  K  T  W  F  Q  N  R  R  T  K  W  K  K  Q  L  T  S  S  I  R  Q  M  V  K
AATCTAACTGAGACCCAGATAAAAACATGGTTTCAAAATAGACGAACAAAATGGAAGAAACAATTAACATCATCAATACGTCAAATGGTAAAA   279
 D  A  P  T  S  T  S  V  G  V  P  F  Q  S  L  L  T  P  P  T  P  P  T  T  L  A  C  H  V  N  S
GATGCTCCAACATCCACGTCAGTTGGTGTTCCTTTTCAGTCACTTCTCACTCCGCCCACTCCACCTACAACTCTTGCCTGTCACGTCAACTCC   372
 L  F  A  C  E  Q
CTGTTCGCTTGTGAGCAATAGCTTCCTCATGTTTATGAAATTTATATTACATTGAGAATTCTATTGCACTTTTTTAATCTCGAAACATGTTAA   465
ATACTCCAGTACAAAATTGATAAAAGAAAATATTTTTTCGTAATAAAAATTGTTGAGAACCaaaaaaaaaaaaaaaaaaaaaa             548
```

Fig. 1. Nucleotide and amino acid sequences of *ceh-19*. Positions of introns are indicated by closed triangles. The homeobox region is boxed. The polyA-addition signal is underlined. The accession number is Z11795.

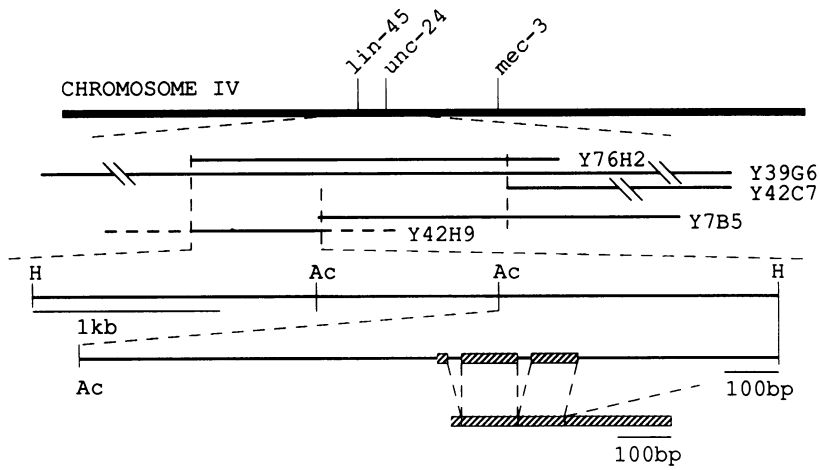

Fig. 2. Physical map indicating the location of *ceh-19* on chromosome IV of *C. elegans*. In the hybridizations of YAC clones with the *ceh-19* probe, Y76H2, Y39G6 and Y42H9 were positive but Y42C7 and Y7B5 were negative. The 4-kb *Hin*dIII fragment from Y42H9 was recloned and the nucleotide sequence of *Acc*I-*Hin*dIII fragment was determined (accession number, Z11794). The exon encoding the 3'-terminal region of *ceh-19* was not included in this fragment. The polarity of this gene was not determined.

```
                        helix 1          helix 2          helix 3  helix 4
Antp     RKRGRQTYTR YQTLELEKEF HFNRYLTRRR RIEIAHALCL TERQIKIWFQ NRRMKWKKEN Chromosome

mec-3    -RGP-T-IKQ N-LDV-NEM- SNTPKPSKHA -AKL-LETG- SM-V-QV--- ---S-ERRLK  IV
mab-5    S--T----S- S--------- -YHK----K- -Q--SET-H- ----V----- -----H---A  III
lin-11   -RGP-T-IKA K-LET-KNA- AATPKP--HI -EQL-AETG- NM-V-QV--- ---S-ERRMK  I
unc-86   K--K-TSIAA PEKR---QF- KQQPRPSGE- IAS--DR-D- KKNVVRV--C -Q-Q-Q-RDF  III
unc-4    -R-T-TNFSG W-LE---SA- EASH-PDVFM -EAL-MR-D- L-SRVQV--- ---A--R-RE  II

ceh-1    MR-A-TAF-Y E-LVA--NK- KTS---SVVE -LNL-IQ-Q- S-T-V----- ---T----H-  X
ceh-2                                                V-V--- ---T-H-RVR  I
ceh-3    ADKY-MV-SD --R------- -TSPFI-SD- KSQLSTM-S- ---------- --        III
ceh-4                                                V--- ---A--R-RE  II
ceh-5    P--P-TDNAD E-LEK--ES- NTSG--SGST -AKL-ES-G- SDN-V-V--- ---T-Q--ID  I
ceh-6    KRKK-TSIEV NVKSR--FH- QS-QKPNAQE ITQV-ME-Q- EKEVVRV--C ---Q-E-RIA  I
ceh-7    IP-R-T-F-V E-LYL--MY- AQSQ-VGCDE -ERL-RI-S- D-Y-V----- ---IRMRR-A  II
ceh-8                                                V--- ---A-FRRQE  I
ceh-9    --KA-T-FSG K-VF----Q- EAKK--SSSD -S-L-KR-DV --T-V----- ---T----IE
ceh-10   KR-H-TIF-Q --ID----A- QDSH-PDIYA -EVL-GKTE- Q-DR-QV--- ---A--R-TE  III
ceh-11   S-K-----Q- ---SV--AK- QQSS-VSKKQ -E-LRLQTQ- -D-------- -----A---K  III
ceh-12                                               ------ -----N-RCP  I
ceh-13   NGTN-TNF-T H-LT------ -TAK-VN-T- -T---SN-K- Q-A-V----- -----E--RE  III
ceh-14                  AY QTSSKPA-HV -EQL-SETG- DM-VVQV--- ---A-E-RLK  X
ceh-15   E--Q-TA--- N-V------- -THK----K- ---V--S-M- ----V----- -----H----  III
en-3/5                     ----EK- -Q-L--E-G- N-S-
prd-1/2                          T -E-L-MRID- --

ceh-19   ERKP--A-SA R-LDR--T-- QTDK--SVNK --QLSQT-N- --T---T--- ---T----QL  IV
```

Fig. 3. Comparison of amino acid sequences of the homeoboxes of *C. elegans* The amino acid sequence of the *Antp* homeobox is included for reference. Bars indicate amino acids identical to those of the *Antp* homeobox. The references for these sequences are indicated in the text. Closed triangles indicate the positions of introns.

homeodomains, those of *ceh-3*, *-13* and *-15* and *mab*-5, have ELEKEF sequences. Genomic DNA of *C.elegans* was subjected to amplification by PCR. In the case of *C.elegans*, short introns are frequently found to have been inserted into homeobox-encoding regions. The products of PCR were of heterogeneous size and were distributed as bands from 130 to 460 bp on a 5% acrylamide gel. Each band was eluted from the gel, cloned and sequenced. Sixteen clones contained homeobox sequences and were classified into four types. Three of them corresponded to published genes, *ceh-3* (3), *ceh-13* (3) and *ceh-15* (7). The numbers in parentheses indicate the numbers of isolated clones. We did not obtain a clone that corresponded to *mab*-5, contrary to our expectations. The other three clones were identical and had a novel sequence. Using this DNA as probe, we screened a cDNA library constructed from mRNAs that had been prepared from a mixture of cells that represented all the developmental stages of *C.elegans*. Three positive clones were isolated. Figure 1 shows the nucleotide sequence and the deduced amino acid sequence of one of them, λCeh19-1, which was the longest cDNA clone. It contains a typical homeobox. As summarized in Fig. 3, we concluded that the newly identified gene, *ceh-19*, is a new homeogene. We also isolated a genomic clone and determined its nucleotide sequence. The coding sequence was interrupted by three introns, as indicated in Fig. 1.

Localization of *ceh-19* between *lin-45* and *unc-24* on chromosome IV

Nearly all of the total genome of *C.elegans* has already been covered by YAC clones, DNAs of which were dotted on a YAC 'polytene' filter (14). Hybridization of the *ceh-19* probe with this filter revealed that two YAC clones, Y76H2 and Y39G6, hybridized with this probe Another YAC clone, Y42C7, was negative. To further localize this gene, two other YAC clones which covered the overlap between Y76H2 and Y39G6 were analyzed . As indicated in Fig. 2, the *ceh-19* gene was located on Y42H9. This locus corresponds to a region between *lin-45* and *unc-24* on chromosome IV. As summarized in Fig. 3, many of the homeogenes have been mapped. *C.elegans* has six chromosomes, I to V plus X. On chromosome IV, only *mec-3* has been mapped to date, although three cosmid clones mapped on chromosome IV possibly containing homeogenes were reported (3).

CONCLUSION

Using PCR, we isolated a new homeogene from *C.elegans*. Preconditions for PCR are apparently lower than those for cross-hybridization. Even if the sequences of the primers do not perfectly match those of the target gene, provided that several nucleotides of the 3′ ends are complementary to the template DNA, a particular DNA can be amplified. In fact, *ceh-19* encoded the sequence RLETEF instead of ELEKEF. The advantages of a strategy incorporating PCR over the cross-hybridization method are not only high sensitivity with respect to detection but also direct acquisition of sequence information. With further appropriate primer sequences, more homeogenes will be identified in *C.elegans*, as predicted by Burglin *et al.* (3).

ACKNOWLEDGEMENTS

We thank Dr. A.Coulson for providing us with YAC clones of *C.elegans*. This work was supported in part by grants from the Ministry of Education, Science and Culture in Japan; Fujita Health University.

REFERENCES

1. McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984) Nature **308**, 428–433.
2. Scott, M.P., Tamkun, J.W. and Hartzell III, G.W. (1989) Biochem. Biophys. Acta **989**, 25–48.
3. Burglin, T.R., Finney, M., Coulson, A. and Ruvkun, G. (1989) Nature **341**, 239–243.
4. Way, J.C. and Chalfie, M. (1988) Cell **54**, 5–16.
5. Costa, M., Weir, M., Coulson, A., Sulston, J. and Kenyon, C. (1988) Cell **55**, 747–756.
6. Freyd, G., Kim, S.K. and Horvitz, H.R. (1990) Nature **344**, 876–879.
7. Finney, M., Ruvkun, G. and Horvitz, H.R. (1988) Cell **55**, 757–769.
8. Miller, D.M., Shen, M.M., Shamu, C.E., Burglin, T.R., Ruvkun, G., Dubois, M.L., Ghee, M. and Wilson, L. (1992) Nature **355**, 841–845.
9. Schaller, D., Wittman, C., Spicher, A., Muller, F. and Tobler, H. (1990) Nucleic Acids Res. **18**, 2033–2036.
10. Hawkins, N.C. and McGhee, J.D. (1990) Nucleic Acids Res. **18**, 6101–6106.
11. Kamb, A., Weir, M., Rudy, B., Varmus, H. and Kenyon, C. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 4372–4376.
12. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science **239**, 487–491.
13. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467.
14. Coulson, A., Waterston, R., Kiff, J., Sulston, J. and Kohara, Y. (1988) Nature **335**, 184–186.
15. Benton, W.D. and Davis, R.W. (1977) Science **196**, 180–182.
16. Otting, G., Qian, Y.Q., Billeter, M., Muller, M., Affolter, M., Gehring, W.J. and Wuthrich, K. (1990) EMBO J. **9**, 3085–3092.
17. Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo C.O. (1990) Cell **63**, 579–590.
18. Frohman, M.A., Boyle, M. and Martin, G.R. (1990) Development **110**, 589–607.
19. Murtha, M.T., Leckman, J.F. and Ruddle, F.H. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 10711–10715.